



System-of-systems modeling and simulation for the risk analysis of industrial installations and critical infrastructures

Elisa Ferrario

► To cite this version:

Elisa Ferrario. System-of-systems modeling and simulation for the risk analysis of industrial installations and critical infrastructures. Engineering Sciences [physics]. Ecole Centrale Paris, 2014. English. NNT : 2014ECAP0046 . tel-01127194

HAL Id: tel-01127194

<https://theses.hal.science/tel-01127194>

Submitted on 7 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**ÉCOLE CENTRALE DES ARTS
ET MANUFACTURES
« ÉCOLE CENTRALE PARIS »**

THÈSE
présentée par

Elisa FERRARIO

pour l'obtention du

GRADE DE DOCTEUR

Spécialité : Génie Industriel

Laboratoire d'accueil : Laboratoire de Génie Industriel

SUJET :

**System-of-systems modeling and simulation for the risk analysis
of industrial installations and critical infrastructures**

**Simulation et modélisation de système des systèmes pour l'analyse des risques
des installations industrielles et des infrastructures critiques**

soutenue le :10 septembre 2014

devant un jury composé de :

Terje Aven	University of Stavanger	Reviewer
Frank Guarnieri	MINES ParisTech	Reviewer
Mohamed Hibti	Électricité de France R&D	Examiner
Alois J. Sieber	European Reference Network for Critical Infrastructure Protection	Examiner
Enrico Zio	École Centrale Paris	Supervisor

2014ECAP0046

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my thesis supervisor Professor Enrico Zio that believed in me from the beginning. He gave me the opportunity to pursue my Ph. D. studies, helping me to start my path and develop my skills as a researcher. During these three years, he has encouraged me to work ever harder and take on new challenges with his strong enthusiasm and passion for the scientific research and his rigorous organization of the work.

I would like to extend my sincere gratitude to my co-supervisor Doctor Nicola Pedroni for being always present and available to discuss, for his insightful advice, patience and great support (also human) that has helped me to overcome the difficulties encountered along this way. It has been interesting, fruitful and a pleasure working with him.

My deepest appreciation goes to all the jury members, Professor Terje Aven, Professor Frank Guarnieri, Doctor Mohamed Hibti and Doctor Alois J. Sieber, who agreed to be part of the committee. I am very pleased to have presented my Ph. D. work in front of such high qualified jury. Special thanks to the reviewers, Professors Terje Aven and Frank Guarnieri, for the evaluation of the manuscript and all the constructive and helpful remarks.

I would like to acknowledge Professor Jean-Claude Bocquet, Director of the Industrial Engineering Laboratory (LGI), for a three-year hosting: it has been a pleasure working in this lively and culturally diverse environment.

My warmest recognition goes to Corinne Ollivier, Delphine Martin and Sylvie Guillemain, the secretaries of LGI, because they have been lovely in helping me to overcome the initial difficulties of integration in France and their kind work has eased mine.

I would like to thank my friends and colleagues of LGI with whom I have spent unforgettable moments. In particular, I remember my office mates Yiping Fang and Tairan Wang with whom I have shared the daily successes and failures of my research activities and the associated feelings of happiness and frustration. They have been great friends to me, ready to

give me a hand if needed, to close the window if I was cold, and to remind me to drink hot water and walk a little bit after long hours spent in front of the computer. 谢谢 (xiè xie).

From the bottom of my heart, thank to Rodrigo that has never stopped to encourage me. With his smile and jokes he has made this journey lighter and with his world view and opinions he has contributed to broaden my horizons.

Finally, I wish to dedicate this thesis to my family that has always been there for me with a truthful support. In particular, to my parents that have been my example and they have taught me fundamental values like commitment, honesty and thoughtfulness, important for the achievement of any goal in my life like this Ph. D. degree; and to my brother that with his openness to experiences, his ability of taking initiative and his love for travelling boosted me to do my doctoral studies abroad. The last words are for my grandmother that when I was a child she always persuaded me to study (opportunity that she could not have), be good at school and become a teacher.

ABSTRACT

This thesis propounds and develops a system-of-systems (SoS) framework for the risk analysis of industrial installations and critical infrastructures. System representation, modeling and simulation methods are developed to capture the peculiar features of SoS, with respect to their vulnerability and physical resilience to random failures and natural hazards. Several representation techniques of literature, i.e., Fault Tree, Muir Web, Hierarchical Modeling, Goal Tree Success Tree – Dynamic Master Logic Diagram, are explored and originally extended/tailored to fit the purpose of SoS analysis. One representation method is developed ex-novo, namely the Hierarchical Graph. Within these representation frameworks, binary and multiple states are used to model the performances of the SoS under analysis. Monte Carlo simulation and interval analysis are combined for the quantitative evaluation of the SoS models in presence of uncertainty (due to both randomness and lack of knowledge). Examples of analyses are carried out within two application areas: external event risk assessment and vulnerability of critical infrastructures.

Keywords: system-of-systems, critical infrastructures, risk analysis, natural external events, physical resilience, vulnerability, system representation, binary and multi-state models, Monte Carlo simulation, interval analysis, uncertainty, random failures, Fault Tree, Muir Web, Hierarchical Modeling, Goal Tree Success Tree – Dynamic Master Logic Diagram, Hierarchical Graph.

RÉSUMÉ EN FRANÇAIS

Le travail de recherche propose et développe un cadre de système des systèmes (SdS) pour l'analyse de risques des installations industrielles et des infrastructures critiques. Les méthodes pour la représentation, la modélisation et la simulation d'un système sont développées pour identifier les particularités du SdS quant à leur vulnérabilité et leur résilience physique à des défaillances aléatoires et risques naturels. Plusieurs techniques de représentation, telles que l'arbre de défaillances, le *Muir Web*, la modélisation hiérarchique, le *Goal Tree Success Tree – Dynamic Master Logic Diagram*, sont étudiées et approfondies depuis l'origine pour s'adapter aux objectifs de l'analyse de SdS. Une méthode de représentation est développée *ex novo*, à savoir, le graphe hiérarchique. Dans ces cadres de représentation, des états binaires et multiples sont utilisés pour modéliser les performances des SdS à analyser. La simulation Monte Carlo et l'analyse d'intervalle sont combinées pour évaluer quantitativement des modèles de SdS en présence d'incertitude (due à la variabilité naturelle d'un phénomène ou au manque d'information). La mise en œuvre de ces approches est illustrée dans deux domaines d'application : l'évaluation du risque d'événements externes et la vulnérabilité d'infrastructures critiques.

Mots clés : système des systèmes, infrastructures critiques, analyse de risques, événements externes naturels, résilience physique, vulnérabilité, représentation du système, états binaires et multiples, simulation Monte Carlo, analyse d'intervalle, incertitude, défaillances aléatoires, arbre de défaillances, *Muir Web*, modélisation hiérarchique, *Goal Tree Success Tree – Dynamic Master Logic Diagram*, graphe hiérarchique.

RÉSUMÉ ÉTENDU EN FRANÇAIS

Contexte général et objectif du travail

Les changements sociaux, politiques, économiques et culturels parallèlement au développement de la technologie ont mené le monde à devenir de plus en plus complexe et interdépendant, et en évolution constante. Les progrès technologiques (par exemple, des télécommunications) ont déterminé une dépendance sur le fonctionnement des systèmes vitaux (par exemple, le réseau électrique) dont les performances ont été considérablement améliorées. Cependant, l'augmentation des relations complexes entre les différents systèmes crée de nouvelles vulnérabilités : une défaillance d'un système peut se propager et provoquer une panne dans un système connecté, conduisant à des effets en cascade qui peuvent frapper des zones aussi très loin de la zone d'impact [Nozick et al., 2005 ; Bouchon, 2006].

Il est une croyance émergente selon laquelle une approche holistique devrait être prise en compte afin de comprendre et de réduire ces vulnérabilités, adoptant un point de vue de système des systèmes (SdS) [Bouchon, 2006; Katina and Keating, 2014].

Lorsque l'on aborde la vulnérabilité et l'analyse de risques de SdS, on fait face aux « vieux » problèmes qui se développent dans de nouveaux défis quant à la représentation et à la modélisation de SdS, la quantification des modèles de SdS, la représentation appropriée et la quantification de l'incertitude du comportement et de la modélisation de SdS, ainsi que la propagation de l'incertitude sur la réponse de SdS.

L'objectif de la présente thèse de doctorat se porte sur la représentation, la modélisation et la simulation de SdS quant à leur vulnérabilité et résilience physiques aux défaillances aléatoires et aux risques naturels. L'application de cette recherche concerne les infrastructures critiques. Le travail a été effectué au Laboratoire Génie Industriel de l'École Centrale Paris dans la Chaire Sciences des Systèmes et Défis Énergétiques, Fondation Européenne pour les Énergies de Demain – Électricité de France – École Centrale Paris – Supélec, France.

Problématiques et définitions

Le concept de « système » a plus d'acceptation universelle, cependant, la définition des « systèmes de systèmes » dépend des domaines d'application et de leurs objectifs.

Un système est un groupe d'éléments (ou des sous-systèmes) qui interagissent et possèdent une structure interne qui les relie en un ensemble unifié. La frontière d'un système doit être définie, ainsi que la nature (physique, logique, etc.) de la structure interne reliant ses éléments [Dupuy, 1985; Kröger and Zio, 2011]. Par ailleurs, il existe nombreuses définitions de SdS ; Boardman en a comparé plus de quarante [Boardman et al., 2006].

Maier a identifié cinq propriétés (appelées également « critères de Maier ») pour caractériser les SdS [Maier, 1996 ; Maier, 1998] : indépendance opérationnelle (chaque système est indépendant et atteint ses fins par lui-même), indépendance managériale (chaque système est géré en grande partie pour ses propres fins plutôt qu'aux fins d'un SdS), dispersion géographique (le SdS est distribué sur une large zone géographique), comportement émergent (le SdS exécute des fonctions coopératives et réalise des tâches précises qui ne résident pas dans n'importe quel composant système) et développement évolutif (le développement du SdS évolue dans le temps suivant la structure, la fonction, l'ajout ou la suppression de composants systèmes).

Dans le contexte de cette recherche, on considère la définition suivante de DeLaurentis [2007] qu'encapsule les critères de Maier et capture des aspects supplémentaires comme l'hétérogénéité de composants des systèmes et la structure à plusieurs niveaux : « un SdS se compose de systèmes d'exploitation multiples, hétérogènes, dispersés, parfois indépendants, intégrés dans les réseaux à plusieurs niveaux et qui évoluent dans le temps » [DeLaurentis, 2007].

Les pratiques d'ingénierie traditionnelles résultent insuffisantes pour construire et gérer les traits des systèmes de systèmes. Selon [Fisher, 2006], les systèmes monolithiques dépendent d'un contrôle central, d'une visibilité globale, des structures hiérarchiques, et des activités coordonnées, mais ces caractéristiques ne peuvent pas être attendues dans les SdS qui sont liés à un contrôle distribué, à la coopération, à l'influence, aux effets en cascade, aux comportements émergents [Béjar et al., 2009]. Pour faire face à ces questions, de nouvelles approches doivent être identifiées.

Les SdS ont été étudiés dans de nombreux domaines d'application ; dans le présent travail, nous nous focalisons sur les infrastructures critiques.

Les infrastructures critiques sont essentielles pour le bien-être de la société moderne car elles fournissent les biens (comme l'énergie, l'eau, les données) et services (comme les réseaux de transport, les services financiers et la santé) à travers les frontières locales, régionales et nationales. Ces infrastructures sont de plus en plus automatisées et fortement interconnectées en raison de leur expansion croissante à grandes échelles et aux progrès dans la technologie de l'information ; par exemple, le fonctionnement des réseaux électriques, qui sont distribués et constitués d'une variété de technologies de production comme le nucléaire, la thermique et l'hydraulique, est uniquement possible en utilisant significativement des systèmes d'information et de communication [Gheorghe and Schlapfer, 2006]. Si, d'un côté, ces progrès ont augmenté l'efficacité des infrastructures critiques, ils ont créé d'autre part de nouvelles vulnérabilités à des défaillances aléatoires, des risques naturels et des événements malveillants. En effet, dans les dernières décennies, un nombre accru d'événements perturbateurs (événements externes d'origine naturelle, actes de malveillance et pannes à grande échelle) affectant les infrastructures critiques a eu lieu : par exemple, la tempête de verglas au Canada (en 1998 et 2013, probablement les deux tempêtes de verglas les plus catastrophiques de notre ère), les attaques du WTC (New York, 2001), l'explosion de l'usine AZF (Toulouse, 2001), le blackout de l'Amérique du Nord (est des USA et Canada, 2003), les attaques terroristes dirigées contre le transport ferroviaire (Madrid, 2004 et Londres, 2007), l'ouragan en Floride (2004), les inondations au Royaume-Uni (2007), le séisme de magnitude 8.8 au Chili (2010), le séisme de magnitude 9.0 et le tsunami (Japon, 2011), etc.

Les infrastructures critiques sont de nature diverse, par exemple, des systèmes physiques, cybernétiques ou organisationnels [Kröger and Zio, 2011]. Dans cette recherche, les infrastructures physiques critiques en réseau (ICs) sont analysées ; par exemple, elles sont celles qui fournissent :

- l'énergie (la fourniture d'électricité, de pétrole et de gaz) ;
- le transport (ferroviaire, routier, aérien, de marchandise) ;
- l'information et les télécommunications (comme Internet) ;
- l'eau potable, y compris le traitement des eaux usées.

Katina et Keating ont proposé une perspective basée sur le concept de SdS pour traiter les ICs [Katina and Keating, 2014]. Tout particulièrement, ils ont classé les ICs interconnectés comme des systèmes de systèmes, montrant qu'ils présentent les caractéristiques de Maier illustrées ci-dessus.

Dans ce point de vue, pour évaluer la vulnérabilité et la résilience des ICs, de nouveaux défis émergent quant à l'analyse des interdépendances et à la représentation, à la modélisation et à la simulation du SdS. En particulier, l'identification des interconnexions est une tâche difficile surtout pour les « interdépendances cachées » qui sont invisibles pendant le fonctionnement normal et deviennent évidentes et critiques en cas d'urgence. Pour cette raison, il est difficile de les identifier avant l'occurrence d'un événement perturbateur et, généralement, elles sont dérivées a posteriori à partir de leurs conséquences.

Le concept de vulnérabilité a été introduit pour mettre au centre la perception du danger des catastrophes dont la compréhension est limitée en ce qui concerne le risque. Il est convenu que l'estimation des probabilités de défaillance adoptées dans l'analyse de risques pour éclairer les décisions de gestion de risques peut être faible à cause du manque d'information et des hypothèses inappropriées et, en outre, pour la survenance des événements inattendus, comme les mécanismes de défaillance inconnus [Johansson and Hassel, 2010].

Il y a deux principales interprétations de la vulnérabilité : une est liée à une propriété globale du système et l'autre à la qualification des composants du système. Dans la première interprétation (plus proche de la définition du risque), l'objectif est d'évaluer l'extension des conséquences adverses causées par la survenance d'un événement dangereux spécifique [Johansson and Hassel, 2010; Kröger and Zio, 2011] ; dans la seconde, le but est d'identifier les composants critiques comme ceux qui, en cas de défaillance, provoquent de grandes conséquences négatives sur le système.

Il existe un large éventail d'approches pour l'évaluation de la vulnérabilité des ICs ; cependant, l'analyse de ces SdS ne peut pas être effectuée uniquement par des méthodes classiques basées sur la décomposition et sur l'analyse de la logique du système : un cadre pour intégrer un certain nombre de méthodes capables de déceler le problème à partir de différentes perspectives comprenant les incertitudes existantes est nécessaire [Zio, 2014]. Il existe quatre principales perspectives. Elles considèrent [Zio, 2014] : 1) les méthodes structurelles/topologiques utilisées pour décrire la connectivité d'un système complexe et

l'analyse de ses effets sur la fonctionnalité du système, sur la propagation en cascade d'un échec et sur sa récupération (résilience) et pour identifier les composantes du système qui doivent être plus contrôlées à cause de leur rôle central dans le système ; 2) les méthodes logiques capables d'identifier la logique de fonctionnement/dysfonctionnement d'un système complexe et les combinaisons des défaillances des composants (matériel, logiciel et humain) qui conduisent à la perte de la fonction du système ; 3) les méthodes phénoménologiques/fonctionnelles capables d'identifier la dynamique de fonctionnement entre les composants (matériels, logiciels et humains) d'un système complexe et avec l'environnement, à partir duquel le fonctionnement dynamique du système émerge, et 4) les méthodes de flux basées sur des modèles mécanistes détaillés (et des codes informatiques) des processus qui se produisent dans le système et qui sont capables de décrire la physique du fonctionnement du système, sa surveillance et son contrôle.

Le concept de résilience varie selon la discipline et l'application [Henry and Ramirez-Marquez, 2012 ; Ouyang et al., 2012]. La résilience peut être décrite comme la capacité du système à réduire les risques de choc, à absorber un choc s'il se produit et à récupérer rapidement après un choc [Bruneau et al., 2003] ; elle est caractérisée par quatre propriétés telles que la robustesse (capacité de résister à un certain stress) ; la redondance (mesure dans laquelle les éléments de SdS peuvent se substituer les uns aux autres) ; la capacité de prise en charge (capacité de mobiliser les ressources nécessaires afin de respecter ses priorités lorsqu'on est menacé) ; et la rapidité (capacité de respecter ses priorités et d'atteindre ses objectifs rapidement et efficacement) ; et par quatre dimensions interdépendantes comme les aspects techniques, organisationnels, sociaux et économiques [Bruneau et al., 2003].

En général, les évaluations du risque, de la vulnérabilité et de la résilience des systèmes sont réalisées, d'abord, par une représentation appropriée et une modélisation de la réalité et, ensuite, par une quantification des métriques définies, souvent à travers d'un processus de simulation.

Donc, la représentation, la modélisation et la simulation du système sont les trois étapes fondamentales qui doivent être effectuées pour capturer l'essence du système considéré et répondre aux questions spécifiques. En fait, le principal défi est de comprendre, prévoir, évaluer des variables, des mesures, des indicateurs qui peuvent donner des informations sur le

comportement d'un système qui est caractérisé par des entrées/paramètres spécifiques et qui est modifié dans le temps par des actions, des événements, des phénomènes physiques.

La représentation du système est adoptée pour identifier les principales caractéristiques du système réel et mettre en évidence les connexions structurelles, logiques et fonctionnelles entre les composants. La modélisation du système consiste en la construction d'un modèle mathématique qui décrit les actions, les événements et les phénomènes physiques qui peuvent provoquer des défaillances du système. Enfin, la simulation est consacrée à simuler le comportement du système d'intérêt dans diverses conditions (par exemple, des transitions opérationnelles et scénarios d'accidents) [USNRC, 2009].

Cependant, en pratique, toutes les caractéristiques du système ne peuvent pas être totalement décrites dans le modèle mathématique : en conséquence, l'incertitude est toujours présente dans les valeurs des paramètres et des variables d'entrée du modèle. Cela se traduit par la variabilité dans les résultats du modèle dont l'incertitude doit être estimée pour une évaluation réaliste du risque lié au fonctionnement du système.

Pour le traitement de l'incertitude dans l'analyse de risque, il existe la distinction classique entre incertitude aléatoire (objectif, stochastique, irréductible) et épistémique (subjective, réductible). La première est due à la variabilité intrinsèque du comportement du système, tandis que la seconde est due au manque de connaissances et des informations sur le système [Apostolakis, 1990; Helton, 2004]. L'incertitude aléatoire est liée à des phénomènes aléatoires, comme la survenance d'événements imprévus (par exemple, la défaillance d'un composant mécanique) qui définissent les différents scénarios d'accidents possibles ; alors que l'incertitude épistémique résulte d'un manque de connaissance des phénomènes et des processus, ou de la pénurie de données opérationnelles et expérimentales disponibles [Ferson and Ginzburg, 1996; Helton and Oberkampf, 2004].

Traditionnellement, les deux types d'incertitude sont représentés par des distributions de probabilités. Cependant, une représentation probabiliste de l'incertitude épistémique ne peut pas être possible quand une quantité suffisante de données n'est pas disponible pour l'analyse statistique. En conséquence, des méthodes alternatives ont été développées comme l'analyse d'intervalle, la théorie de l'évidence, la théorie de la possibilité et la théorie des intervalles flous [Klir and Yuan, 1995 ; Aven and Zio, 2011 ; Aven et al., 2014].

Dans ce travail, la théorie des probabilités est utilisée pour représenter l'incertitude aléatoire : par exemple, la magnitude du séisme est représentée par une distribution exponentielle tronquée, et le temps de récupération d'un composant est décrit par une distribution log-normale ; par contre, les intervalles sont utilisés pour représenter l'incertitude épistémique (par exemple, dans les transitions de probabilités entre différents états du composant).

L'incertitude aléatoire est propagée à travers le modèle mathématique par la simulation Monte Carlo [Kalos and Whitlock, 1986; Zio, 2013] qui se base sur un échantillonnage aléatoire et répété des possibles entrées du modèle et sur l'évaluation du modèle de système pour différentes valeurs des entrées échantillonnées. Par contre, l'incertitude épistémique est propagée dans le cadre de l'analyse d'intervalle [Buckley, 2004].

Contribution

Dans cette thèse de doctorat, des défis en matière de représentation, de modélisation et de simulation de SdS sont abordés. Des applications liées i) à l'évaluation du risque d'événements externes et ii) aux infrastructures critiques (ICs) sont analysées. En particulier, la première application concerne la sûreté et la résilience physique d'une installation critique (comme une centrale nucléaire) exposée au risque d'événements externes d'origine naturelle (comme les séismes et leurs répliques). La probabilité que l'installation critique reste ou non dans un « état de sûreté » (c'est-à-dire dans une condition qui ne provoque pas des dommages à la santé ou à l'environnement) après la survenance d'un événement extérieur est considérée comme un indicateur quantitatif de la sûreté. Le temps correspondant pour récupérer l'état de sûreté est pris en compte dans l'évaluation de la résilience physique. Les limites de l'analyse sont étendues aux infrastructures interdépendantes, c'est-à-dire aux réseaux de distribution d'énergie et d'eau, ainsi qu'aux réseaux de transport, qui sont liés à l'installation et qui peuvent fournir les services nécessaires pour maintenir ou rétablir la sûreté de l'installation critique, au cas où les équipements de secours internes se tombent en panne. La seconde application analyse la robustesse et la capacité de récupération des ICs interdépendantes (comme les réseaux de gaz et d'électricité et le système de contrôle et d'acquisition de données) dans un cadre de systèmes des systèmes. La robustesse est mesurée comme la probabilité que la fourniture de produit (gaz et électricité) soit livrée aux utilisateurs finaux à l'état stationnaire. La capacité de récupération est mesurée au niveau du temps nécessaire

pour récupérer les SdS à partir du pire scénario, dans lequel le SdS est en condition très dégradé et les utilisateurs finaux ne peuvent pas recevoir la fourniture de produit demandée, jusqu'à le meilleur scénario, dans lequel tous les utilisateurs finaux sont fournis du produit dont ils ont besoin.

Ensuite, les contributions de l'œuvre sur la représentation, la modélisation et la simulation sont synthétiquement indiquées.

Représentation

Différents types de cadres de représentation du système ont été étudiés et comparés qualitativement. Certains d'entre eux sont entièrement extraits de la littérature, comme l'arbre de défaillances [Zio, 2007] et le *Muir Web* [Sanderson, 2009 ; La Rocca et al., 2011] ; d'autres sont étudiés et approfondis depuis l'origine pour s'adapter aux objectifs de l'analyse de système de systèmes, comme la modélisation hiérarchique et le *Goal Tree Success Tree – Dynamic Master Logic Diagram* (GTST-DMLD). Enfin, une méthode de représentation est développée *ex novo*, à savoir, le graphe hiérarchique.

De manière très synthétique, l'arbre de défaillances représente les combinaisons possibles d'événements qui permettent la réalisation d'un événement indésirable prédéfini. Une telle représentation graphique met donc en évidence les relations de cause à effet [Zio, 2007]. Le *Muir Web* a été introduit dans le contexte de l'écologie [Sanderson, 2009] et récemment appliqué aux systèmes d'infrastructure [La Rocca et al., 2011]. Il s'agit d'une technique de représentation de réseau qui permet l'analyse par la théorie des graphes et une représentation explicite de la structure de dépendance des éléments physiques sur les facteurs qui influent sur leurs fonctionnalités [Sanderson, 2009]. La modélisation hiérarchique permet de décrire le système à différents niveaux de précision, de faciliter la compréhension du système, de fournir des informations pour soutenir le processus de prise de décision et de réduire le coût de calcul de l'analyse du système [Gómez et al., 2013].

Le GTST-DMLD offre une description claire et efficace de la complexité du système à travers différents niveaux hiérarchiques des objectifs et fonctions du système (à travers le GT), des objets et des parties (à travers le ST), et il met en évidence leurs relations à travers le DMLD qui se traduit par une matrice de dépendance et portes logiques redéfinies, comme « ET » et « OU », qui revêtent une signification différente par rapport à un modèle d'état binaire [Hu

and Modarres, 1999], comme l'arbre de défaillances. Cette méthode a été approfondie soit pour l'application sur l'évaluation du risque d'événements externes soit pour l'application sur les ICs. Le graphe hiérarchique a été introduit pour les réseaux d'infrastructures interdépendantes et il organise les arcs des réseaux en niveaux hiérarchiques sur la base du nombre de demandes fournies par un arc ou un groupe d'arcs. Cela facilite l'identification des parties du réseau qui sont plus « importantes/critiques ». Il peut également prendre en charge l'évaluation de la partition du produit dans le réseau, considérant de différentes priorités des utilisateurs finaux.

Modélisation

L'état des différents composants et leurs interactions déterminent l'état de la performance du SdS (par exemple la sûreté, la robustesse, le temps de la récupération) dont l'évaluation est soutenue par les représentations introduites ci-dessus. Pour la modélisation de l'état des éléments individuels, deux types de modèles ont été analysés et comparés au niveau des composants : l'état binaire et multiple. En conséquence, l'état des performances en ce qui concerne les SdS est respectivement binaire ou multiple.

Les modèles à états binaires permettent une simulation plus rapide, mais ils peuvent produire des résultats trompeurs ; les modèles à états multiples, quant à eux, offrent une approximation plus précise de la réalité, même s'ils ont besoin de plus d'informations sur le système à modéliser (par exemple, la définition d'états, la probabilité d'entrer dans un certain état et les probabilités de transition d'un état à un autre).

En particulier, en référence à l'application sur l'évaluation des risques dus aux événements externes, le modèle à état binaire fait une distinction entre les éléments défaillants ou non défaillants au niveau des composantes, qui se traduit par un état de sûreté ou de non-sûreté de l'installation critique au niveau du système des systèmes ; par contre, le modèle à multiples états est capable de décrire deux aspects au niveau des composantes, comme le degré de dommages structurels et fonctionnels (par exemple, la performance), qui se traduit par le degré de sûreté de l'installation critique au niveau du système des systèmes.

Quant à l'application sur les ICs, le modèle multi-état se réfère aux performances fonctionnelles des éléments qui sont décrits par des processus de Markov et semi-Markov.

Simulation

La simulation Monte Carlo (SMC) a été adoptée pour évaluer la vulnérabilité et la résilience des systèmes de systèmes. En particulier, en référence à l'application sur l'évaluation des risques dus aux événements externes, la SMC a été intégrée dans le cadre de l'évaluation probabiliste du risque sismique comprenant aussi la survenance des répliques. Quant à l'application sur les ICs, la SMC a été combinée avec l'analyse d'intervalle pour traiter l'incertitude épistémique dans les probabilités de transition d'état.

Dans la Figure 1, le travail effectué dans la présente thèse de doctorat est résumé en référence à l'application sur l'évaluation du risque d'événements externes (sur la gauche) et l'application sur les ICs (sur la droite).

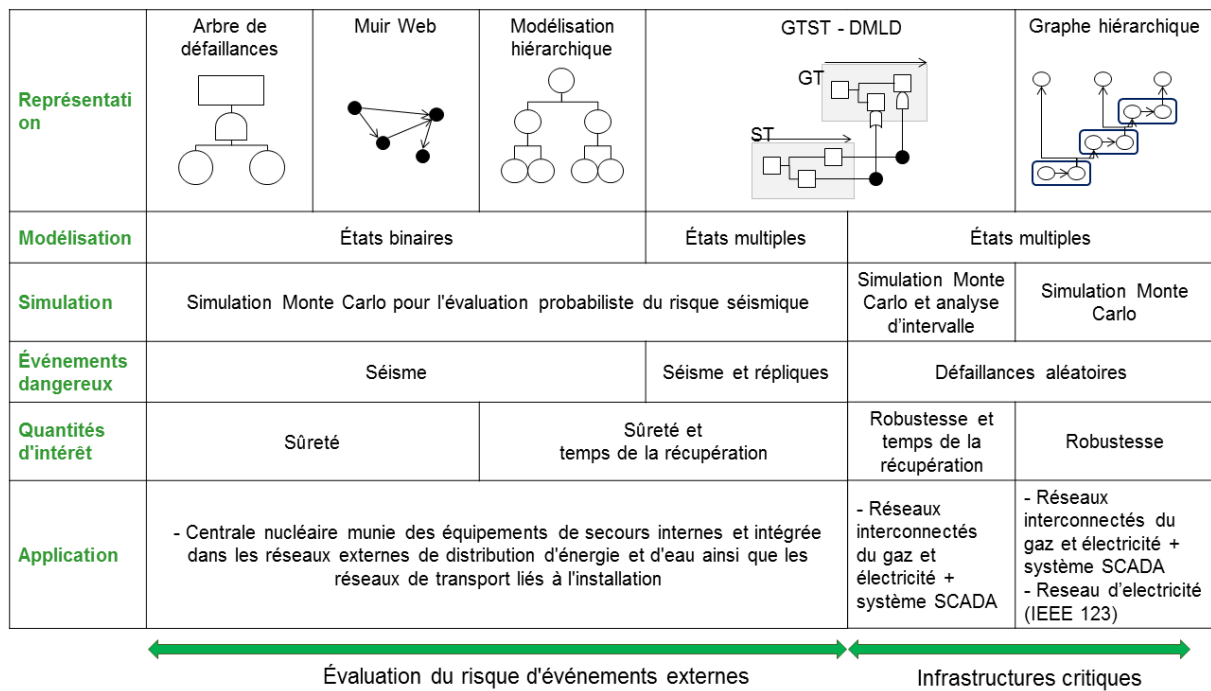


Figure 1 : Synthèse des travaux menés dans cette thèse de doctorat.

LIST OF FIGURES

Figure 1.1: Concepts of representation, modeling and simulation with respect to a real system.	22
Figure 1.2: Uncertain input variables and output variables.	23
Figure 1.3: Synthesis of the work carried out in this Ph. D. thesis.	28
Figure 1.4: Pictorial view of the flow (topic; focus, applications and outputs) of the present Ph. D. work on reliability and resilience analysis of system of systems.	30
Figure 2.1: Example of two logic gates: AND gate on the left, OR gate on the right.	34
Figure 2.2: Muir Web representation of a system of systems made of a critical plant, H (dotted-rectangular shape) whose safety is identified in the state of its critical element E, and four interdependent systems S_i , $i = 1, \dots, 4$, whose elements (represented by circles, squares, rhombs and hexagons, respectively) are connected by direct dependencies (solid lines) and support dependencies (dashed lines). The systems S_1 and S_2 are inside the critical plant, whereas the systems S_3 and S_4 are outside. The links to the critical element E (star) of the critical plant are the bold lines.	38
Figure 2.3: Top: dependencies among the components of the system of systems; the links represent the intra-systems dependencies (solid lines), the inter-systems dependencies (dashed lines) and the dependencies of the critical plant H on its interconnected systems (bold lines). Middle: graphical representation of their grouping; the rectangular, dashed, dotted and solid oval shapes represent the increasing resolution in the hierarchical level. Bottom: corresponding hierarchical representation; L: Level.	41
Figure 2.4: Corresponding fault tree representations of the state matrices reported in Table 2.4. On the left, $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in series (OR gate); in the middle, $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in parallel (AND gate); on the right, $S_5^{(3)}$ depends only on $S_{10}^{(4)}$ (INHIBIT gate without condition).	43
Figure 2.5: Computation of recovery time (RT) of the system $S_5^{(3)}$ with reference to three different configurations of the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ represented in the fault tree. On the left: OR gate, the recovery time of $S_5^{(3)}$ is the maximum recovery time of $S_{10}^{(4)}$ and $S_2^{(4)}$. In the middle: AND gate, the recovery time of $S_5^{(3)}$ is the minimum recovery time of $S_{10}^{(4)}$ and $S_2^{(4)}$. On the right, INHIBIT gate: the recovery time of $S_5^{(3)}$ is the recovery time of $S_{10}^{(4)}$ but if the condition $S_2^{(4)} = 1$ is verified, the recovery time is the sum between the recovery times of $S_{10}^{(4)}$ and $S_2^{(4)}$. 1 represents the failure state.	44
Figure 2.6: Conceptual sketch of GTST-DMLD.	45
Figure 2.7: Example of an element C that depends on two elements A and B by an “AND” gate.	46
Figure 2.8: Safety levels in a system-of-systems framework considering a critical plant in emergency conditions. The first level (top) considers internal barriers; the second one (middle) extends to the external supports; the third one (bottom) accounts for the elements supporting the recovery.	48
Figure 2.9: Scheme of GTST-DMLD for a system of systems.	49
Figure 2.10: Example of the use of the “AND” logic gate together with the dot- and square- dependencies for computing the state and the recovery time of the function F^*	51
Figure 2.11: Scheme of the hierarchies of the qualities (left) and parts (right) of a generic system of systems. The auxiliary functions and parts are connected by a dashed line to the hierarchy branch that they support. The indices $\alpha, \beta, \gamma, a, b, c$ are used to indicate the systems/elements in the hierarchies; $n^{MI}, n^{IB}, n^{ES}, n^{RS}$ refer to the number of main inputs, internal barriers, external supports and recovery supporting elements, respectively.	52
Figure 2.12: Hierarchy of the qualities for the simple example proposed.	53
Figure 2.13: Graph of the physical components (parts) for the simple example proposed.	54
Figure 2.14: Hierarchic representation of the parts of the simple example proposed: $n^{MI}, n^{IB}, n^{ES}, n^{RS}$ refer to the number of main inputs, internal barriers, external supports and recovery supporting elements, respectively.	54
Figure 2.15: GTST-DMLD with respect to the simple example of Figure 2.13.	55
Figure 2.16: Examples of dot- and hexagon-dependencies with respect to possible graph representations.	58

Figure 2.17: Examples of triangle-dependencies with respect to possible graph representations.....	58
Figure 2.18: Top: graph of the components of the system of systems; the links represent the exchange of physical product (solid lines) and influence/support relationships (dotted lined). Bottom: corresponding Hierarchical Graph; LV: Level.....	62
Figure 2.19: Hierarchical Graph of the system of systems in Figure 2.18, highlighting the path from the input to demand node S3(2); LV: Level.....	62
Figure 3.1: Relations between the structural, g_i^n , $i = 1, 2, \dots, G$, and functional z_j^n , $j = 1, 2, \dots, Z$, states for a component η	70
Figure 3.2: Three types of relations between the structural, g_i^n , $i = 1, 2, \dots, G$, and functional z_j^n , $j = 1, 2, \dots, Z$, states of a component η	72
Figure 3.3: Exemplification of the combination of $S^{(a)}$, $a = 1, \dots, 4$, systems into 3 redundant logic paths ξ_k^F , $k = 1, \dots, 3$, that attain the same function F^*	72
Figure 4.1: Example of series (left) and parallel (right) configurations between two components.....	79
Figure 4.2: Hierarchical Graph of a generic example taken as reference to illustrate the algorithm; LV: Level.....	93
Figure 4.3: Exemplification of step 2 of the algorithm with respect to the example proposed in Figure 4.2.....	95
Figure 4.4: Exemplification of step 4 of the algorithm with respect to the example proposed in Figure 4.2.....	96
Figure 4.5: Exemplification of the clustering procedure.....	98
Figure 4.6: Left: sketch of the decomposition of a system in five hierarchical levels (L) where the last one (L = 5) coincides with the actual nodes of the system; right: Hierarchical Graph of the corresponding hierarchical level 3; LV: Level of the Hierarchical Graph.....	99
Figure 4.7: Top: artificial system composed by two clusters C_1 and C_2 supplied by one input node $S_1^{(1)}$. Bottom: illustration of the real nodes inside the fictitious clusters: two arcs of C_1 are needed to supply C_2	100
Figure 5.1: Physical representation of the system. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, RP: River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator, R: Road access.	102
Figure 5.2: Fault tree of the system of systems of interest: upper levels. The elements in the triangular shape are not detailed. NPP: Nuclear Power Plant.....	105
Figure 5.3: Muir Web of the system of systems of interest: the elements in the dashed box are not considered in the present study.	106
Figure 5.4: Representation of the physical components of the Muir Web of Figure 5.3, highlighting the different types of dependencies. The interconnected systems S_i , $i=1,\dots,5$, can provide services relevant to the safe state of the nuclear power plant (NPP). The links represent the direct dependencies (solid lines), the support dependencies (dashed lines) and the dependencies of the nuclear reactor (star) on its interconnected systems (bold lines). GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, S_1 : internal power system, S_2 : internal water system, S_3 : external power system, S_4 : external water system, S_5 : Road transportation. The name of the components may differ with respect to the physical representation of Figure 5.1 since this representation is referred to the case study A.....	106
Figure 5.5: Hierarchical representation of the system of systems. NPP: Nuclear Power Plant, EE: External Energy system, EW: External Water system, IE: Internal Energy system, IW: Internal Water system, GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, L: Level. The name of the components may differ with respect to the physical representation of Figure 5.1 since this representation is referred to the case study A.....	107
Figure 5.6: GTST-DMLD of the system of systems of Figure 5.1. MFW: Main Feedwater System; HPCI: High Pressure Coolant Injection System; LPCI: Low Pressure Coolant Injection System; IE: Internal Energy System; DS: Depressurization System; EW: External Water System; EE: Offsite power system; R: Road access; GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, Cond: Condenser; RP: River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator.....	110

Figure 5.7: Left: estimate of the probability that the nuclear power plant reaches a risk (1), marginal (2) and healthy (3) state upon occurrence of an earthquake of moment magnitude equal to 5.5, in the case of multi-state (grey) and binary state (black) models. Right: same as Figure on the left, but considering also occurrence of subsequent aftershocks, in the case of multi-state (grey) and binary state (black) models.	112
Figure 5.8: Left: probability density function (PDF) of the time (RT) necessary to restore the marginal state (2) of the nuclear power plant (NPP) from a risk state (1). Right: comparison of the probability density function (PDF) of the time (RT) necessary to restore the healthy state (3) of the nuclear power plant (NPP) from a risk state (1), in the case of a multi-state (solid line) and binary state (dashed line) model.	112
Figure 6.1: Interdependent gas (solid lines) and electric (dashed lines) infrastructures and SCADA system (dotted lines) [Nozick et al., 2005]. The possible states of the arcs are given in square brackets; the quantities demanded by the end-nodes D1, D2, L1, L2 are reported in bold.	114
Figure 6.2: GTST-DMLD of the system of systems of Figure 6.1 (case study A).	115
Figure 6.3: Hierarchical Graph of the system of systems depicted in Figure 6.1 (case study A); LV: Level.	117
Figure 6.4: IEEE 123 node test feeders adapted to the purposes of the present analysis	119

LIST OF TABLES

Table 1.1: Main approaches for the vulnerability assessment of CIs with respect to the logical, phenomenological/functional, structural/topological and flow perspectives.	13
Table 1.2: Structure of the thesis.	30
Table 2.1: Advantages and limitations of Fault Tree Analysis.	35
Table 2.2: Advantages and limitations of Muir Web.	38
Table 2.3: Advantages and limitations of Hierarchical Modeling.	42
Table 2.4: Three possible state matrices for the system $S_5^{(3)}$ of Figure 2.3 (middle) on the basis of the states of the systems $S_{10}^{(4)}$ and $S_2^{(4)}$. On the left: $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in series; in the middle: $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in parallel; on the right: $S_5^{(3)}$ depends only on $S_{10}^{(4)}$; 1 represents the failure state.	43
Table 2.5: Advantages and limitations of Goal Tree Success Tree – Dynamic Master Logic Diagram.	46
Table 2.6: Advantages and limitations of Hierarchical Graph.	63
Table 2.7: Comparison of the Fault Tree Analysis (FT), Muir Web (MW), Hierarchical Modeling (HM), Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD), Hierarchical Graph (HG).	64
Table 6.1: Steady-state probabilities of (i) delivering the (optimal) required product to the demand nodes (top) and (ii) delivering a quantity of product exceeding the 90% of the corresponding demands (bottom).	118

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	iii
RÉSUMÉ EN FRANÇAIS.....	v
RÉSUMÉ ÉTENDU EN FRANÇAIS.....	vii
LIST OF FIGURES	xvii
LIST OF TABLES	xix
TABLE OF CONTENTS	xxi
ACRONYMS	xxv
1. INTRODUCTION	1
1.1. Systems of systems	2
1.2. Critical infrastructures as systems of systems	4
1.3. Interdependencies	7
1.4. Vulnerability and resilience.....	11
1.5. Representation, modeling and simulation	20
1.6. Uncertainty	22
1.7. Synthesis of the contribution of the thesis.....	24
1.8. Structure of the thesis	28
2. SYSTEM REPRESENTATION	31
2.1. Overview on the existing system representation techniques	31
2.2. Fault Tree.....	33
2.3. Muir Web.....	36
2.4. Hierarchical Modeling	39
2.5. Goal Tree Success Tree – Dynamic Master Logic Diagram	44
2.5.1. GTST-DMLD for the safety and physical resilience of a critical plant	47
2.5.2. GTST-DMLD for evaluating the robustness and recovery capacity of interdependent critical infrastructures.....	55
2.6. Hierarchical Graph	59
2.7. Comparisons of the representation techniques adopted	64
3. SYSTEM MODELING	67
3.1. Overview on the existing modeling approaches.....	67
3.2. Binary state models	68

3.3.	Multi-state models	68
3.3.1.	Structural damage and functionality.....	69
3.3.2.	Markov and semi-Markov processes.....	73
4.	SYSTEM SIMULATION AND UNCERTAINTY PROPAGATION	75
4.1.	Overview on the existing simulation and uncertainty propagation techniques	75
4.2.	Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a binary system-of-systems framework.....	76
4.2.1.	Operative simulation steps considering Fault Tree or Muir Web system-of-systems representations.....	77
4.2.2.	Operative simulation steps considering Hierarchical Modeling system-of-systems representation	79
4.3.	Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a multi-state system-of-systems framework.....	81
4.4.	Monte Carlo simulation and interval analysis within a multi-state system-of-systems framework.....	85
4.4.1.	Operative simulation steps considering GTST-DMLD system-of-systems representation: evaluation of robustness	86
4.4.2.	Operative simulation steps considering GTST-DMLD system-of-systems representation: evaluation of recovery capacity	88
4.5.	Monte Carlo simulation and Hierarchical Graph within a multi-state system-of-systems framework.....	91
4.5.1.	Operative simulation steps combining Monte Carlo method and Hierarchical Graph system-of-systems representations for robustness evaluation.....	92
4.5.2.	Combination of the Hierarchical Graph representation and a clustering algorithm for managing large-sized critical infrastructures.....	97
5.	APPLICATION 1: EXTERNAL EVENT RISK ASSESSMENT	101
5.1.	Case studies A and B: description	102
5.2.	Case study A: system-of-systems representations and main results.....	104
5.2.1.	Fault Tree representation.....	104
5.2.2.	Muir Web representation.....	105
5.2.3.	Hierarchical Modeling representation	107
5.3.	Case study B: system-of-systems representations and main results.....	108
6.	APPLICATION 2: CRITICAL INFRASTRUCTURES	113
6.1.	Case study A: interconnected gas and electricity networks	113
6.1.1.	GTST-DMLD representation and main results of case study A.....	115
6.1.2.	Hierarchical Graph representation and main results of case study A.....	116
6.2.	Case study B: electric power distribution network.....	118
7.	CONCLUSIONS.....	121
	REFERENCES.....	127

Part II

Paper I

Ferrario E., Zio E. (2012). “A system-of-systems framework of nuclear power plant Probabilistic Seismic Hazard Analysis by Fault Tree analysis and Monte Carlo simulation.” *Proceedings of the joint 2012 International Conference on Probabilistic Safety Assessment and Management (PSAM 11) & European Safety and RELiability Conference (ESREL 2012)*, Helsinki, Finland, 25-29 June.

Paper II

Zio E., Ferrario E. (2013). “A framework for the system-of-systems analysis of the risk for a safety-critical plant exposed to external events.” *Reliability Engineering & System Safety*. V. 114, pp. 114-125.

Paper III

Ferrario E., Zio E. (2014). “Assessing nuclear power plant safety and recovery from earthquakes using a system-of-systems approach.” *Reliability Engineering & System Safety*. V. 125, pp. 103-116.

Paper IV

Ferrario E., Zio E. (2014). “Goal Tree Success Tree–Dynamic Master Logic Diagram and Monte Carlo simulation for the safety and resilience assessment of a multistate system of systems.” *Engineering Structures*. V. 59, pp. 411-433.

Paper V

Ferrario E., Pedroni N., Zio E. “Analysis of the robustness and recovery of critical infrastructures within a multi-state system-of-systems framework, in presence of epistemic uncertainty.” Submitted to *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*.

Paper VI

Ferrario E., Pedroni N., Zio E. “Hierarchical Graph representations for the evaluation of the robustness of critical infrastructures within a multi-state system-of-systems framework.” Under submission.

ACRONYMS

CDF	Cumulative Distribution Function
CIs	Critical Infrastructures
EPCIP	European Program for Critical Infrastructure Protection
FT	Fault Tree
GEOSS	Global Earth Observation Systems of Systems
GTST-DMLD	Goal Tree Success Tree – Dynamic Master Logic Diagram
LGI	Laboratoire de Génie Industriel (Industrial Engineering Laboratory)
MC	Monte Carlo
PCCIP	President’s Commission of Critical Infrastructure Protection
PDF	Probability Density Function
PRA	Probabilistic Risk Assessment
SoS	System of Systems
SCADA	Supervisory Control and Data Acquisition
SPRA	Seismic Probabilistic Risk Assessment

1. INTRODUCTION

Social, political, economic and cultural changes together with the development of technology have led the world to become more and more complex and interconnected, and in continuous evolution. The advancements in the telecommunication and technology field, for example, have determined a dependence on its operation of vital systems, as the electric power grid, producing visible important improvement in their performances. However, the increasing of complex relationships among different systems creates new vulnerabilities: a failure in one system can propagate and cause a failure in a connected system leading to cascading effects that can strike areas also very far from the impact zone [Nozick et al., 2005; Bouchon, 2006].

It is an emerging belief that a holistic approach should be taken into account in order to understand and reduce these vulnerabilities, adopting a system-of-systems (SoS) view [Bouchon, 2006; Katina and Keating, 2014].

When addressing the vulnerability and risk analysis of SoS, one faces old problems which develop into new challenges with respect to the representation and modeling of the SoS, the quantification of the SoS models, the proper representation and quantification of the uncertainty in SoS behavior and modeling, and the propagation of the uncertainty to the SoS response.

The focus of the present Ph. D. thesis is on the representation, modeling and simulation of SoS with respect to their vulnerability and physical resilience to random failure and natural hazards. The application of this research regards critical infrastructures (CIs). The work has been performed at the Laboratoire Génie Industriel (LGI, Industrial Engineering Laboratory) at École Centrale Paris in the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy – Electricité de France, at École Centrale Paris – Supélec, France.

This Chapter is organized as follows. In Section 1.1, the SoS definitions, characteristics and applications are introduced; in Section 1.2, CIs are described with respect to a SoS view; in

Section 1.3, the concept of interdependence is discussed; in Section 1.4, the definitions of vulnerability and resilience are reported, and existing approaches for their evaluation are summarized; in Section 1.5, the issues of representation, modeling and simulation are illustrated; in Section 1.6, the problem of uncertainty in risk assessment is stated, and a distinction is made between the aleatory and epistemic components of uncertainty; in Section 1.7, the contribution of the thesis are highlighted; and in Section 1.8, the structure of the thesis is given.

1.1. Systems of systems

While the concept of “system” has more universal acceptance, the definition of “systems of systems” (SoS) depends on the application areas and their focus.

A *system* is a group of interacting elements (or subsystems) having an internal structure which links them into a unified whole. The boundary of a system is to be defined, as well as the nature of the internal structure linking its elements (physical, logical, etc.). Its essential properties are autonomy, coherence, permanence, and organization [Dupuy, 1985; Kröger and Zio, 2011]. With respect to the SoS, Boardman et al. collected around 40 definitions taken from academic literature, conference proceedings/presentations, and documentation that have been independently published by industry, government and academia [Boardman et al., 2006].

For example, Kotov defines them using the expression “complex systems” [Kotov, 1997]: “SoS are large scale concurrent and distributed systems that are comprised of complex systems”. A complex system [Guckenheimer and Ottino, 2008] is made by many components interacting in a network structure. Most often, the components are physically and functionally heterogeneous, and organized in a hierarchy of subsystems that contributes to the system function. This leads both to structural and dynamic complexity [Zio, 2014]. *Structural complexity* derives from i) heterogeneity of components across different technological domains due to increased integration among systems and ii) scale and dimensionality of connectivity through a large number of components (nodes) highly interconnected by dependences and interdependences. *Dynamic complexity* manifests through the emergence of (unexpected) system behavior in response to changes in the environmental and operational conditions of its components. In addition, uncertainty is considered “pervasive in complex

systems” and its quantification and propagation is a “key aspect in reliable prediction and control” [Guckenheimer and Ottino, 2008].

Another interpretation of the concept of SoS is given by Maier that identifies five properties (also known as “Maier’s criteria”) [Maier, 1996; Maier, 1998]: i) operational independence, i.e., each system is independent and it achieves its purposes by itself, ii) managerial independence, i.e., each system is managed in large part for its own purposes rather than the purposes of the SoS, iii) geographic distribution, i.e., SoS are distributed over a large geographic extent, iv) emergent behavior, i.e., SoS have capabilities and properties that do not reside in the component systems, and v) evolutionary development, i.e., SoS evolve with time and experience. The last three properties (iii – v) are the same as those that characterize a complex system: thus, it can be drawn that, according to Maier’s definition, the difference between a complex system and a SoS is determined by the first two properties (i and ii).

Based on Maier’s criteria, Sage and Cuppan specify that SoS exist when there is a presence of a majority of the mentioned five characteristics [Sage and Cuppan, 2001]. In this work, we focus on the following definition by DeLaurentis [2007] that encapsulates the Maier’s criteria and captures additional aspects, such as heterogeneity of component systems and multi-level structure: “*A SoS consists of multiple, heterogeneous, distributed, occasionally independently operating systems embedded in networks at multiple levels that evolve over time*” [DeLaurentis, 2007].

The SoS traits make them difficult to build and manage with traditional engineering practices. According to [Fisher, 2006], monolithic systems depend on central control, global visibility, hierarchical structures, and coordinated activities, but these characteristics cannot be expected in SoS that are related to distributed control, cooperation, influence, cascade effects, emergent behaviors [Béjar et al., 2009]. To deal with these issues, new approaches have to be identified.

Examples of application areas of SoS include [Jamshidi, 2009]: *service industry* (e.g., infrastructure systems), *electric power grids* (large-scale, complex, dynamical systems that must operate reliably to supply electrical energy to customers, and, in addition, are experiencing an increasing integration of renewable energy resources), *transportation systems* (a SoS approach is needed for a more complete model and understanding of the national

transportation system [DeLaurentis, 2009]), *healthcare systems* [Wickramasinghe et al., 2009], *national defense* (military systems were designed and developed individually, but nowadays the changes of operations and technologies call for the need of systems that work together [Dahmann, 2009]), *aeronautical field*, (e.g., e-enabling aircraft design as a SoS at Boeing Commercial Aircraft Division [Wilber, 2009]), *sensor networks* (multiple sensing devices that work cooperatively and collaboratively [Sridhar et al., 2009]), *space explorations* (that deal with extremely large, complex, and intertwined command and control, and data distribution ground networks [Jolly and Muirhead, 2009]), *communication and navigation networks for the space* (that require system interoperability, enhanced reliability, common interfaces, dynamic operations, and autonomy in system management [Bhasin and Hayden, 2009]), *sustainable environmental management* (e.g., solutions to global warming problems [Hipel et al., 2009]), *robotic swarms* [Sahin, 2009], *Global Earth Observation Systems of Systems* (GEOSS – a global project with more than 60 nations involved to improve the coordination of strategies and systems for Earth observations [Shibasaki and Pearlman, 2009]), and others.

The application area of this work focuses on CIs; in the next Section 1.2, an introduction to CIs is given within a SoS framework of analysis.

1.2. Critical infrastructures as systems of systems

The welfare of modern society relies on the continuous operation of CIs that are essential in providing goods (such as energy, water, data) and services (such as transportation, banking and health care) across local, regional and national boundaries. These infrastructures are getting more and more automated, and strongly interconnected due to their increasing extension on large scales and the progressive advances in information technology; for example, “today’s ability to run largely distributed power networks with a variety of generation technologies e.g., nuclear, thermo, hydro etc. is only possible through the intense use of information and communication systems” [Gheorghe and Schlapfer, 2006]. If, on one hand, these advances have increased their efficiency (e.g., provide better measurements, allow quicker operations, more powerful control schemes and broad access to data [Gheorghe and Schlapfer, 2006]), on the other hand, they have created new vulnerabilities to component

failures, natural and manmade events. Actually, in the last decades an increased number of disruptive events (natural external events, malicious acts, large scale blackouts) affecting CIs has occurred: for example, Ice Storm in Canada (in 1998 and 2013, probably the two worst ice storms in the recent history), World Trade Center attack (New York, 2001), explosion of the AZF factory (Toulouse, 2001), North American blackout (eastern USA and Canada, 2003), terrorist attacks targeting underground transportation (Madrid, 2004 and London, 2007), hurricane in Florida (2004), UK floods (2007), M_w 8.8 earthquake in Chile (2010), M_w 9.0 earthquake and subsequent tsunami (Japan, 2011), etc.

To conduct a comprehensive review on CIs and recommend a national policy for protecting and assuring their continued operation, the President's Commission of Critical Infrastructure Protection (PCCIP) was created in 1996 [PCCIP, 1997]. Several directives and executive orders followed (e.g., the Presidential Decision Directive 63 (1998) to protect from deliberate attacks) and the US Department of Homeland Security (2002) was established with the primary responsibilities of protecting the United States and its territories from terrorist attacks, manmade accidents, and natural disasters. Other international federations and countries followed with some delay, such as the European Union that introduced the European Program for Critical Infrastructure Protection (EPCIP) of 2004 to "assure the continued functioning of Europe's critical infrastructure" [COM, 2004]. The infrastructures defined "critical" are those "whose services are so vital that their incapacity or destruction would have debilitating impact on the defense or economic security of any state" [COM, 2004].

During the last decades the meaning of CIs have expanded: in 1997 CIs included telecommunications, electric power systems, natural gas and oil, banking and finance, transportation, water supply systems, emergency services, and continuity of government [PCCIP, 1997], i.e., those infrastructures which prolonged disruptions could impact significantly military and economic areas; in 2003, they extended to chemical and hazardous, postal and shipping industries and, in addition, to national monuments and icons, where an attack might affect the nation's morale [Moteff et al., 2003]. This continuing evolution and extension of the concept of critical infrastructure with the changing of modern society implies an evolution of the traditional concept of protection, management and controlling of infrastructures.

CIs are various in nature, e.g., physical-engineered, cybernetic or organizational systems [Kröger and Zio, 2011]. *Engineered, physically networked* CIs are considered in this thesis; examples are those providing:

- energy (electricity, oil and gas supply as subsectors);
- transportation (by rail, road, air, shipping);
- information and telecommunication (such as the internet);
- drinking water, including wastewater treatment.

CIs are considered large scale, spatially distributed, complex systems that mainly operate in the open. Actually, they are enormous networks that transcend national borders and are composed by a multitude and variety of nodes representing physical hard components (e.g., road, railway, pipelines, etc.), soft components (e.g., SCADA, information and telecommunication systems) and human and organizational components. They are highly interconnected and mutually dependent in complex ways (a detailed description of the types of interdependencies and how to identify them is provided in Section 1.3), and a failure in one infrastructure can propagate to other infrastructures provoking cascading failures that produce large consequences well beyond the impact zone. CIs are triggered by various sources of hazards due to exogenous and endogenous stressors like natural events, terrorism, criminal activities, malicious behavior, market and policy factors, human factors and technical failures components. CIs are affected by large uncertainties in the characterization of the failure behavior of their components, their interconnections and interactions [Zio and Aven, 2011]; this makes the vulnerability analysis a challenging task to quantify the uncertainty and predict how it propagates throughout the system. They present a dynamic structure, they evolve and adapt themselves responding to environmental changes to continue providing for their functionality and they show emergent behavior. Indeed, the overall behavior emerges from the interactions among single parts of a complex system: in other words, synergies emerge from the interactions among these components and the whole critical infrastructure is more than the sum of its parts.

Not just the concept of critical infrastructure itself has expanded over time but also the related concerns. At the beginning, they were associated with deterioration, technological obsolescence and insufficient capacity to serve future growth [Moteff et al., 2003].

Afterwards, the manmade events and the information warfare (due to the increasing dependency on information and technology) have shifted the focus on maintaining and sustaining public wellbeing. As a response to these concerns, a SoS perspective to look at CIs took shape [Bouchon, 2006; Tolone et al., 2009; Eusgeld et al., 2011; Kröger and Zio, 2011]. Katina and Keating introduce a SoS based perspective of CIs [Katina and Keating, 2014]. In particular, they classify interconnected CIs as SoS showing that they exhibit the Maier's characteristics (see Section 1.1). Actually, CIs present: i) operational independence (even if they do not operate in isolation since they are interdependent) due to their capacity of achieving their goals mainly independently; ii) managerial independence, since they are owned and operated independently; iii) geographical distribution, since they are often geographically distributed, iv) emergent behavior since the final goal of maintaining and sustaining public wellbeing is achieved by multiple CIs and not evident in any individual infrastructure, and v) evolutionary development, since CIs evolve over time in response to environmental changes [Katina and Keating, 2014].

The SoS view allows embracing a holistic approach where multiple complex systems are integrated and their interdependencies evaluated [Katina and Keating, 2014]. In this view, new challenges emerge with respect to the analyses of interdependencies (Section 1.3), and the representation, modeling and simulation of the SoS (Section 1.5) in order to evaluate its safety, vulnerability, robustness and resilience (Section 1.4).

1.3. Interdependencies

Infrastructure systems interact on the basis of relationships that cross the single infrastructure boundary, giving rise to a large-scale complex network system. Identifying, understanding and analyzing these complex interactions represent a challenge to the evaluation of the real vulnerability of each infrastructure system in consequence of an initiating event [Rinaldi et al., 2001; Pederson et al., 2006].

Parshani et al. distinguish between dependency and connectivity links to study the robustness of real networks: dependency links propagate the failure of a node to its connected neighbors, whereas connectivity links disconnect the nodes from the network [Parshani et al., 2011]. The authors show that networks with high density of dependency links are extremely vulnerable to

random failure; on the contrary, networks with low density of dependency links are more robust.

Rinaldi et al. distinguish between dependency and interdependency [Rinaldi et al., 2001]: dependency is a unidirectional relationship between two infrastructures, i.e., infrastructure i depends on j through a link, but j does not depend on i through the same link, while interdependency is a bidirectional relationship, i.e., infrastructure i depends on j through some links, and j likewise depends on i through the same and/or other links.

Rinaldi et al. identify six dimensions for describing infrastructure interdependencies [Rinaldi et al., 2001]: i) *environment* that includes business, economic, public policy, health and safety, security, economic, legal/regulatory, technical and social/political aspects, ii) *state of operation* that considers normal, repair/restoration and stressed/disrupted states, iii) *coupling and response behavior* with respect to the degree of connections and the corresponding response to changes (adaptive or inflexible), iv) *infrastructure characteristics* (spatial, temporal, operational and organizational), v) *types of failure* (common cause, cascading, escalating), and vi) *types of interdependencies* (physical, cyber, logical and geographical).

In particular, *physical interdependencies* exist when the state of each infrastructure depends on the material output of the other infrastructure (e.g., a rail network depends on the coal-fired electrical generation plant for the supply of electricity and, vice versa, the electrical generator requires the rail network for the delivery of the coal for fuel and other activities like repair/replacement parts to the plant); *cyber interdependencies* connect infrastructures to one another by informational links (e.g., a supervisory control and data acquisition (SCADA) system monitors and controls components of the electric power grid); *geographical interdependencies* consider elements that are in spatial proximity and so can potentially be affected by the same local environment (e.g., earthquake occurrence); *logical interdependencies* include all the connections that are not physical, cyber and geographical [Rinaldi et al., 2001].

With respect to the taxonomy of the types of interdependencies extensions or modifications have been proposed. The logical interdependencies have been expanded to policy/procedural and societal interdependencies by Pederson et al. [Pederson et al., 2006]. Kröger and Zio change “cyber” interdependencies with “informational” interdependencies to include hard- and software, and they extend the term “geographical” interdependencies with “geospatial” interdependencies [Kröger and Zio, 2011].

Other types of interdependencies have been identified in the academic literature. Zimmerman distinguishes between *spatial* and *functional interdependencies*: the first one refers to proximity between infrastructures, whereas the second one identifies an infrastructure that is necessary for the operation of another infrastructure, (e.g., water pumps need electricity to work) [Zimmerman, 2001]. Wallace et al. distinguish between input, shared, exclusive-or, co-located and mutual interdependencies [Wallace et al., 2003]. *Input interdependency* illustrates an infrastructure that needs one or more inputs that are provided by another infrastructure; *shared interdependency* highlights components and/or activities of an infrastructure that are shared with one or more infrastructures; *exclusive-or interdependency* exists when only one of two or more infrastructures can provide service; *co-located interdependency* occurs when components of two or more infrastructure are situated within a prescribed geographical region; *mutual interdependency* is with respect to a collection of infrastructure and it exists if at least one of the activities of any infrastructure among the set of infrastructures is dependent upon each of the other infrastructure in the set. Zhang and Peeta extend the concept to the business scenarios and consider functional, physical, budgetary, and market and economic interdependencies [Zhang and Peeta, 2011]. *Functional interdependency* means that the operation of one system needs inputs from another system, or can be replaced, to a certain extent, by other systems; *physical interdependency* occurs when some infrastructures are coupled through shared physical attributes; *budgetary interdependencies* are due to the dependence of many infrastructures to some level of public financing (especially under a centrally-controlled economy or during disaster recovery); *market and economic interdependencies* consider for example that infrastructure systems interact with each other in the same economic system, or serve the same end-users who determine the final demand for each commodity [Zhang and Peeta, 2011].

The analysis of interdependencies, their identification and quantification, is necessary in order to assure proper infrastructure protection and resilience [Ruzzante et al., 2010].

The identification task is hard due to the so called “hidden interdependencies”, i.e., those connections that are hidden during normal operation and become evident and critical during emergency. For this reason, it is difficult to identify them before the occurrence of a disruptive event and, usually, they are derived a posteriori from its consequences. Pederson et

al. consider a dependency matrix whose rows and columns are represented by the CIs and the cells are filled by the terms “high”, “medium” and “low” to express the degree of the dependencies (if they exist) [Pederson et al., 2006]. However, this approach is suitable for small scale networks of systems since it requires too many resources for large scale systems [Pederson et al., 2006]. Another approach to discover interdependencies comes from “data mining”, whose fundament is to extract information from huge data sets and that only recently has been applied to analyze interdependency of CIs [Oliveira et al., 2009; Shih et al., 2009; Chou and Tseng, 2010].

With respect to the measure of interdependencies, Ruzzante et al. suggest a data-driven metric for the analysis and estimation of interdependencies in dynamic system networks defining the strength of the dependences [Ruzzante et al., 2010]. They identify five properties for a dependency relation that are direction, position (i.e., geographical setting with respect to a reference point), delay (i.e., reaction time interval), order (i.e., number of links connecting two nodes, e.g., source and target), and strength; this last one is identified from the observations of input-output relations among nodes and without considering the internal node dynamics.

The Methodology for Interdependencies Assessment European Union project [D'Agostino et al., 2010; Fioriti et al., 2010; Casalicchio et al., 2011] has investigated methods to assess measures of interdependencies between the information and telecommunication technology and electric power systems. Since a unique metric can lead to incorrect considerations, the Methodology for Interdependencies Assessment European Union project proposes to define a metric for each type of interdependency classified on the basis of its nature (physical, geographical and cyber) and analyzed along its temporal scale [D'Agostino et al., 2010; Casalicchio et al., 2011]. In particular, the main metrics identified are [Casalicchio et al., 2011]: i) topological robustness to measure the intensity of cyber and physical dependencies; ii) module and phase of the frequency response (modeled by transfer functions) and poles placement (that evaluates the stability margins of the systems) to measure the cyber and physical dependencies from a dynamic viewpoint; iii) ratio of interdependency (ratio of the number of malfunctioning in a set of infrastructures and the number of malfunctions occurring in another infrastructure) to measure the cyber and physical dependencies from a

static and dynamic viewpoints, and iv) temporal scale of interdependency to measure the time scale and, thus, the dynamics of cyber and physical dependencies. The module of the transfer function can represent a good metric when an exhaustive knowledge of the system is available and the possible disturbances are small; otherwise, a topological metric such as degree, clustering, and centrality indices, or a “spectral metric”, based on the maximum eigenvalue of the adjacency matrix of the whole network, can be suitable alternatives [D'Agostino et al., 2010; Fioriti et al., 2010].

1.4. Vulnerability and resilience

As mentioned earlier, the SoS which make up CIs raise concerns with respect to their risk, vulnerability and resilience. CIs are getting more and more interdependent manifesting emergent behavior that cannot be predicted from the behavior of individual elements. In addition, large uncertainties exist that make their predictions difficult to achieve reliably [Zio and Aven, 2011]. In this context, reductionist methods for vulnerability and risk analysis may fail to capture heterogeneity, structural and dynamic complexity. New approaches are needed [Kröger and Zio, 2011].

While the concept of risk is fairly mature and consensually agreed, the concepts of vulnerability are still evolving and not yet established [Kröger and Zio, 2011]. Risk refers to the probability of occurrence (frequency) of a specific (mostly undesired/adverse) event leading to loss damage or injury, and its extent. These quantities and their associated uncertainties are considered as being numerically quantifiable [Kröger and Zio, 2011]: e.g, for CIs, risk can be computed as the loss of service with its resulting consequences for the people concerned.

Vulnerability has been introduced as the hazard-centric perception of disasters that is revealed as being too limited to understand in terms of risks. It is claimed that the estimation of the failure probabilities adopted in risk analysis to inform risk management decisions, may be poor since they are affected by insufficient knowledge and inappropriate assumptions and, moreover, unexpected events can occur, like unknown failure mechanisms [Johansson and Hassel, 2010]. The level of vulnerability makes the difference between a hazard of low

intensity that could have severe consequences and a hazard of high intensity that could have negligible consequences [White, 1974].

There are two main interpretations of vulnerability: one is related to a global system property whereas the other one qualifies directly system components. In the first interpretation (closer to the definition of risk), the goal is the evaluation of the extent of adverse effects caused by the occurrence of a specific hazardous event [Johansson and Hassel, 2010; Kröger and Zio, 2011]; in the second one, the objective is to identify critical components, i.e., those components whose failure causes large negative effects to that system.

The concept of vulnerability as global system property embeds three other concepts [Kröger and Zio, 2011]: i) degree of loss and damages due to the impact of a hazard, ii) degree of exposure to the hazard (defined as the likelihood of being exposed to hazards and as the susceptibility of an element at risk to suffer losses and damages), and iii) degree of resilience. In this view, resilience can be seen as an aspect of vulnerability. Actually, vulnerability and resilience are two sides of the same coin, where the first one focuses more on system protection and the second one on system recovery [Haimes, 2009b].

A broad spectrum of approaches exists for the vulnerability assessment of CIs; however, the analysis of these SoS cannot be carried out only with classical methods of system decomposition and logic analysis; a framework is needed to integrate a number of methods capable of viewing the problem from different perspectives under the existing uncertainties [Zio, 2014]. The main perspectives include [Zio, 2014]:

- *Logical methods* based on system analysis, hierarchical and logic trees, etc.; these methods are capable of capturing the logic of the functioning/dysfunctioning of a complex system, and of identifying the combinations of failures of elements (hardware, software and human) which lead to the loss of the system function.
- *Phenomenological/Functional methods*, based on transfer functions, state dynamic modeling, input-output modeling and control theory, agent-based modeling etc.; these methods are capable of capturing the dynamics of interrelated operation between elements (hardware, software and human) of a complex system and with the environment, from which the dynamic operation of the system itself emerges.

- *Structural/topological methods* based on system analysis, graph theory, statistical physics, etc.; these methods are capable of describing the connectivity of a complex system and analyzing its effects on the system functionality, on the cascade propagation of a failure and on its recovery (resilience), as well as identifying the elements of the system which must be most robustly controlled because of their central role in the system.
- *Flow methods*, based on detailed, mechanistic models (and computer codes) of the processes occurring in the system; these methods are capable of describing the physics of system operation, its monitoring and control.

Table 1.1 illustrates the main approaches for the vulnerability assessment of CIs with respect to the perspectives above introduced. In the following, their principal features are briefly described.

Table 1.1: Main approaches for the vulnerability assessment of CIs with respect to the logical, phenomenological/functional, structural/topological and flow perspectives.

LOGICAL methods
- Risk analysis <i>Fault and event trees and core methods of probabilistic risk assessment</i>
- Probabilistic modeling <i>Markov Chains</i> <i>Markov/Petri nets</i> <i>Bayesian network</i>
PHENOMENOLOGICAL/FUNCTIONAL methods
- Agent based modeling and simulation - System dynamic model - Economic-based approaches <i>Input-output model</i> <i>Computable general equilibrium</i>
- Others <i>Dynamic control system theory</i> <i>High level architecture</i>
STRUCTURAL/TOPOLOGICAL methods
- Network based approaches <i>Topology-based</i>
FLOW methods
- Network based approaches <i>Flow-based</i>

Logical methods

This perspective includes i) risk analysis approaches that evaluate the result of adverse events affecting a system by means of the potential negative consequences and their associated likelihoods, and it provides suggestions on how to reduce vulnerability, improve resilience and mitigate consequences and ii) probabilistic modeling adopted for the characterization of CIs.

Risk analysis is carried out by qualitative [Moore, 2006; Piwowar et al., 2009] and quantitative assessments [Apostolakis and Lemon, 2005; Flammini et al., 2009] with the further goal of ranking system components on the basis of their criticality [Koonce et al., 2008]. Traditional methods for risk analysis, e.g., fault and event tree methodology and core methods of probabilistic risk assessment (PRA), have been applied to the vulnerability analysis of CIs for protecting the systems against malevolent actions [Piwowar et al., 2009]. The approach comprises a step-by-step process typical of PRA [Kröger and Zio, 2011]. However, it implies drawbacks for use on safety-related issues of large-scale infrastructures due to “i) the high complexity and interconnectedness of modern SoS that cannot be adequately modeled; ii) all kinds of human factors and the full spectrum of threats, including malicious behavior and attacks that cannot be taken into account; iii) the dynamic or even the non-linear behavior of systems that cannot be easily handled; and iv) independence from contextual factors that has to be assumed” [Kröger, 2008].

Probabilistic modeling approaches include Markov Chains, Markov/Petri nets and Bayesian networks. The first two rely on the definition of transition probabilities of the system components among their reachable states. An achieved configuration of the component states determines the system state. A limitation of these methods is the exponential growth of the possible configuration of the system when the number of components increases and/or the number of states for each component is high. Bayesian networks can be used for modeling and predicting the behavior of a system, based on observed stochastic events. Drawbacks of this methodology arise from its complexity that leads to significant efforts in logic modeling and quantification, and from the limited capability of providing an exhaustive analysis.

Phenomenological/Functional methods

This category of methods includes i) agent based model, ii) system dynamic model, iii) economic-based approaches, iv) others (e.g., dynamic control system theory and high level

architecture).

Agent based modeling is a simulation methodology coming from the field of complexity science. It is used to evaluate the dynamic operational behavior of infrastructure network and its associated economic entity. An agent based model is composed by three elements: i) agents, i.e., technical and non-technical components, ii), environment, i.e., abstract space where the agents can interact, and iii) rules, i.e., behavior patterns for the agent and the environment, they can include physical law. The behavior of the infrastructure emerges from the behaviors of the individual agent and their interactions [Kröger and Zio, 2011]. The main advantages of the agent based modeling are the possibility of representing heterogeneous components and capturing all types of interdependencies among CIs, capturing the emerging behavior, create a space where the agents interact according to distance, provide a scenario-based what-if analysis and the effectiveness assessment of different control strategies, and can be also integrated with other modeling technique to provide more comprehensive analysis [Borshchev and Filippov, 2004; Ouyang, 2014]. However, two main limitations are with respect to i) the challenge of calibrating the simulation parameters due to the lack of significant data and the difficulties in model the agent behavior, and ii) the dependence of the quality of the simulation on the assumptions made that are difficult to justify theoretically and statistically [Ouyang, 2014].

System dynamic models take a top-down analysis for interdependent CIs to characterize their functions such as production, transmission and consumption. It uses a series of differential equations to describe the system level behaviors of the CIs. The key concepts are i) feedback loops to indicate connection and direction of effects between CIs components, ii) stocks to represent the states of the system and iii) flow rates between stocks.

Several are the advantages of this approach: for example, it allows capturing important causes and effects under disruptive scenarios, providing investment recommendations, and including multi-attribute utility functions to compare protection strategies. On the contrary, they cannot analyze component-level dynamics (such as change of infrastructure topologies), it is difficult to calibrate parameters (huge amount of data are needed) and perform a validation of the model [Ouyang, 2014].

Economic-based approaches include two types of economic theories employed to model CIs interdependencies: input-output model and computable general equilibrium.

The first one is based on a static and linear model whose output is interpreted as the risk of inoperability of a CI, i.e., its inability to perform its function. It is based on the large-scale databases and measures the interdependencies among infrastructure sectors by economic relationships. The input-output model allows analyzing the propagation of perturbations between interdependent infrastructures and, thus, implementing effective mitigation strategies; in addition, it can provide analytical solutions that facilitate the sensitivity analysis of parameters [Ouyang, 2014]. However, it cannot analyze the interdependencies at component levels and it can give a good approximate result only when the disturbances have small impact on the economic sectors (since the interdependent matrix is derived from economic database and its elements measure the interdependent strength in normal economic operations), otherwise it will provide large errors [Ouyang, 2014].

The computable general equilibrium is an extension of the input-output model to capture non-linear connections among CIs.

Other approaches exist like dynamic control system theory [Casalicchio et al., 2011] and high level architecture that integrate all the other methods (e.g., agent based modeling) [Eusgeld et al., 2011; Wang et al., 2011].

Structural/topological methods

This class of approaches models the interdependent CIs on the basis of their topologies under different types of hazards [Ouyang, 2014]. They represent CIs by networks (actually, they are network-based approaches) where nodes are the components and links are the physical and relational connections among them. These topology-based methods consider two possible states for the components, failed and functioning, and can measure the strength of the connections by including weighted links. Topological analyses focus on the static structural property of the network by: i) highlighting the role played by its components, ii) making preliminary vulnerability assessments based on the simulation of faults (mainly represented by the removal of nodes and arcs) and the subsequent re-evaluation of the network topological properties [Kröger and Zio, 2011]. Topological-based studies can be analyzed by analytical methods and simulation methods depending on the level of detail the CIs topologies are modeled. The analysis can be performed analytically if the node heterogeneity is not taken into account so that each critical infrastructure can be described by its node degree distribution represented by a generating function [Ouyang, 2014]. Application in this research

area regards, for example, analyses of the spread of epidemic disease on networks [Newman, 2002], the breakdown of network under intentional and random attack [Albert et al., 2000], cascading failure in interdependent network [Buldyrev et al., 2010]. A simulation method is preferred when node heterogeneity is considered. Many performance metrics can be quantified, e.g., number of normal or failed components, connectivity loss, fraction of costumers affected, lost service hour, and they can be used to evaluate interdependent effects (as the absolute differences between the independent and interdependent responses normalized by the maximum interdependent response [Dueñas-Ororio et al., 2007b]) to facilitate the assessment of mitigation actions and cascading failure effects [Ouyang, 2014]. The topology-based methods are suitable to evidence relevant structural properties of interconnected networks by identifying critical components and supporting the improvement of system robustness, but they cannot capture the dynamic complexity of real CIs, i.e., they cannot give sufficient information about their flow performance, that instead is analyzed by the flow-based methods.

Flow methods

This class of approaches models the interdependent CIs on the basis of their flow patterns under different types of hazards [Ouyang, 2014]. As the structural/topological methods, they represent CIs by networks (they are network-based approaches too). These methods can be based on uniform network descriptions [Lee et al., 2007], physical rules that provide a more realistic modeling on interdependencies [Ouyang et al., 2009], oriented stochastic modeling methods [Bobbio et al., 2010], dynamic functional model [Trucco et al., 2012], maximum flow model [Nozick et al., 2005], and others [Ouyang, 2014]. They capture the flow characteristics of interdependent CIs, identify critical component, and suggest improvement for the emergency protection; however, their computational cost can be prohibitive when the components and links are described in detail [Ouyang, 2014].

Finally, another approach not included in the above perspectives is the statistical analysis. Statistical analysis is suitable when rich data sets about the system operation and performance are available. It has been used as a tools for decision support in the diagnosis and rehabilitation of CIs, e.g., water supply system [Yamijala et al., 2009], and for the identification of the system most critical parameters. However, ranking of the system

components is not possible by resorting to statistical technique only and no identification of the topological structure is accounted for. Several drawbacks of this analysis are due to: i) data that are collected in specific past operating conditions and may not reflect the same conditions at present and in the future; ii) relationships between the measures of the operating conditions and system performance that may be complicated and poorly understood; iii) very big data sets that do not allow drawing clear insights; iv) data presented in aggregated form that hide the structure of the critical infrastructure under analysis; v) improper choice of the most suitable model for the specific critical infrastructure which best fits the physics of the provided service [Kröger and Zio, 2011].

Recently, due to the increasing importance of the resilience concept, a new field of research in “resilience engineering” is emerging [Zio, 2009b]. The concept of resilience varies by discipline and application [Henry and Ramirez-Marquez, 2012; Ouyang et al., 2012]. Bruneau et al. provide a general framework to define seismic resilience for communities, including all actions that reduce losses from hazards, including effects of mitigation and rapidity of recovery [Bruneau et al., 2003]. Resilience can be understood as “the ability of the system to reduce the chances of shock, to absorb a shock if it occurs and to recover quickly after a shock (re-establish normal performance)” and it is characterized by four properties, i.e., robustness, redundancy, resourcefulness and rapidity, and four interrelated dimensions i.e., technical, organizational, social, and economic [Bruneau et al., 2003]. Hollnagel et al. consider resilience engineering as the new paradigm for safety engineering [Hollnagel et al., 2006]. Manyena views resilience as the capacity of a system to survive to aggressions and shocks by changing its non-essential attributes and rebuilding itself [Manyena, 2006]. The U.S. Department of Homeland Security defines resilience as “the capacity of an asset, system, or network to maintain its function during or to recover from a terrorist attack or other incident” [U.S. Department of Homeland Security, 2009]. Haimes considers resilience as the “ability of the system to withstand a major disruption within acceptable degradation parameters and to recover within an acceptable time and composite costs and risks” [Haimes, 2009b]. According to Aven, resilience is closely related to the concept of robustness and the key difference is the initiating event: actually, robustness (and vulnerability) relate to a specific initiating event, whereas resilience relates to any initiating event [Aven, 2011]. Resilience can be interpreted as “the uncertainty about and severity of the consequences of the activity given the occurrence

of any type of initiating event [Aven, 2011]. Alessandri and Filippini consider resilience as a structural property, i.e., as the ability to resist to internal drifts and cascading failures, and recover back to the initial operation state [Alessandri and Filippini, 2013].

In this work, the concept of resilience is limited to the physical characteristics of the components and systems and we do not include organizational, social and economic aspects: then, we refer to physical resilience as the underlying concept.

Ouyang classified the approaches for the vulnerability assessment of CIs illustrated above on the basis of their supporting in the defined resilience improvement strategies that are distinguished according the three resilience capacities, i.e., resistant, absorptive and restorative [Ouyang, 2014]. He found out that only the high level architecture (that is an hybrid method and can integrate all the other approaches) can support all the resilience improvement strategies identified, followed by the agent-based and flow-based methods [Ouyang, 2014]. However, the high level architecture is not a mature approach and still many challenges to apply it in practice exist. Also for the resilience analysis as for the vulnerability analysis the necessity of a uniform framework to integrate the existing methods is emerging.

Numerous attempts have been done to determine a metric for measuring resilience [Ouyang et al., 2012]. For example, Reed et al. identify a simple methodology to quantify engineering resilience for systems of interdependent networked infrastructures exposed to extreme natural hazards, combining resilience measures of fragility and quality with an input-output model [Reed et al., 2009]; Cimellaro et al. propose a framework for quantitative definition of resilience using an analytical function that may fit both technical and organizational issues [Cimellaro et al., 2010]; Henry and Ramirez-Marquez propose generic metrics and formulae for quantifying system resilience as a function of time [Henry and Ramirez-Marquez, 2012]; Ouyang et al. introduce a three-stage resilience analysis framework to consider system resistive, absorptive and restorative capacities together, taking into account different hazard types [Ouyang et al., 2012].

In general, the evaluations of risk, vulnerability and resilience of systems are carried out, first, by a proper representation and modeling of the reality and, then, by a quantification of the

defined metrics, often through a simulation process (Section 1.5). The system representation and modeling is such that the characteristics of the system under analysis cannot all be fully captured, and uncertainty is always present and has to be represented, quantified and propagated (Section 1.6).

1.5. Representation, modeling and simulation

Representation, modeling and simulation are three fundamental steps that have to be carried out to capture the relevant essence of the system under consideration and answer specific questions. Actually, the main challenge is to understand, predict, evaluate some variables, measures, indicators that can give information about the behavior of a system that is characterized by specific inputs/parameters and modified over time by actions, events, physical phenomena.

The first step is devoted to the *system representation*, which the subsequent steps of modeling and simulation rely on. The representation step aims at capturing the main features of the real system and depends on the type of the system and the outputs of interest. Actually, different types of systems can be better described by different representation frameworks, e.g., complex network theory may be more suitable for large distributed systems [Dueñas-Osorio et al., 2007a], whereas fault and event trees can be used for industrial, safety-critical plants [Zio, 2007]. The outputs of the analysis also lead the choice of the suitable kind of representation: for example, with respect to the representation of a geographic area, there are several thematic maps that convey information about a single topic or theme, such as population density or geology, etc. [ESRI, 2014], thus, each one is more suitable to answer different questions.

In this view, the representation of a system is a picture of the information needed to answer relevant questions: it should be synthetic, efficient, easy to read, immediate to understand without the support of further text description, and without the necessity of (possibly confusing) details. In addition, the system representation may facilitate the analysis of scenarios, including those that are new and emergent.

For a quantitative evaluation of risk, vulnerability and resilience, the representation of the system of interest should be supported by a quantitative mathematical model. In general, actions, events and physical phenomena that may provoke system failures are described by *complex mathematical models*, which are then implemented in computer codes to simulate the behavior of the system of interest under various conditions (operational transitions and accident scenarios) [USNRC, 2009].

A quantitative model for the risk analysis of complex, (safety-critical) engineered systems may be viewed as composed of three main elements: an input vector $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_n\}$, a computer code (simulating the behavior of the system of interest) and an output vector $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_p\}$. The elements of the input vector \mathbf{Y} are all the *model parameters* and *input variables* needed to calculate one realization of the output variables \mathbf{Z} describing the system response. The computer code can be regarded as a *black box* which implements the complex, multi-dimensional, non-linear (possibly unknown) deterministic mathematical function $f(\cdot)$ that maps the input vector \mathbf{Y} into the output vector \mathbf{Z} :

$$\mathbf{Z} = f(\mathbf{Y}) = f(Y_1, Y_2, \dots, Y_n) \quad (1.1)$$

For *fixed* values of \mathbf{Y} , the output values \mathbf{Z} are *deterministically* computed.

The choice of a proper model is a trade-off between complexity and accuracy: "any model should be as simple as possible, and as complex as needed to answer the expected questions" [Haimes, 2009a].

Finally, *simulation* is the step used to evaluate the outputs of the model and to provide answers to the expected questions. Usually, the predictive models of the complex real-world systems include a large number of parameters and hypotheses, many of which are uncertain (see Section 1.6). This so called "high-dimensionality" problem constitutes a major challenge for the simulation of system behavior.

For illustration purposes, Figure 1.1 shows the concepts of representation, modeling and simulation with respect to a given real system at the top of the Figure.

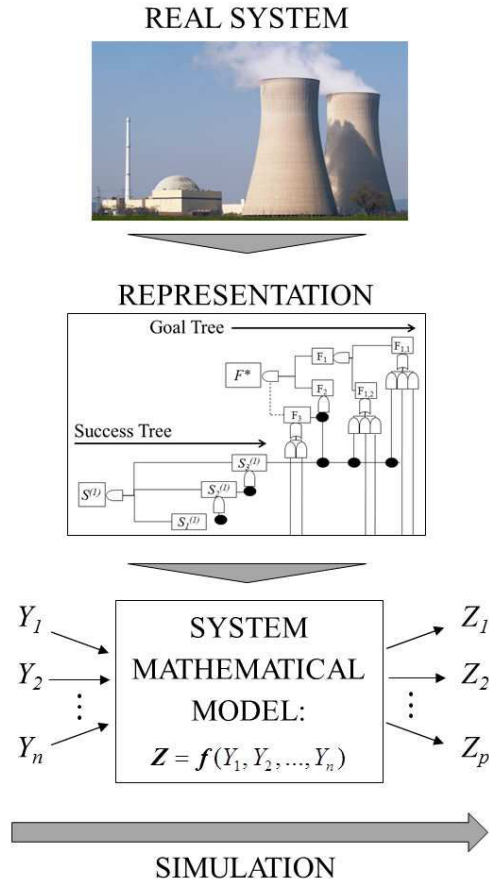


Figure 1.1: Concepts of representation, modeling and simulation with respect to a real system.

1.6. Uncertainty

In practice, *not all* the characteristics of the system under analysis can be fully captured in the mathematical model: as a consequence, *uncertainty* is always present in the values of the model *input parameters* and *variables* Y (see Section 1.5 for the notation). This translates into variability in the *model outputs* $Z = \{Z_1, Z_2, \dots, Z_p\}$ whose uncertainty must be estimated for a realistic assessment of the risk associated to system functioning. This concept is pictorially shown in Figure 1.2: the input parameters/variables $Y = \{Y_1, Y_2, \dots, Y_n\}$ are not fixed, known values, but they are uncertain and they can assume more than one value (in this case, only for illustration purposes, their uncertainty is represented by probability distribution functions $\{p^{Y_1}(Y_1), p^{Y_2}(Y_2), \dots, p^{Y_n}(Y_n)\}$); as a consequence of the uncertainty affecting the input variables $Y = \{Y_1, Y_2, \dots, Y_n\}$, also the output variables $Z = \{Z_1, Z_2, \dots, Z_p\}$ are uncertain (in this case, their uncertainty is described by proper probability distribution functions

$\{p^{Z_1}(Z_1), p^{Z_2}(Z_2), \dots, p^{Z_p}(Z_p)\}$ depending on the input probability distributions and on the mathematical model $f(\cdot)$.

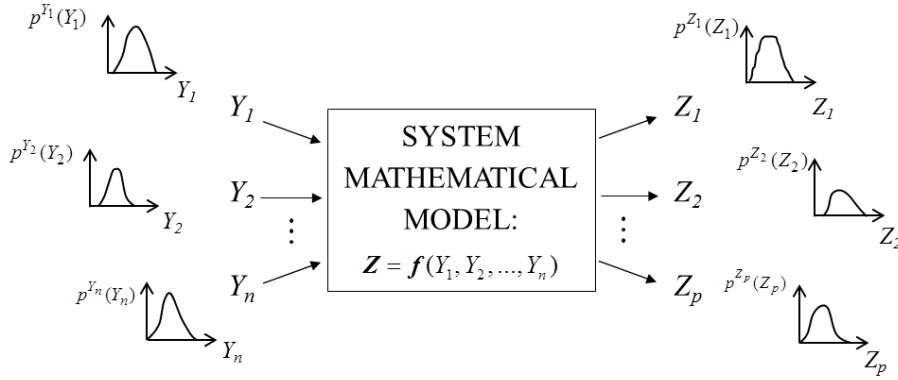


Figure 1.2: Uncertain input variables and output variables.

For the treatment of this uncertainty in risk assessment, it is often convenient to distinguish two types: *randomness* due to *inherent variability* in the system behavior and *imprecision* due to *lack of knowledge* and *information* on the system. The former type of uncertainty is often referred to as objective, aleatory, stochastic, irreducible uncertainty whereas the latter is often referred to as subjective, epistemic, state of knowledge, reducible uncertainty. The adjective “reducible” highlights that a gain of information about the system or environmental factors can lead to a reduction of epistemic uncertainty: this is possible because the epistemic component of uncertainty is not an inherent property of the system like the aleatory one [Apostolakis, 1990; Helton and Oberkampf, 2004].

The distinction between aleatory and epistemic uncertainty plays a particularly important role in the risk assessment framework applied to complex engineered systems that are critical from the safety viewpoint, e.g., in the nuclear, aerospace, chemical and environmental fields. In these contexts, the aleatory uncertainty is related to *random phenomena*, like the occurrence of unexpected events (e.g., the failure of a mechanical component) which define the various possible accident scenarios; whereas, epistemic uncertainty arises from a *lack of knowledge* as some phenomena and processes, and/or from the paucity of related operational and experimental data available [Ferson and Ginzburg, 1996; Helton and Oberkampf, 2004].

In the current risk assessment practice, both types of uncertainty are represented by means of probability distributions. However, resorting to a single probabilistic representation of epistemic uncertainty may not be possible when sufficient data is not available for statistical analysis, even if one adopts expert elicitation procedures to incorporate diffuse information into the corresponding probability distributions, within a subjective view of probability. Indeed, an expert may not have sufficiently refined knowledge or opinion to characterize the relevant epistemic uncertainty in terms of probability distributions [Helton and Oberkampf, 2004].

As a result of the potential limitations associated to a probabilistic representation of epistemic uncertainty under limited information, a number of alternative representation frameworks have been proposed, e.g., fuzzy set theory, evidence theory, possibility theory and interval analysis [Klir and Yuan, 1995; Aven and Zio, 2011; Aven et al., 2014].

In this work, probability theory is used to represent aleatory uncertainty: for example, the earthquake magnitude is described by a double truncated exponential distribution, or the time to recover of a component is described by a lognormal distribution; on the contrary, intervals are employed to represent epistemic uncertainty (e.g., in the transition probabilities between different components states).

Probabilistic uncertainty is propagated through the mathematical model by Monte Carlo simulation (MC) [Kalos and Whitlock, 1986; Zio, 2013], based on the repeated random sampling of possible model inputs and the running of the system model for the different input values sampled. A very large number of random realizations of the uncertain input parameters are typically necessary for a deep exploration of their ranges and a robust estimation of the model output uncertainty; for each input realization sampled, the computer code simulating the system behavior must be run: thus, the resulting computational cost may be very high and at times impractical. Interval uncertainty is instead propagated within the framework of interval analysis [Buckley, 2004].

1.7. Synthesis of the contribution of the thesis

In this Ph. D. thesis, challenges with respect to representation, modeling and simulation of SoS are tackled and applications related to i) external event risk assessment and ii) CIs are

considered. In particular, the first application deals with the safety and the physical resilience of a critical plant (i.e., a nuclear power plant) exposed to the risk of natural external events (i.e., earthquakes and aftershocks). The probability that the critical plant remains or not in a “safe state” (i.e., in a condition that does not cause health and/or environmental damages) upon the occurrence of an external event is taken as a quantitative indicator for safety, and the corresponding time to recover the safe condition is considered for the physical resilience assessment. The boundaries of the study are extended to the connected power and water distribution and transportation networks that can support the safety and the physical resilience of the plant in emergency condition, under a SoS framework. The second application considers the robustness and the recovery capacity of interdependent CIs (i.e., gas and electricity networks and SCADA system), under a SoS framework. Robustness is measured as the steady-state probability of the supply of product (i.e., gas and electricity) at the demand nodes. The recovery capacity is computed as the time needed to recover the SoS from the worst scenario to a level in which the demand nodes are satisfied.

In the following, the contributions of the work on representation, modeling and simulation are synthetically listed.

Representation

Different types of system representation frameworks have been considered and qualitatively compared. Some of them are entirely taken from the literature (e.g., the Fault Tree (FT) [Zio, 2007] and Muir Web [Sanderson, 2009; La Rocca et al., 2011]); others represent original extensions/modifications of existing techniques tailored to the framework of analysis of interest (as the Hierarchical Modeling and Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) [Hu and Modarres, 1999]). Finally, one representation method has been developed ex-novo, i.e., the Hierarchical Graph.

In extreme synthesis, the FT is a systematic, deductive technique which allows developing the causal relations leading to a given undesired event. It can provide qualitative information, e.g., how a particular event can occur and what consequences it leads to, and quantitative information, e.g., the probability of events of interest [Zio, 2007]. The Muir Web has been introduced in the context of ecological human community [Sanderson, 2009] and recently applied to infrastructure systems [La Rocca et al., 2011]. It is a network representation

technique, which allows analysis by graph theory and an explicitly representation of the structure of dependence of physical elements on the factors which influence their functionalities [Sanderson, 2009]. The Hierarchical Modeling allows describing the system at different levels of precision, facilitating the understanding of the system itself, providing information to support the decision-making process and reducing the computational cost of the system analysis [Gómez et al., 2013]. The GTST-DMLD provides an efficient and clear description of the system complexity through different hierarchical levels of system goals and functions, by the GT, and objects and parts, by the ST, and it highlights their relationships by the DMLD that translates into a dependency matrix and redefined logic gates, e.g., “AND” and “OR”, that assume a different meaning with respect to a binary state model [Hu and Modarres, 1999], e.g., FT. This approach has been extended both for the external-event-risk-assessment and the critical-infrastructure applications. The Hierarchical Graph has been introduced with respect to interdependent infrastructure networks and it structures network arcs in hierarchical levels on the basis of how many demands (i.e., loads) are served by an arc or group of arcs. This facilitates the identification of which parts of the network are more “important/critical” and can also support the evaluation of the partition of the product in the network, accounting for different importance criteria of the loads.

Different representations allow supporting different analyses and evaluating different outputs that are described in detail in Chapter 2.

Modeling

The state of the individual components and their interactions determine the state of the SoS performance (e.g., safety, robustness, recovery time), whose evaluation is supported by the representations introduced above. For the modeling of the state of the individual elements, two types of models have been considered and compared at component level: binary and multi-state. They result into binary and multi-state performances at SoS level, respectively. Binary state models allow faster simulation, but they can produce misleading results; on the contrary, multi-state models offer a more accurate approximation of reality, even though they require more information about the system to be modeled (e.g., the definition of the states, the probability to enter in a certain state and the transition probabilities from one state to another). In particular, with respect to the external-event risk-assessment application, the binary state model distinguishes between failed or not failed elements at component level, which

translates into a safe or not safe critical plant at SoS level; on the contrary, the multi-state model is able to describe two aspects at component level, i.e., the degrees of structural damage and functionality (i.e., performance), that translate into degree of safety of the critical plant at SoS level. Details of the structural and functional multi-state are provided in Section 3.3.1.

With respect to the critical-infrastructures application, the multi-state model refers to the functional performances of the elements that are described by Markov and semi-Markov processes. Further details are given in Section 3.3.2.

Simulation

MC simulation has been adopted to evaluate the vulnerability and resilience of SoS. In particular, with respect to the external-event risk-assessment application, MC simulation has been integrated into the Seismic Probabilistic Risk Assessment (SPRA) framework including also the occurrence of aftershocks. With respect to the CI application, MC simulation has been combined with interval analysis to process the epistemic uncertainty in the state transition probabilities.

Figure 1.3 sums up the work performed in the present Ph. D. thesis with respect to the external-event risk-assessment application (on the left) and the CI application (on the right).

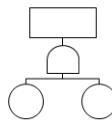

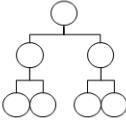
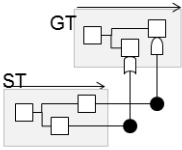
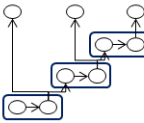
Representation						
Modeling	Binary state			Multi-state	Multi-state	
Simulation	Monte Carlo simulation for Seismic Probabilistic Risk Assessment				Monte Carlo simulation and interval analysis	Monte Carlo simulation
Hazardous events	Earthquake			Earthquake and aftershocks	Random failures	
Quantities of interest	Safety		Safety and recovery time		Robustness and recovery time	Robustness
Application	- Nuclear power plant provided with proper internal emergency devices and embedded in the connected power and water distribution, and transportation networks				- Interconnected gas and electricity network and SCADA system	- Interconnected gas and electricity network and SCADA system - IEEE 123 node test feeders
<div><div>External event risk assessment</div><div>Critical infrastructures</div></div>						

Figure 1.3: Synthesis of the work carried out in this Ph. D. thesis

1.8. Structure of the thesis

The thesis is composed of two parts. Part I, subdivided in seven Chapters, introduces and addresses the problems in further details and illustrates the methodological approaches developed and employed in this Ph. D. work. Part II is a collection of six selected papers published, submitted for publication or under submission as a result of the work and which the reader is referred to for further details.

Table 1.2 summarizes the thesis structure. Chapter 2 gives an overview of the system representation methods and illustrates in details those employed in this thesis. Chapter 3 tackles the issue of the mathematical modeling under the binary and multi-state perspectives. Chapter 4 provides the details of the simulation and uncertain propagation approaches, and describes the specific algorithms implemented in this work. Chapter 5 illustrates the applications related to external event risk assessment (i.e., to a nuclear power plant embedded in the power and water distribution and transportation networks, exposed to the risk of earthquakes and aftershocks). Chapter 6 describes the applications to CIs (i.e., to gas and electricity networks and SCADA system). Figure 1.4 provides a pictorial view of the issues

and the approaches considered in the present Ph. D. work on vulnerability and resilience analysis of SoS.

Part II includes the collection of papers published, submitted or under submission as a result of the research performed. Papers I – IV present the applications related to the external event risk assessment. In particular, papers I – III adopt the FT, Muir Web and Hierarchical Modeling representations, respectively, considering a binary state model to describe a nuclear power plant exposed to the risk of earthquakes. In paper I and II, the safety of the plant is evaluated, while in paper III, also the physical resilience is considered in terms of time to recover safety. In paper IV, the GTST-DMLD is employed within a multi-state modeling framework. The evaluation of the safety and physical resilience is carried out for a nuclear power plant exposed to the risk of earthquakes and subsequent aftershocks. The case study is extended with respect to the case study of papers I – III. The quantitative evaluation of the analysis is performed by MC simulation for SPRA.

Paper V presents the applications related to CIs: a GTST-DMLD is adopted to describe the interconnected gas and electricity networks and SCADA system, and evaluate i) the robustness of the SoS measured in terms of probability of the product (gas and electricity) that can be given to the load nodes at steady state and ii) the recovery time of the system starting from the worst scenario. Random failures modelled as Markov and semi-Markov processes are used within a multi-state model of the arc capacities of the gas and electricity networks. In addition, epistemic uncertainty in the state transition probabilities is taken into account and propagated into the model outputs.

A paper VI is under submission: it deals with a Hierarchical Graph representation developed in the research of the Ph. D. thesis and applied to the same case study of paper V and to the IEEE 123 node test feeders to demonstrate its application on large systems.

Table 1.2: Structure of the thesis.

Topic	PART I Chapter(s)	PART II Paper(s)
Representation Fault Tree Muir Web Hierarchical Modeling GTST-DMLD Hierarchical Graph	2	I – VI I II III IV – V VI
Modeling Binary state Multi-state	3	I – VI I – III IV – VI
Simulation and uncertain propagation Monte Carlo Monte Carlo and interval analysis	4	I – VI I – IV, VI V
Applications External event risk assessment Critical infrastructures	5 – 6	I – VI I – IV V – VI

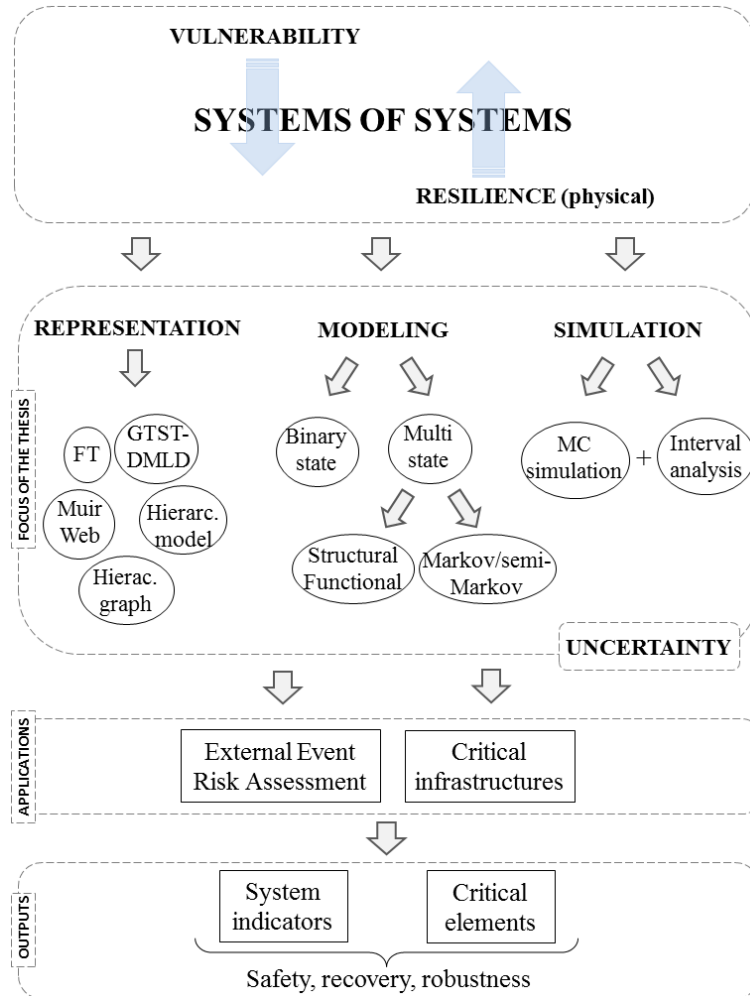


Figure 1.4: Pictorial view of the flow (topic; focus, applications and outputs) of the present Ph. D. work on reliability and resilience analysis of system of systems.

2. SYSTEM REPRESENTATION

In this Chapter, the issue of system representation is tackled. First, a brief overview of the existing system representation techniques is carried out (Section 2.1); then, the representation approaches adopted in this thesis, i.e., Fault Tree (FT), Muir Web, Hierarchical Modeling, Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) and Hierarchical Graph, are explained in detail (Sections 2.2 – 2.6) under a system-of-systems (SoS) framework; finally, the techniques are compared (Section 2.7).

2.1. Overview on the existing system representation techniques

Several types of system representation approaches exist in literature and they rely mainly on a hierarchy or graph structure.

Hierarchical Modeling have been often adopted to represent and model complex systems, since many organizational and technology-based systems are hierarchical in nature [Haimes, 2012]. “Frequently, complexity takes form of hierarchy, whereby a complex system is composed of interrelated subsystems that have in turn their own subsystems, and so on, until some lowest level of elementary components is reached” [Courtois, 1985]. This approach can be based on different perspectives, e.g., functional, technical, organizational, geographical, political, etc., and can allow simplifying the modeling process and the ultimate management of the system as a whole [Haimes, 2012].

Hierarchical functional models include Goal Tree Success Tree (GTST) – also combined with Master Logic Diagram (MLD) – and Multilevel Flow Modeling (MFM). The GTST is a functional hierarchy of a system organized in levels starting with a goal at the top; the MLD, developed and displayed hierarchically, shows the relationships among independent parts of the systems: the combined GTST – MLD provides a powerful functional/structural description method. Finally, the dynamic version of the approach, namely the GTST – Dynamic MLD (GTST-DMLD), allows describing the temporal behavior of the systems [Hu and Modarres, 2000]. Further details about the GTST-DMLD are provided in Section 2.5. Multilevel Flow Models [Lind, 2011a; Lind, 2011b], developed in the field of artificial intelligence, have been proposed for qualitative reasoning, i.e., for representing and structuring knowledge about physical phenomena and systems. They consider cause-effect

relations and facilitate the reasoning at different levels of abstraction on the basis of "means-end" and "whole-part" decomposition and aggregation procedures. Goals, functions and flow of material, energy and information are connected to form a hyper graph. They are mainly used for measurement validation (e.g., for checking the measurement of a mass or energy flow), alarm analysis (e.g, for the identification of primary and secondary alarm), and fault diagnosis (i.e., for the identification of the consequences and the root causes of a disturb in the system functioning) [Larsson, 1992].

In risk analysis, common representation techniques are hierarchical trees that are commonly used to identify i) the initiating causes of a pre-specified, undesired event or ii) the accident sequences that can generate from a single initiating event, through the development of structured logic trees, i.e., fault and event trees, respectively [Zio, 2007].

In complex network theory, instead, complex systems are represented by networks where the nodes stand for the components and the links describe the physical and relational connections among them. Network-based approaches model interdependent critical infrastructures (CIs) on the basis of their topologies or flow patterns by topology-based and flow-based methods, respectively. The first class of approaches allows identifying relevant structural properties of the system. Topological approaches consider two possible states for the components, failed and functioning, and can measure the strength of the connections by including weighted links. They can be solved analytically or by simulation depending on the level of detail used to model the CIs topologies [Ouyang, 2014]. Flow-based methods, instead, capture the dynamic complexity of real CIs by considering the evolution in time of physical variables of flow (e.g., material, energy, information) [Fang et al., 2014].

Probabilistic modeling includes Petri nets, Bayesian networks and flowgraphs. A Petri net is a directed graph that consists of places (i.e., conditions), transitions (i.e., events that may occur) and directed arcs describing which places are pre- and/or post-conditions for each transition. They are well suited for modeling the behavior of distributed systems. Laprie et al. have adopted Petri nets to describe and analyze high level scenarios that may take place when failures occur in two interdependent infrastructures (i.e., in information and electricity infrastructures), considering the effect that failures in one infrastructure have on the other and accounting also for malicious attacks [Laprie et al., 2007]. Bayesian networks are based on directed acyclic graphs where nodes are random variables representing the state of components and edges are conditional dependencies, reflecting the causal relationships

among adverse events. Classical Bayesian networks provide static models of the system at each time step; however, recently, dynamic Bayesian networks have been introduced [Ouyang, 2014]. Differently, flowgraphs model the outcomes of random variables: in this framework, nodes identify the actual physical state of a system and edges model the allowable transitions, the probabilities of different outcomes, and waiting times until the occurrence of outcomes of interest [Huzurbazar, 2005].

2.2. Fault Tree

FT analysis is a systematic, deductive technique which allows developing the causal relations leading to a given undesired event [Zio, 2007]. It is deductive in the sense that it starts from a defined system failure event (called top event) and unfolds backward its causes down to the primary (basic) independent faults. The method focuses on a single system failure mode and can provide qualitative information on how a particular event can occur and what consequences it leads to, while at the same time allowing the identification of those components which play a major role in determining the defined system failure. Moreover, it can be solved in quantitative terms to provide the probability of events of interest starting from knowledge of the probability of occurrence of the basic events which cause them [Zio, 2007].

To build a FT it is necessary to understand how the system functions and to select a system failure event of interest with respect to the analysis is performed. Then, the contributing events that may directly cause the occurrence of the top event are identified and in turn each event must be examined to decide whether it is to be further decomposed in more elementary events contributing to its occurrence. If the event is a primary failure, the corresponding branch of the tree is terminated and this primary event is symbolically represented by a circle. This also implies that the event is independent of the other terminating events which will be eventually identifies and that a numerical value for the probability of its occurrence is available if a quantitative analysis of the tree is to be performed. Otherwise, if the event is to be broken down further in more primary failure causes, it must be examined to identify the sub-events which contribute to its occurrence [Zio, 2007].

The events in the tree are connected logically via OR, AND, INHIBIT (and others) functions graphically shown through corresponding gates; for illustration purpose the AND and OR gates are reported in Figure 2.1: the AND gate is used for output events that occur if all input events occur simultaneously, whereas, the OR gate is adopted if any one of the input events occurs.



Figure 2.1: Example of two logic gates: AND gate on the left, OR gate on the right

A FT can be described by a set of Boolean algebraic equations, one for each gate of the tree. For each gate, the input events are the independent variables and the output event is the dependent variable. It is possible to solve these equations using the rules of Boolean algebra and obtain an expression (called *switching* or *structure* function) for the top event based on primary events only [Zio, 2007]. Considering N system components which state (functional or failure) is described by the corresponding variables g_1, g_2, \dots, g_N , the structure function Φ can be expressed as: $\Phi(g) = \Phi(g_1, g_2, \dots, g_N)$.

The structure function can be written as the union of the fundamental products (i.e., the products containing all the input variables, complemented or not) which correspond to the combinations of the variables which render the function true (canonical expansion [Barlow, 1998]).

Coherent structure functions are considered in this work; they are characterized by several properties including: i) the system is successful if all components are in their success state, ii) the system is failed if all the components are failed, iii) the system cannot change from the success to the failed state if the performance of a component is improved [Zio, 2007].

Coherent structure functions can be expressed in reduced expressions in terms of *minimal cut sets* (M_1, M_2, \dots, M_{mcs}), i.e., a set of components whose failure ensure the failure of the system and that does not have another cut as a subset. By definition, the system structure function is the union of the *mcs* minimal cut sets [Zio, 2007]:

$$\Phi(g) = 1 - (1 - M_1) (1 - M_2) \dots (1 - M_{mcs}) \quad (2.1)$$

Minimal cut sets can be identified by inspection if few components are involved, however the number of minimal cut sets increases rapidly with the increasing of the system complexity and a systematic approach is needed. They allow performing qualitative analysis by determining the criticality of the various components and quantitative analysis by computing the probability of the top event [Zio, 2007].

Further details are not given here for brevity sake, the interested reader is referred to [Barlow, 1998; Bedford and Cooke, 2001; Aven, 2003; Zio, 2007].

In Table 2.1, the main advantages and limitations of FT Analysis are given.

Table 2.1: Advantages and limitations of Fault Tree Analysis.

Fault Tree Analysis		
	Advantages	Limitations
Qualitative analysis	The physical elements are represented in a well-defined structure, according to the logic of the system that leads to the identification of the Minimal Cut Sets.	Additional factors (operational, organizational, etc.) are not included. The exhaustive identification and manipulation of the Minimal Cut Sets can be difficult for large systems.
	The structured representation allows a rigorous and transparent analysis.	Difficult to build the FT, in particular in the case of large number of components and complicated logic, dependencies, etc. No flexibility: the addition of a new component can change the entire structure of the FT.
	The representation is clear and allows understanding which combinations of components cause the failure of the top event. The modelization is straightforward via few, simple, logic operators.	The Boolean-logic based approach does not allow considering the strength of the relationships.
Quantitative analysis	Numerical calculation of the probability of occurrence of the top event by transforming the logical structure into an equivalent probability form.	Difficulty in treating the dynamics of failures.

In this work, the FT representation is adopted to evaluate the safety of a critical installation exposed to the risk of earthquake briefly, presented in Section 5.1; the boundary of the study includes the connected infrastructure services in which the plant is embedded and that support its operation adopting a SoS framework; the reader is referred to Section 5.2.1 and to paper I of Part II for the application of the FT on the SoS of interest. A binary (failed-functioning) modeling framework is considered to model the states of the system (Section 3.2); finally, the operative steps used to simulate the system behavior are given in Section 4.2.1.

2.3. Muir Web

Muir Web is a system analysis technique used to represent a complex system and the relationships among its elements. It has been first introduced in the context of ecological human community to explicitly represent the structure of dependence of the physical elements on factors which influence their functionalities [Sanderson, 2009]. Actually, in this field, traditionally only the major interactions are taken into account in the system modeling: for example, with reference to the food chain, only the connections between predator and prey are usually considered, whereas other relevant and influencing relationships exist between organisms, e.g., one species may take cover for another, and other factors contribute to the food chain, e.g., abiotic elements like water, sun, soil, rainfall, wind [Sanderson, 2009]. By the representative power of Muir Web, the traditional picture of dependencies is extended through a graph where the nodes represent all the system elements (e.g., species and abiotic factors in the ecological case) and the edges represent their dependency structure.

The concept of Muir Web has been recently applied also to infrastructure systems, exploiting some similarities which exist between the ecological and the infrastructure networks [La Rocca et al., 2011]: both are large scale systems with complex interactions and can fail when an external event occurs. In the case of infrastructure systems, the nodes of the web are system components, e.g., a pump, and other factors which influence the infrastructure state, e.g., a stable soil with respect to seismic hazard.

The Muir Web is a network representation technique, which allows analysis by graph theory. It is a tool to visualize, capture and understand the relations among physical elements and factors of a system, and it organizes the knowledge in a comprehensive way through its multi-dimensional structure. It is inspired by the view of John Muir, the famous naturalist [Muir et al., 1985]: “When we try to pick out anything by itself we find that it is bound fast by a thousand invisible cords that cannot be broken, to everything in the universe”. The original purpose behind the introduction of the Muir Web was to recreate the landscape and the wildlife of the city of Mannahatta four hundred years ago to see how that place was before it became a city and to reimagine the city’s development taking into account the natural cycles and processes [Sanderson, 2009]. For this aim, the Muir Web can be converted into maps by an iterative computer program that works through all the relationships and find the right layers in a Geographic Information System [Sanderson, 2009].

In the Muir web representation there is no difference among the types of relations: they are depicted by arrows that are directed from an element to another dependent on it. Applying it to an engineered system means to consider all the other elements (physical, operational, organizational, etc.), which each single element depends on, including, for example, the type of soil, the maintenance task, the presence of operators, etc. One main objective of the Muir Web is to visualize all the connections among elements. This gives the basis for performing further analysis to characterize the types of relations, the way a failure of an element can affect the state of another connected element, the elements with significant influence on the system functionality and those with little influence.

For a general representation of a SoS based on the Muir Web framework (Figure 2.2), let us consider a plant H that is critical from the safety viewpoint, i.e., if it is not provided with the necessary service inputs it can reach a condition which causes health and environmental damages. The state of the critical plant H is the state of its critical element, E . Connections exist to N_S interdependent systems S_i , $i = 1, \dots, q, q + 1, \dots, N_S$, numbered in order in such a way that the first q are those inside the plant and the last $N_S - q$ belong to systems outside the plant. The systems internal to the plant S_i , $i = 1, \dots, q$, are designed to provide inherent safety, i.e., the input services required to keep E in a safe state. Each system is composed by N_{c_i} , $i = 1, \dots, N_S$, components and the overall SoS is therefore formed by $N = N_{in} + N_{out}$ components, where $N_{in} = \sum_{i=1}^q N_{c_i}$ and $N_{out} = \sum_{i=q+1}^{N_S} N_{c_i}$. For the sake of clarity of the representation, we distinguish the intra-system and inter-systems links, i.e., the links among components of the same system and of different systems, respectively, into two types here called “direct dependency” and “support dependency” on the basis of their physical meaning: for the first type, when a component fails, its direct neighbors also fail; for the second one, when a component fails, it does not cause the failure of its neighbors because it assumes the role of “support”, i.e., it is useful to the neighbors when these fail for other reasons. In addition, the links between the interconnected infrastructure systems and the critical plant have been considered. They represent unidirectional dependency, but if a connected system fails, it does not mean that the critical plant fails too; identification, specification and joint analysis of all these dependencies have to be performed to determine their effect on the critical plant, as explained in Section 4.2.1. The Muir Web of Figure 2.2 shows an example in which the element E (star) of the critical plant H (dotted-rectangular shape) is connected to

four interdependent systems S_i , $i = 1, \dots, 4$ with $N_{c_1} = 5$, $N_{c_2} = 6$, $N_{c_3} = 7$ and $N_{c_4} = 3$ components. The systems S_1 and S_2 are inside the plant and the systems S_3 and S_4 are outside. The direct dependencies are represented by solid lines, the support ones by dashed lines and the connections to the critical plant by bold lines.

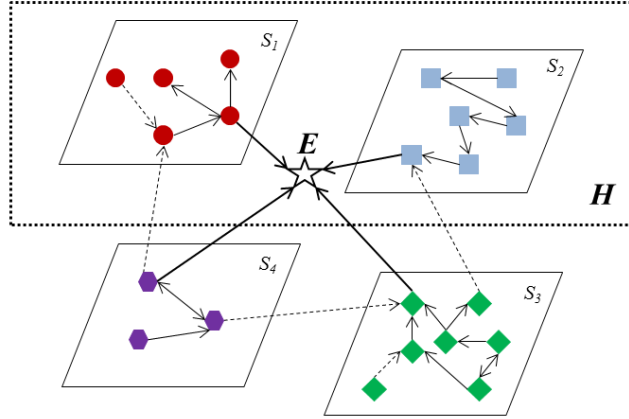


Figure 2.2: Muir Web representation of a system of systems made of a critical plant, H (dotted-rectangular shape) whose safety is identified in the state of its critical element E , and four interdependent systems S_i , $i = 1, \dots, 4$, whose elements (represented by circles, squares, rhombs and hexagons, respectively) are connected by direct dependencies (solid lines) and support dependencies (dashed lines). The systems S_1 and S_2 are inside the critical plant, whereas the systems S_3 and S_4 are outside. The links to the critical element E (star) of the critical plant are the bold lines.

The structure of dependence represented in the Muir Web drives the identification of the functional logic relations among the components within each system (intra-system links) and among different systems (inter-system links).

In Table 2.2, the main advantages and limitations of Muir Web are illustrated.

Table 2.2: Advantages and limitations of Muir Web.

Muir Web		
Advantages		Limitations
Qualitative analysis	Representing the invisible: in addition to the physical elements, the representation includes the factors (operational, organizational, etc.) which the physical elements depend on. The associated knowledge is organized in a comprehensive way through a multi-dimensional structure.	A large amount of information and competences of different disciplinary fields are needed to build the representation.
	Easy to build the network answering the question “why” to identify the depending elements.	A further analysis is needed to identify the logic structure of the system.
	Extendable/Flexible: the addition of a new component is possible without changing all the structure.	A further analysis is needed to identify the logic structure of the system.

	The representation clearly illustrates the dependencies among the components: arrows are directed from one element to another dependent on it. In addition, there is the possibility of including the strength of the relationship ¹ .	A further analysis is needed to identify the logic structure of the system.
	Possibility to be converted into maps, resorting to the support of Geographic Information Systems.	
Quantitative analysis	Simulation: propagation of failures in the network.	High computational cost of simulation for large systems.

In this work, the Muir Web is adopted to evaluate the safety of a critical installation exposed to the risk of earthquake, briefly presented in Section 5.1; the reader is referred to paper II of Part II for further details and examples of the functional logic relations in a Muir Web. A binary modeling framework is considered for the components/system states (Section 3.2); the operative simulation steps of the analysis are given in Section 4.2.1.

2.4. Hierarchical Modeling

The idea behind the use of hierarchical representation (also adopted in the GTST-DMLD described in Section 2.5) is that most complex systems are in the form of a hierarchy (see the previous Section 2.1). The Hierarchical Modeling allows studying the system at different level of details and extracting from each level the groups of components that are critical from the safety viewpoint.

For a formal conceptualization of this approach, let us denote a system i at the level L of the hierarchy as $S_i^{(L)}$ and by $N_S^{(L)}$ the number of systems at the level L . In the hierarchical representation of a SoS view of a critical plant, H , at the top of the hierarchy there is only $N_S^{(1)} = 1$ system, the critical plant itself, and it is denoted as $S_1^{(1)}$. At the second level, $L = 2$, this is connected to $N_S^{(2)}$ systems, $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, inside and outside the plant, that provide it with the necessary inputs for its operation. The systems $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, at level $L = 2$, can, in turn, be broken down into subsystems $S_i^{(3)}$, $i = 1, \dots, N_S^{(3)}$ at the third level of the hierarchy, $L = 3$. The Hierarchical Modeling is built by identifying the elements (or

¹ The strength of the relationship has not been included in the present thesis but it has been considered in [Sanderson, 2009], characterizing the relationships by modifiers like “especially” or “often”.

groups) that are “part of” the parent objects, and continuing up to the desired level $L = N_L$, where N_L is the number of levels of the hierarchy. For the analysis of interest here, the hierarchy is continued down to the level of details of the individual components of the SoS. However, following this procedure for building the hierarchical model, some components may not be considered. Actually, some elements of the SoS i) may not provide the critical plant H with the inputs necessary for its operation, thus, they cannot be represented in the level-2 of the hierarchy, and ii) may not be part of any system $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, thus, they cannot be identified by the decomposition criteria described above. This is the case, e.g., of those elements (hereafter called “recovery supporting elements”) that provide only the components (or groups) of the systems $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, with the inputs necessary for their functioning or recovery (i.e., they are not physically part of any system $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$): for these elements, the original Hierarchical Modeling framework detailed above is extended in the sense that they are here represented as a part of the systems (groups) they support.

By way of example, Figure 2.3 depicts the graph of the system (top), the grouping of its components (middle) and its hierarchical representation (bottom). The intra-system dependencies (solid lines), the inter-system ones (dashed lines) and the connections to the critical plant H (bold lines) are identified (Figure 2.3, top). The increasing resolution in the four levels considered is illustrated (Figure 2.3, middle): in the first level (square shape), the critical plant H is represented; in the second level (dashed oval shape), the three interdependent systems, $S_i^{(2)}$, $i = 1, \dots, 3$ are reported; in the third and fourth levels (dotted and solid oval shapes, respectively), the grouping of the elements within the systems of level 2 are specified. In Figure 2.3, top, the recovery supporting elements are those not connected to the critical plant H but linked to other components by dashed lines (i.e., $S_1^{(4)}$ and $S_2^{(4)}$); in Figure 2.3, middle, they are grouped in the systems to which they provide support, e.g., $S_1^{(4)}$ is both in the systems $S_1^{(2)}$ and $S_2^{(2)}$ and $S_2^{(4)}$ is in the system $S_3^{(2)}$; in Figure 2.3, bottom, they are represented in the last levels of the hierarchy according to the grouping of the Figure 2.3 in the middle.

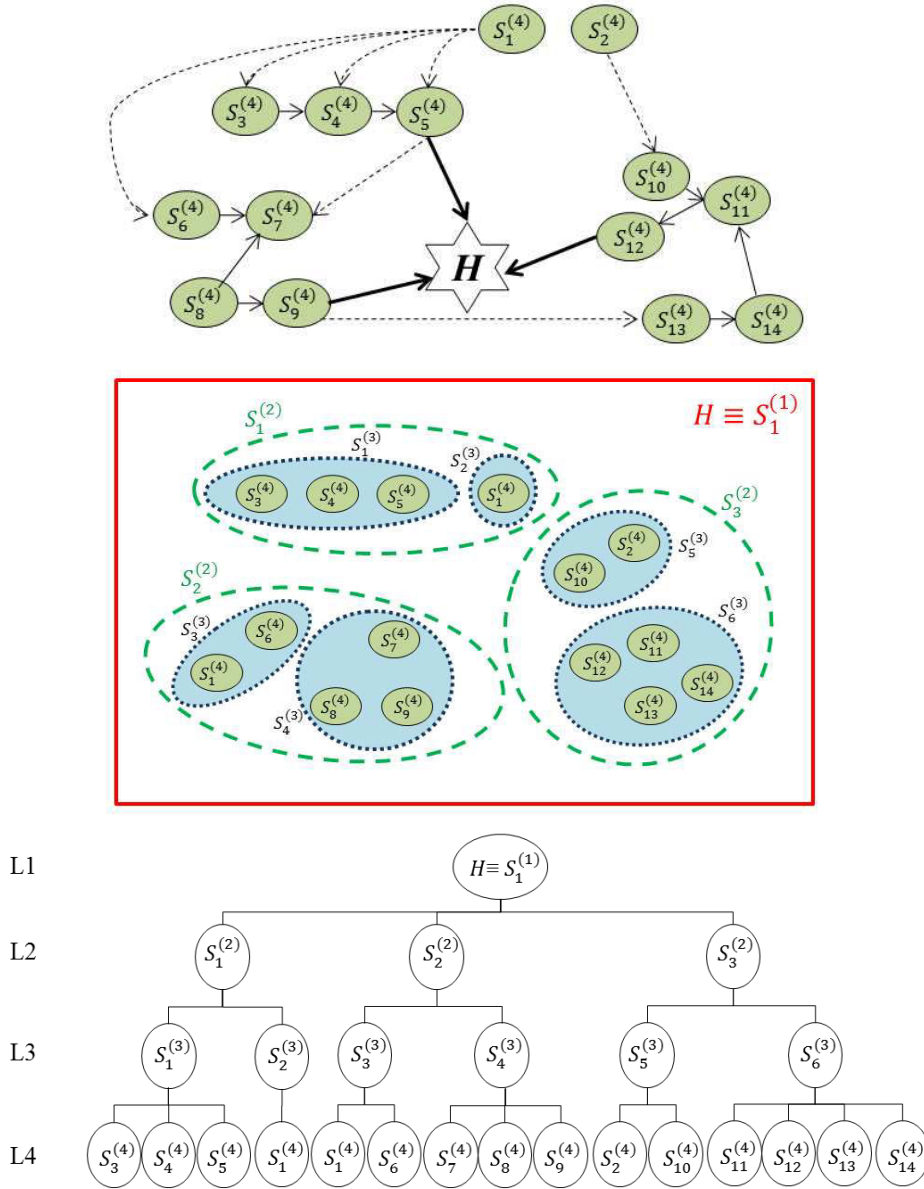


Figure 2.3: Top: dependencies among the components of the system of systems; the links represent the intra-systems dependencies (solid lines), the inter-systems dependencies (dashed lines) and the dependencies of the critical plant H on its interconnected systems (bold lines). Middle: graphical representation of their grouping; the rectangular, dashed, dotted and solid oval shapes represent the increasing resolution in the hierarchical level. Bottom: corresponding hierarchical representation; L: Level.

Notice that the recovery supporting elements can belong to more systems (or groups) since they can be a support to different components (or groups), whereas all the others components (or groups) appear within just one system, since the hierarchy is mainly built following the criterion “to be a part of”. A final remark is in order with respect to the top-down approach adopted to build the hierarchical model. It is possible that, before reaching the bottom of the hierarchy, some components cannot be subdivided further (e.g., $S_2^{(3)}$ coincides with $S_1^{(4)}$)

leading to an incomplete hierarchical representation. Therefore, in this circumstance a copy of those elements is reported in the levels they are absent [Gómez et al., 2011].

In Table 2.3, the main advantages and limitations of Hierarchical Modeling are reported.

Table 2.3: Advantages and limitations of Hierarchical Modeling.

Hierarchical Modeling		
	Advantages	Limitations
Qualitative analysis	The system is broken up according to its parts and it is analyzed at different levels of detail.	Additional factors (operational, organizational, etc.) are not included.
	It is easy to build the hierarchy and to identify the parts of the system with increasing level of detail.	A further analysis is needed to identify the logic structure of the system.
	Analyzing the system at different levels of detail allows a good understanding of the system structure.	The addition of a new component can change the hierarchy.
	The representation is clear and allows understanding the composition of the system at different levels of detail.	The representation does not show the relationships among the components at each level of the hierarchy and a further analysis is needed for that.
Quantitative analysis	Simulation: propagation of failures bottom-up through the hierarchy.	High computational cost of simulation for large systems.

In this thesis, we have used the Hierarchical Modeling to evaluate the safety and the recovery capacity of a critical plant exposed to the risk of earthquake, briefly presented in Section 5.1, adopting a binary state modeling (Section 3.2); the operative steps used to simulate the system behavior are given in Section 4.2.2. The analysis proceeds from the bottom to the top of the hierarchy: the hierarchical levels at the top depend on those at the bottom that show a higher level of detail. In extreme synthesis, given the state (operational or failure) of the components of the bottom of the hierarchy and their logic connections, it is possible to determine the state of the grouped components at the previous hierarchical level and proceeding back to the top. This concept is explained more precisely in the following paragraphs first with respect to the safety analysis and then to the quantification of the recovery capacity of a critical plant. The reader is referred to Section 5.2.2 and to paper III of Part II for the application of the Hierarchical Modeling on the SoS under analysis.

A system $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, at level $L - 1$, $L = 2, \dots, N_L$, can be in an operational or in a failure state depending on the states of the systems at the level L , on their functionality and

on their logic connections. A state (truth) matrix is associated to each system $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, $L = 2, \dots, N_L$, where the first columns represent the states of the systems $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, at level L and the last column represent the state of the system $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, at level $L - 1$. The entries are equal to 1 or 0 according to whether the states are in a failure state or not.

By way of example, refer to Table 2.4 and Figure 2.4 where three state matrices and the corresponding FTs are reported, with reference to the system $S_5^{(3)}$ at level $L = 3$ of Figure 2.3 (middle) composed by the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ at level $L = 4$. The first two state matrices represent, respectively, the series and parallel configurations between the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ (illustrated by the OR and the AND gate in the FTs): in the first case, the state of $S_5^{(3)}$ can assume only one operational state, since the failure of $S_{10}^{(4)}$ or $S_2^{(4)}$ causes its failure; on the contrary, in the second case, $S_5^{(3)}$ is in a failure state when both $S_{10}^{(4)}$ and $S_2^{(4)}$ fail. The third matrix shows a case in which the state of $S_5^{(3)}$ depends only on the state of $S_{10}^{(4)}$. The FT of this last case is represented by an inhibit gate without condition on the system $S_2^{(4)}$.

Table 2.4: Three possible state matrices for the system $S_5^{(3)}$ of Figure 2.3 (middle) on the basis of the states of the systems $S_{10}^{(4)}$ and $S_2^{(4)}$. On the left: $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in series; in the middle: $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in parallel; on the right: $S_5^{(3)}$ depends only on $S_{10}^{(4)}$; 1 represents the failure state.

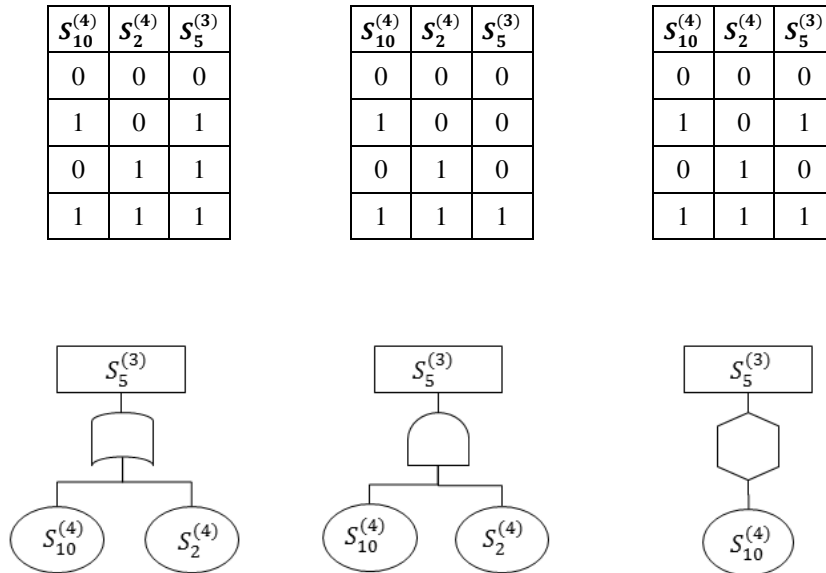


Figure 2.4: Corresponding fault tree representations of the state matrices reported in Table 2.4. On the left, $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in series (OR gate); in the middle, $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in parallel (AND gate); on the right, $S_5^{(3)}$ depends only on $S_{10}^{(4)}$ (INHIBIT gate without condition).

To define the appropriate state matrix for the systems $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, $L = 2, \dots, N_L$, a deep understanding of their functionality is necessary. The dependencies identified in Figure 2.3 (top) are a support for this analysis.

With respect to the recovery capacity, if the systems at level L are connected in series to the system at level $L - 1$, the recovery time of the latter is the maximum recovery time of the systems or components at the lower level L (Figure 2.5, left); if they are connected in parallel, the recovery time is the minimum (Figure 2.5, middle). In other cases, specific evaluations should be performed. For example, if the failure of a given system $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, does not affect the state of another system $S_j^{(L)}$, $j = 1, \dots, N_S^{(L)}$, $j \neq i$, but plays a role in the operations needed for its recovery from failure it should be considered in the analysis like an increasing time for operations of recovery of the system at level $L - 1$ (Figure 2.5, right).

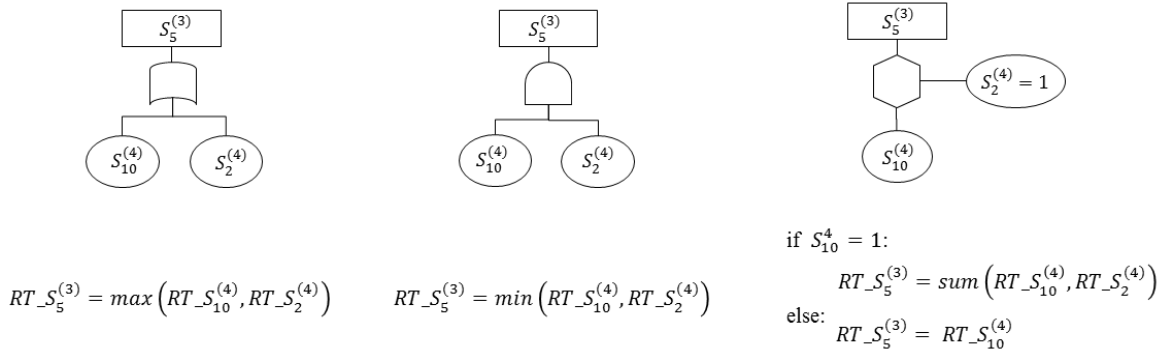


Figure 2.5: Computation of recovery time (RT) of the system $S_5^{(3)}$ with reference to three different configurations of the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ represented in the fault tree. On the left: OR gate, the recovery time of $S_5^{(3)}$ is the maximum recovery time of $S_{10}^{(4)}$ and $S_2^{(4)}$. In the middle: AND gate, the recovery time of $S_5^{(3)}$ is the minimum recovery time of $S_{10}^{(4)}$ and $S_2^{(4)}$. On the right, INHIBIT gate: the recovery time of $S_5^{(3)}$ is the recovery time of $S_{10}^{(4)}$ but if the condition $S_2^{(4)} = 1$ is verified, the recovery time is the sum between the recovery times of $S_{10}^{(4)}$ and $S_2^{(4)}$. 1 represents the failure state.

2.5. Goal Tree Success Tree – Dynamic Master Logic Diagram

The GTST-DMLD is a goal-oriented method based on a hierarchical framework [Hu and Modarres, 1999]. It gives a comprehensive knowledge of the system describing the complex physical systems in terms of functions (qualities), objects (parts) and their relationships (interactions). The first part is developed by the Goal Tree (GT), the second one by the Success Tree (ST) and the third one by the DMLD [Hu and Modarres, 1999].

The GT identifies the hierarchy of the qualities of the system decomposing the objective of the analysis, i.e., the goal, into functions that are in turn divided into other functions and so on. The hierarchy is built by answering questions on “how” the subfunctions can attain the parent functions (looking at the hierarchy from top to bottom) and on “why” the functions are needed (looking at the hierarchy from bottom to top). Two types of qualities, i.e., main and support functions, are considered: the former directly contribute to achieving the goal, whereas, the latter support the realization of the former [Brissaud et al., 2011]. For example, the goal function of safely generating electric power in a nuclear power plant is attained by many functions as heat generation, heat transport, emergency heat transport, heat to mechanical energy transformation, mechanical to electrical energy transformation [Modarres et al., 1999]. Each of these functions require the support of other functions, e.g., emergency heat transport may require internal cooling [Modarres et al., 1999] or a pump whose function is to “provide pressure” require the support functions “provide ac power”, “cooling and lubrication”, “activation and control” [Modarres et al., 1999].

The ST represents the hierarchy of the objects of the system, from the entire system to the parts necessary to attain the last levels of the GT. This hierarchy is built identifying the elements that are “part of” the parent objects. As for the GT, two types of objects are distinguished also in the ST: main and support. The former are directly contributing to achievement of the main functions, whereas the latter are needed for the operation of the former [Brissaud et al., 2011]. For example, generating power plants, electric power transmission and distribution networks are the support objects to provide ac power to a pump.

The DMLD is an extension of the Master Logic Diagram (MLD) [Hu and Modarres, 1999] introduced to model the dynamic behavior of a physical system. It describes the interactions between parts, functions and parts and functions, in the form of a dependency matrix, and it include the dynamics by means of time-dependent fuzzy logic rules [Hu and Modarres, 1999].

A conceptual sketch of GTST-DMLD is given in Figure 2.6.

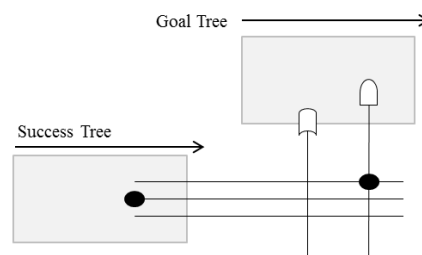


Figure 2.6: Conceptual sketch of GTST-DMLD.

The GT is drawn at the top, the ST on the left and the DMLD is represented by filled dots at the intersections between vertical and horizontal lines, to indicate the possible dependencies between the elements on the left and on the top. Several types of logic gates can be used to represent the time-dependent fuzzy logic rules, and different dependency-matrix nodes to describe the probabilities and degrees of truth in the relationships [Hu and Modarres, 1999]. Figure 2.7 gives an example of dependency of an element C on two elements A and B by the “AND” gate in a DMLD [Hu and Modarres, 1999]. In this case, the output value of the element C is the minimum value between the inputs A and B. Replacing the “AND” gate with an “OR” gate, the output value will be the maximum between the input values.

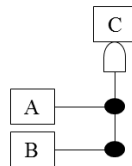


Figure 2.7: Example of an element C that depends on two elements A and B by an “AND” gate.

In Table 2.5, the main advantages and limitations of the GTST-DMLD are reported.

Table 2.5: Advantages and limitations of Goal Tree Success Tree – Dynamic Master Logic Diagram.

Goal Tree Success Tree – Dynamic Master Logic Diagram		
	Advantages	Limitations
Qualitative analysis	Comprehensive knowledge of the system in terms of functions (qualities), objects (parts) and their relationships (interactions)	
	System “decomposed” according to its goals and functions by the GT and according to its parts by the ST, allowing a good understanding of the system.	Difficult to build and manage hierarchies for large-scale systems.
	The representation clearly illustrates the dependency relations distinguishing between logical, physical and fuzzy relationships ² . The last ones allow modeling also probabilistic, linguistic and resolution uncertainty. The strength of the relationship can be included.	
	Dynamic behavior modeling.	
	Full-scale logical reasoning (i.e., cause-effect reasoning) that cannot be included in the classical logic-based system [Hu and Modarres, 1999].	

² Physical and fuzzy relationships are not included in this work, they are considered in [Hu and Modarres, 1999]

	Various representation methodologies (e.g., sensor, Markovian model, neural net) can be combined through a DMLD dependency matrix: the inputs may be connected to other models which have either quantitative or qualitative outcomes [Hu and Modarres, 1999].	
	GTST-DMLD includes a family of models: thus, for a physical behavior there is no unique DMLD model but rather varieties of DMLD models. However, they should yield approximately similar results.	
Quantitative analysis	Simulation that is much faster than a numerical simulator, since the GTST-DMLD estimation is based on logic.	Computer-aid tools are required to handle the creation and reasoning of complex DMLD [Hu and Modarres, 1999].

Further details are not given here for brevity sake: the interested reader is referred to the cited literature [Hu and Modarres, 1999; Brissaud et al., 2011].

In this thesis, we have used the GTST-DMLD to evaluate the safety and the recovery capacity of two applications within a multi-state SoS framework: i) plant external event risk assessment and ii) CIs risk analysis. The first application deals with a critical installation exposed to the risk of earthquake, briefly presented in Section 5.1 and detailed in paper IV of Part II; the second one concerns small-sized interconnected gas and electricity networks and a supervisory control and data acquisition (SCADA) system, illustrated in Section 6.1 and in paper V of Part II. The corresponding operative steps used to simulate the system behavior are given in Sections 4.3 and 4.4, respectively.

In the next Sections, the adaptations of GTST-DMLD for the proper representation of SoS with respect to the plant external event risk assessment (Section 2.5.1) and the CIs risk analysis (Section 2.5.2) are presented.

2.5.1. GTST-DMLD for the safety and physical resilience of a critical plant

When the main inputs to a critical plant stop due to an accident, safety is assured by internal barriers which provide the inputs necessary for satisfying the safety conditions. These barriers are designed to withstand postulated accidents (design basis accidents) and include multiple, independent and redundant layers of defense to compensate for potential human and mechanical failures (defense in depth) [USNRC, 2013]. Under a SoS framework, we extend the analysis to the external supports for emergency management actions and additional, redundant infrastructure systems to provide the safety-required inputs in case of failure of both the main inputs and the first (internal) barriers. In all generality, we consider also

recovery supporting elements, such as physical components (e.g., roads for access to the site) and organizational elements (e.g., technical competence of operators), that provide help in the recovery of the internal and external safety systems. On the basis of this SoS framework, we can identify three levels of safety distinguishing the internal barriers (first level), the external supports (second level) and the recovery supporting elements (third level), as illustrated in Figure 2.8.

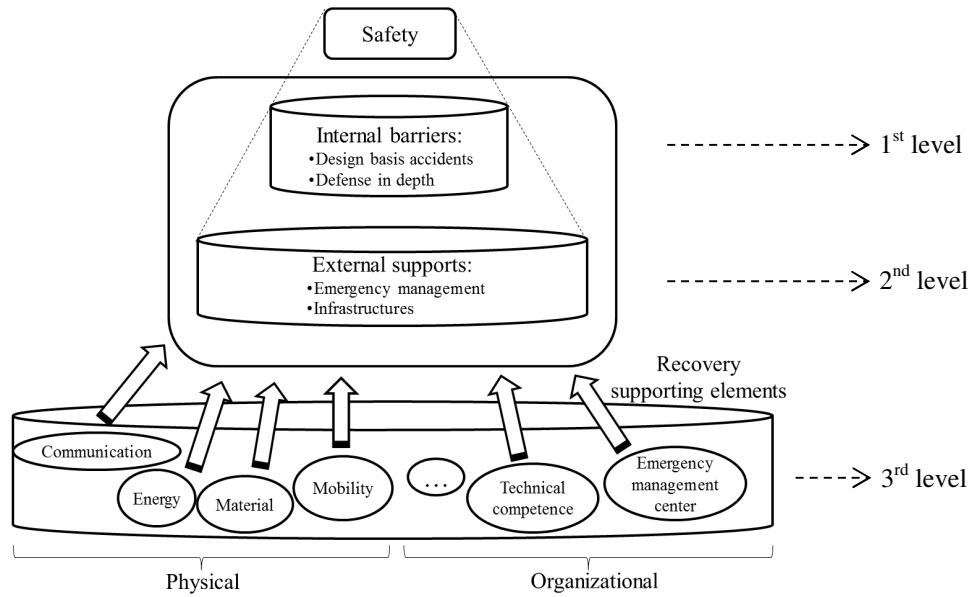


Figure 2.8: Safety levels in a system-of-systems framework considering a critical plant in emergency conditions. The first level (top) considers internal barriers; the second one (middle) extends to the external supports; the third one (bottom) accounts for the elements supporting the recovery.

In the present work, for the sake of simplicity, emergency management and organizational supporting elements are not considered. The concept of resilience is limited to the physical characteristics of the components and systems: then, we refer to physical resilience as the underlying concept. On the other hand, the GTST-DMLD illustrated in Section 2.5 can accommodate elements of fuzzy logic theory to describe imprecisely known characteristics and logic relations of non-physical facets by linguistic fuzzy terms [Hu and Modarres, 1999]. For example, specific inputs like the level of experience of the operators, can have an impact on the degree of safety of the critical plant in emergency condition: these inputs could be described in the GTST-DMLD by including threshold values [Hu and Modarres, 1999].

Figure 2.9 shows a conceptual scheme of GTST-DMLD for a SoS.

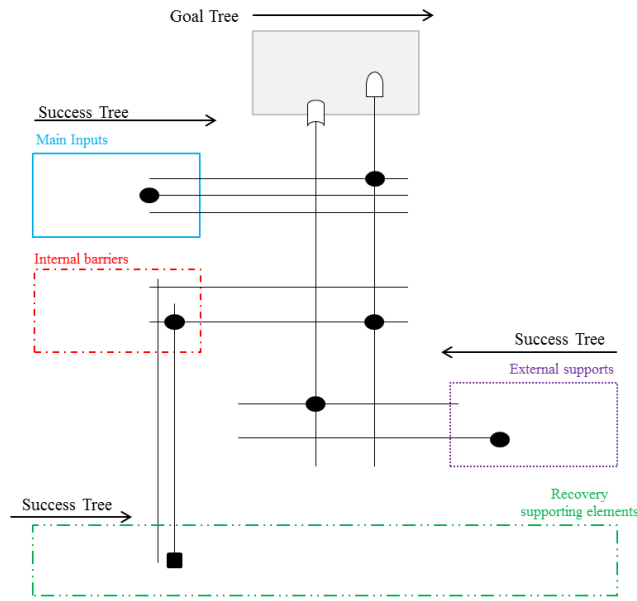


Figure 2.9: Scheme of GTST-DMLD for a system of systems.

The GT is located at the top; the ST, below the GT, is divided into three different parts to put in evidence the different role and importance of the physical elements with respect to the safety levels introduced. The main inputs and the internal barriers are placed on the top-left, the external supports on the middle-right and the recovery supporting elements on the bottom. We call the “main” and “supporting” functions/parts of the original GTST-DMLD representations as “principal” and “auxiliary” functions/parts, respectively, in order to avoid confusion with the main inputs, the external supports and the recovery supporting elements of the SoS framework.

The relationships among elements and functions are illustrated by the MLD. In particular, the connections among components of i) the main inputs, ii) the internal barriers, iii) the external supports are shown; the interdependencies between the systems i), ii), iii) are depicted; the links of the recovery supporting elements with the systems i), ii), iii) are indicated; the connections between the systems i), ii), iii) and the functions of the GT are given. Two types of dependencies have been taken into account: direct and support dependencies. The first ones, identified by a dot in the representation and called in the following “dot-dependencies”, express the need to have the element on the bottom in operation to achieve (with respect to a function) or to let working (with respect to an object) the element on the top. The support dependencies, depicted by a square and called hereafter “square-dependencies”, mean that the element on the bottom is needed for the recovery of the element on the top: its failure does not

cause the failure of the corresponding elements, but it increases the recovery time of the connected element in the case that this fails too. It acts like a delay in the repairing of the connected components. Thus, the square-dependencies are “time dependent”: when a component does not need recovery, they can be neglected; in the opposite case, they become fundamental until the complete restoration of the component: after the restoration, they can be neglected again. They are key elements of the model for the time evolution of the recovery process and they can modify (increase) the total recovery time of the component that needs to be restored.

The dynamic aspect, consisting in the functional multi-state of the components, is represented by the logic gates “AND” and “OR” that assume the same meaning as in [Hu and Modarres, 1999] for the evaluation of the state of the connected components and functions from the bottom to the top of the diagram: the minimum and the maximum values of inputs are the output values in case of “AND” and “OR” gates, respectively. In this state analysis only the dot-dependencies are considered. Differently from [Hu and Modarres, 1999], where the inputs are described by fuzzy intervals, in the present work the inputs are represented by discrete states denoted as z_j^η , $j = 1, 2, \dots, Z$, for a generic component η , $\eta \in \{1, \dots, N\}$, where N is the total number of system components, j represent the state in which the component η enters, and Z is the total number of states associated to the component η . Further details about the multi-state modeling are given in Section 3.3.1.

On the contrary, in the evaluation of the physical resilience both the dot- and square-dependencies are included and the logic gates “AND” and “OR” have an opposite meaning with respect to the state evaluation. In fact, in this case, the inputs are the recovery time values: thus, the output values of the “OR” and “AND” gates are the minimum and the maximum values of the inputs, respectively. For example, refer to Figure 2.10 where two systems $S^{(a)}$, $a = 1, 2$, contribute to the realization of the function F^* (dot-dependencies) and other two systems $S^{(a)}$, $a = 3, 4$, are relevant only to allow the recovery of the system $S^{(a)}$, $a = 2$, (square-dependencies). Assuming that $S^{(1)}$ and $S^{(4)}$ are in functional state 3, (i.e., $z_j^{S(1)}$ and $z_j^{S(4)}$, $j = 3$), with associated recovery time ($RT_{S(1)}$ and $RT_{S(4)}$) equal to 0, and $S^{(2)}$ and $S^{(3)}$ are in state 1, ($z_j^{S(2)}$ and $z_j^{S(3)}$, $j = 1$), with associated recovery times ($RT_{S(2)}$ and $RT_{S(3)}$) equal to 2 and 5, respectively, the function F^* is in state 1 ($z_j^{F^*}$, $j = 1$), since the “AND” gate (G1) means “minimum values between $z_j^{S(1)}$ and $z_j^{S(2)}$ ”. The time needed to realize the function F^* is 7 ($RT_{F^*} = 7$) since the “AND” gate (G1) means “maximum values between $RT_{S(1)}$ and $RT_{S(2)}$ ”,

where the total time needed to recover $S^{(2)}$ depends on the time to recover $S^{(2)}$ itself and the maximum value (“AND” gate G2) between $RT_{S(3)}$ and $RT_{S(4)}$. Replacing the “AND” gate G2 with an “OR” gate, the total time needed to recover $S^{(2)}$ is 2, since the minimum value between $RT_{S(3)}$ and $RT_{S(4)}$ is zero. Replacing both the “AND” gates, G1 and G2, with two “OR” gates, the function F^* is in state 3, $z_j^{F^*}$, $j = 3$, thus, it is not necessary to recover it ($RT_{F^*} = 0$).

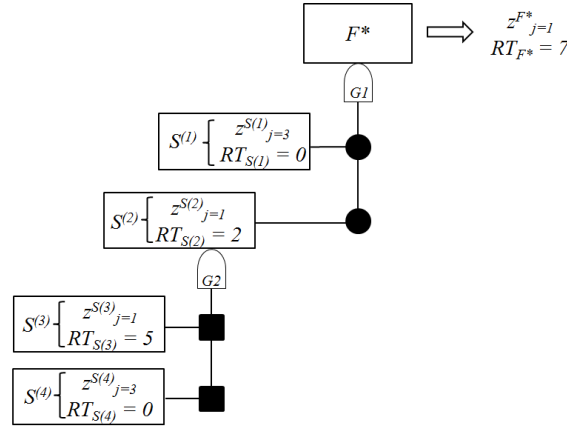


Figure 2.10: Example of the use of the “AND” logic gate together with the dot- and square- dependencies for computing the state and the recovery time of the function F^* .

In the following, we introduce a formal description of the qualities (referred to the goals and functions, i.e., the objectives) and parts (referred to the objects, i.e., the physical elements), which can be organized in hierarchies, with respect to a critical plant H whose state corresponds to the state of its critical element, E (see the previous Section 2.3). However, this description can be adapted also to the interconnected CIs, which framework is presented in the next Section 2.5.2.

The *qualities* are identified by the main goal F^* concerning the safety of H (i.e., E) that is attained by F_α , $\alpha = 1, \dots, N^*$, functions ordered in such a way that the first r directly achieve the goal F^* (i.e., they are principal – main – functions) and the last $N^* - r$ support the first ones (i.e., they are auxiliary – support – functions), as illustrated in Figure 2.11, left. The F_α , $\alpha = 1, \dots, N^*$, functions may be hierarchically divided into other functions that can be further decomposed into other ones until the required level of functional detail is reached. The last $N^* - r$ functions are represented in a parallel branch of the same hierarchy of F^* and they are connected to it by a dashed line to highlight their auxiliary role.

The *parts* are composed by A infrastructure systems $S^{(a)}$, $a = 1, \dots, A$, divided in: n^{MI} infrastructure systems of main inputs, n^{IB} internal barriers, n^{ES} external supports, n^{RS} recovery supporting elements (Figure 2.11, right). Each system $S^{(a)}$, $a = 1, \dots, A$, can be hierarchically decomposed into other systems that can be in turn divided into other ones until the desired level of detail of system components is reached. Some of the n^{MI} , n^{IB} and n^{ES} systems directly provide necessary supplies to the critical element E (i.e., they are principal – main – systems), whereas some others among them are needed for the operation of the principal systems (i.e., they are auxiliary – support – systems); to point out the different role of the last ones, they are connected to the corresponding principal systems by a dashed line (Figure 2.11, right), as for the functional hierarchy. The n^{RS} recovery supporting elements are considered apart from the other n^{MI} , n^{IB} and n^{ES} systems since they are involved in the recovery of system safety.

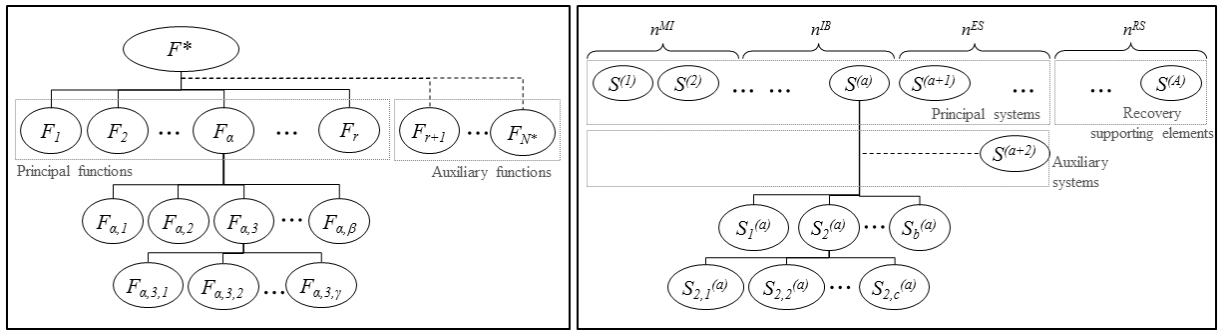


Figure 2.11: Scheme of the hierarchies of the qualities (left) and parts (right) of a generic system of systems. The auxiliary functions and parts are connected by a dashed line to the hierarchy branch that they support. The indices $\alpha, \beta, \gamma, a, b, c$ are used to indicate the systems/elements in the hierarchies; n^{MI} , n^{IB} , n^{ES} , n^{RS} refer to the number of main inputs, internal barriers, external supports and recovery supporting elements, respectively.

Notice that in a SoS view only one main function (F^*) is analyzed, whereas more than one physical systems, involved in achieving that function, are considered ($S^{(a)}$, $a = 1, \dots, A$).

For illustration purpose, let us consider the main function F^* of a critical plant H , i.e., the critical element E , achieved through the success of two principal functions, F_1 and F_2 , where the former is in turn obtained by the combination of functions $F_{1,1}$ and $F_{1,2}$. In addition, we consider an auxiliary function F_3 that is not directly needed for achieving F^* , but it serves function F_2 . In the hierarchy, function F_3 is represented in a parallel branch connected to F^* by a dashed line (Figure 2.12).

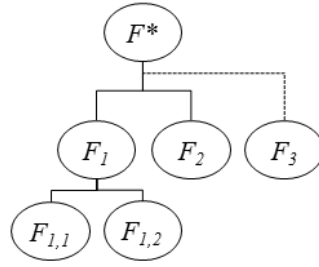


Figure 2.12: Hierarchy of the qualities for the simple example proposed.

Figure 2.13 represents the graph of the components (parts) of this example with respect to the safety levels of Figure 2.8. The links show the relationships among the components; they are directed from an element to another dependent on it. The safety of a critical element E (star) is assured by $A = 8$ systems divided into $n^{MI} = 1$ system of main inputs ($S^{(1)}$) $n^{IB} = 3$ internal barriers ($S^{(2)}$, $S^{(3)}$ and $S^{(4)}$), $n^{ES} = 2$ external supports ($S^{(5)}$ and $S^{(6)}$), $n^{RS} = 2$ recovery supporting elements ($S^{(7)}$ and $S^{(8)}$), represented in dashed oval shape. The components included in these systems are represented in solid oval shape. For example, system $S^{(1)}$ is constituted by 3 components ($S_1^{(1)}$, $S_2^{(1)}$, $S_3^{(1)}$), system $S^{(2)}$ is composed by 1 component ($S_1^{(2)}$), and so on. Notice that there are some components that are directly connected to E , e.g., $S_3^{(1)}$ and $S_1^{(2)}$, and others that are connected to the components of other systems, e.g., $S_1^{(3)}$ is connected to $S_1^{(2)}$. The former type of components belongs to principal systems, whereas the latter belong to the auxiliary systems (an exception is represented by the recovery supporting elements that are considered apart from these systems for their role of recovery, as explained above). Each system $S^{(a)}$, $a = 1, \dots, 8$, can be represented in the form of a hierarchy as illustrated in Figure 2.14.

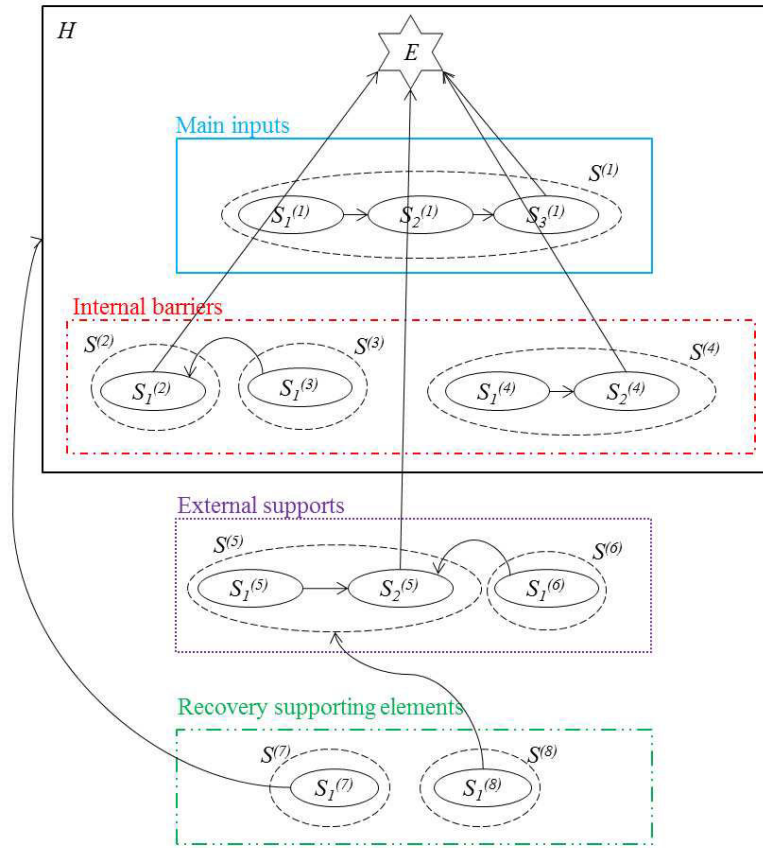


Figure 2.13: Graph of the physical components (parts) for the simple example proposed.

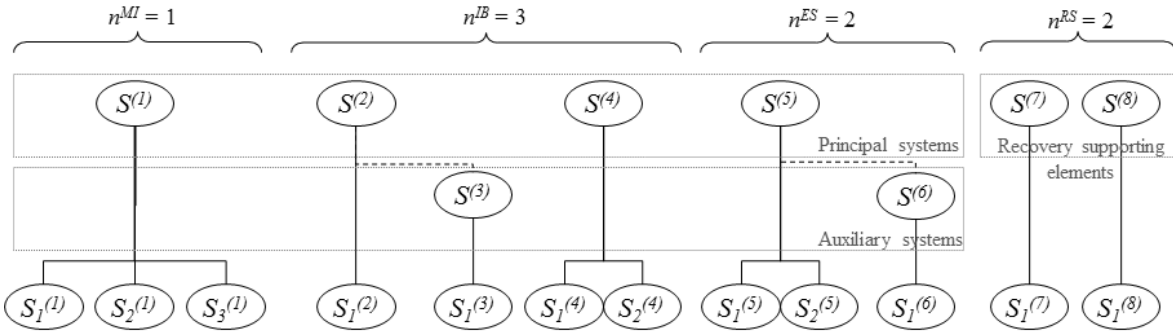


Figure 2.14: Hierarchic representation of the parts of the simple example proposed: n^{MI} , n^{IB} , n^{ES} , n^{RS} refer to the number of main inputs, internal barriers, external supports and recovery supporting elements, respectively.

In Figure 2.15, the GTST-DMLD of the example above is reported. The GT is the hierarchy of Figure 2.12 and the ST is composed by the hierarchies of Figure 2.14. The dot- and square-dependencies detail the connections of the graph of Figure 2.13 and connect the physical elements to the functions.

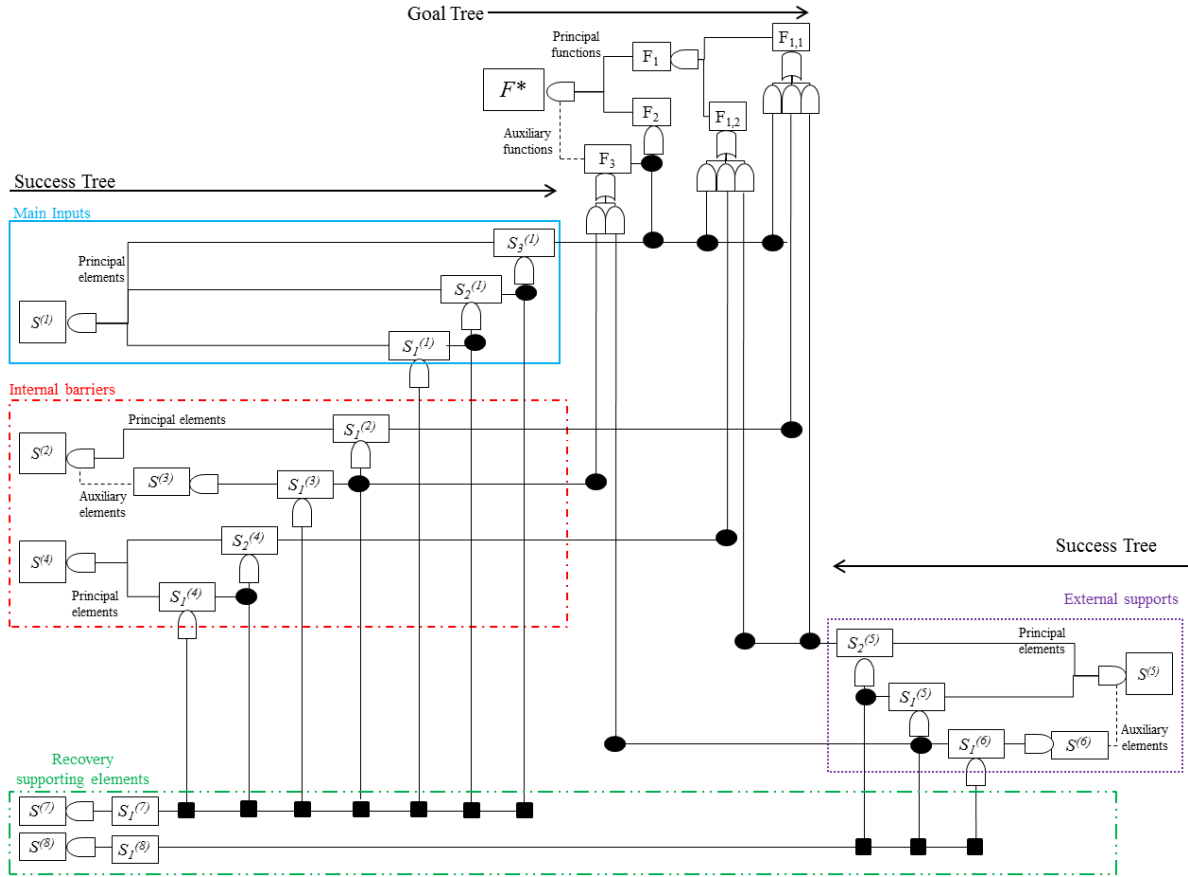


Figure 2.15: GTST-DMLD with respect to the simple example of Figure 2.13.

2.5.2. GTST-DMLD for evaluating the robustness and recovery capacity of interdependent critical infrastructures

In the context of interconnected networked CIs, the objectives of the analysis are the evaluation of the robustness of the SoS and of its recovery capacity. In particular, robustness is measured as the capability of the CI of supplying the required product to the demand nodes; instead, the recovery capacity is computed as the time needed to recover the SoS from the worst scenario to a level in which the demand nodes are satisfied. Thus, differently from the analysis of the safety of a critical plant (Section 2.5.1) where the focus is on a local point (i.e., the critical installation), in this case the focus is on several spatially distributed demands.

In this Section, we adapt the GTST-DMLD presented in Section 2.5 for the representation of interconnected networked infrastructures. In particular, we introduce new concepts in the diagram in order to highlight not only the dependency relations between the components, but also the ways in which the flows of (energy, material, etc.) are partitioned into the network on the basis of i) the importance of the demand nodes, ii) the amount of product necessary to

satisfy each demand, iii) the constraints of the arc capacities, and iv) the role of the supporting elements that are not devoted to the recovery of the main systems but that are needed to control the flows in the arcs and exchange information (e.g., the role provided by a SCADA system). In the following, the notation here adopted in the GTST-DMLD is explained, whereas its application to a case study involving two interdependent infrastructures (gas and electric power networks) and a SCADA system connected to the gas network is shown in Section 6.1.1 and in paper V of Part II.

First of all, since the objectives of the analysis are related to the quantity of product available for the demands, it is of interest analyzing the flows passing through the network; thus, the inputs of an arc are flows and the output is (generally) the sum of the flow inputs. This situation is represented by a “+” in the middle of an “AND” gate, as shown in the example of Figure 2.16 a. where the flows of arcs A and B enter into arc C.

With respect to the dependency relations, we distinguish between three main types: *direct*, *indirect* and *constraint-based* dependencies, as illustrated in Figures 2.16 and 2.17. The former, pictorially represented by dots and hereafter called "*dot-dependencies*", express the fact that the product of the element on the bottom passes straightly into the element on the top. The indirect dependencies, represented by hexagons and called hereafter "*hexagon-dependencies*", are instead needed for the optimal allocation of the product in the network: for example, they are used to describe those cases where the flow exceedance in an arc can be better partitioned into another arc that is not directly connected to it but that shares one of the inputs (see the example of Figure 2.16 b.). Finally, the constraint-based dependencies, depicted by triangles and hereafter called "*triangle-dependencies*", are employed to take into account some physical constraints posed by the problem, like the maximum flow required by a demand node.

For clarity of illustration, in Figure 2.16, examples of two types of dot- and hexagon-dependencies are given, with respect to different graph representations. Figure 2.16 a. shows the dependence of arc C on two input arcs A and B: arc C receives all the input products from A and B (e.g., if the flows in arcs A and B are 50 and 70 units, respectively, the flow in arc C is 120 units); this complete direct dependence is depicted by a *black dot*. Figures 2.16 b. and c. describe the same "physical" situation (i.e., an input arc A and two output arcs B and C), but with different relative importances of the arcs. Two different cases are illustrated. In the

first case (Figure 2.16 b.), arc B is more important than C: thus, in this situation, the flow from A supplies first arc B until its demand is satisfied, and then arc C e.g., if the flows in arc A is 100 units and both arcs B and C need 80 units, arc B will receive 80 units – demand fully satisfied – and arc C the rest, i.e., 20 units, – demand partially satisfied. In the second case (Figure 2.16 c.), arcs B and C are equally important: thus, the input flow (A) is divided into equal parts on the basis of the number of output arcs (i.e., two in this example); with respect to the numeric example above, both arcs B and C will receive 50 units – demands partially satisfied. In the case of Figure 2.16 b., the flow that enters in C is given by the difference between the entire flow from A and the flow given to B; to represent and compute this difference in the DMLD, the hexagon-dependency is adopted to correct the black dot-dependency from arc A to arc C (in fact, it is impossible that the entire flow of A enters at the same time in the arcs B and C as expressed by the black dot-dependency). The *white hexagon* assumes the value of the flow in B with a negative sign; this value is, then, summed to the initial flow of A to obtain the flux to C. The flow given to B can be the entire flow of A or a lower value depending on the constraints and arc capacity (see the following example in Figure 2.17). In the case of Figure 2.16 c., the flow from A is divided into equal parts: this condition is represented by a *grey dot*. However, this equal partition of the flow may not represent the optimal one, since some output arcs may require less flow than the one allocated according to this criterion, e.g., if the flows in arc A is 100 units and arcs B and C need 80 and 20 units, respectively, giving 50 units to both arcs is not a good allocation of the resource since B is partially satisfied and some product (i.e., 30 units) given to arc C is wasted. Thus, to optimize the repartition of the flow, hexagon-dependencies are adopted: they are directed from an output arc to all the other output arcs that share the same input. In this case, the “surplus flow” is a positive quantity and it is represented by a *grey hexagon* (to distinguish it from the “negative” white hexagon of the example in Figure 2.16 b.).

Notice that the graph representation of Figures 2.16 b. and 2.16 c. are identical; however, the partition of the flux from A is completely different in the two cases: this means that the graph representation alone cannot be used to describe the repartition of the flows in the network according to different criteria. On the contrary, the DMLD can capture and represent this aspect, which is useful in the quantitative evaluation of system performance; however, in this respect it exhibits some limitations as illustrated at the end of this Section.

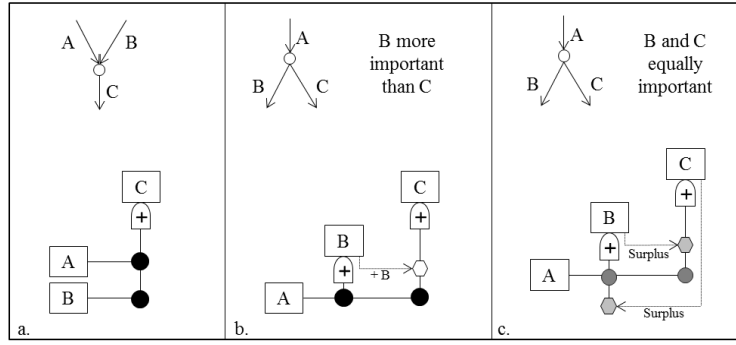


Figure 2.16: Examples of dot- and hexagon-dependencies with respect to possible graph representations.

In Figure 2.17, examples of two types of triangle-dependencies are given, with respect to different possible graph representations. Figure 2.17 a. depicts the same situation as Figure 2.16 a., with an additional arc D whose behavior impacts on the state of arc C (however, notice that D is not an input to C). This dependency is represented by a grey triangle and it means that the output of C can be modified on the basis of the state of arc D. As illustrated in Section 6.1.1, this constraint-based dependency is used in this thesis to model the SCADA system that can decrease the actual flow of the controlled arc if it is in a damage state. Figure 2.17 b. represents the same situation of Figure 2.16 c. with the addition of another arc (D) sequential to arc C. In this case, the capacity (or the demand) of arc D can limit the amount of flow in input to arc C, e.g., if the flows in arc A is 100 units, the capacity of arc C is 50 units and arcs B and D need 80 and 20 units, respectively, the repartition of the flow is as follows: first 100 units from A are equally divided into arcs B and C (50 units each) and the surplus (if there is) is partitioned into arcs B and C, then the triangle constraints is considered (i.e., arc D needs 20 units) and the new surplus is given to arc B (i.e., the exceedance of 30 units from arc C is directed to arc B). This constraint is represented in the DMLD by a black triangle and it is needed to control the input flow partitioned in different arcs and guarantee that it is not higher than necessary.

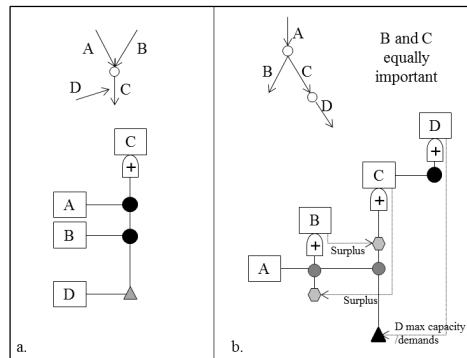


Figure 2.17: Examples of triangle-dependencies with respect to possible graph representations.

Finally, another type of constraint is taken into account, i.e., the one related to the capacity of the arcs: when the flow in input to an arc is higher than the capacity of the arc itself, the output flow will be equal to the capacity of the arc. The arc capacity can be deterministic or stochastic and in the GTST-DMLD it is represented by a grey or dot-filled rectangular, respectively, as illustrated in Figure 6.2 of Chapter 6 with respect to the application considered.

The GTST-DMLD representation here described for the evaluation of the robustness of CIs presents all the advantages already highlighted in Section 2.5 and in Table 2.5; in addition, it allows i) showing how the flow is partitioned in the CIs on the basis of different importance and priority given to the demand nodes, and ii) supporting its mathematical computation. However, the representation may become unclear due to the increasing number of hexagons, triangles and their arrows needed when i) the size of the SoS increases and ii) the importance of the demands considered is not “sequential”: “sequential importance” means that the product is given *first* to the most important demand node until it is *completely supplied*, then to the second one and so on, until the input product is over. In particular, *in this case*, the sequential importance is established *only* on the basis of the *geographical positions* of the demand nodes: those closer to the generation source are the most important; vice versa those farther are the least important. Thus, when a sequential (geographical) importance of the demands is not considered, the representation can lose its efficacy. Accounting for different criteria, e.g., equal or proportional importance, implies that the product should be partitioned into different ways according to some proportions established by the importance of the demand; this requires solving an optimization problem that cannot be tackled by the representation proposed. To overcome this limitation, we have elaborated the Hierarchical Graph described in the next Section.

2.6. Hierarchical Graph

This approach has been here introduced to analyze the robustness of interdependent CIs taking into account the fact that the demand nodes may have different importance, which establishes possibly different priorities in the partitioning of the product through the connections and elements of the CIs.

The proposed representation technique requires that the CIs of interest are first modeled by a directed graph without loops and composed by nodes and arcs; notice that the arcs may represent the components of an infrastructure or the connections between different infrastructures. We then need to distinguish between input, demand (load) and transmission arcs: the “input arcs” connect the sources of product to the network, the “demand arcs” terminate with nodes that require a given amount of product, whereas the “transmission arcs” transfer the product to other components in the network. Notice that the transmission and the demand arcs may coincide: for example, an arc may be needed to supply the connected node and in addition it may be required to transmit the product to other arcs/nodes.

In the Hierarchical Graph representation, the adjective “hierarchical” does not imply a “decomposition of the system into different level of details”, as in the previous approaches (Sections 2.4 and 2.5), but it simply means that the graph of interconnected CIs is structured in hierarchical levels. In extreme synthesis, the representation is built as follows: at the bottom of the graph, the inputs (i.e., the arcs through which the product is injected into the networks) are identified; at the top, the goals (i.e., the demand nodes that have to be satisfied) are reported; in the middle, all the other arcs (transmission and/or load arcs) that provide product to the demand nodes are organized in hierarchical levels. These levels are numbered on the basis of the number of demand nodes that are served by the corresponding arcs: the higher the number of demands supplied by an arc, the higher the hierarchical level of that arc. For example, all the arcs that are required to supply *LV* demand nodes are “placed” at hierarchical level *LV*.

Formally, let us consider A interconnected infrastructure systems $S^{(a)}$, $a = 1, \dots, A$, constituting the overall SoS, numbered in order in such a way that the first q exchange physical product (e.g., energy or material) and the last $(A - q)$ exchange information and are useful for the operation and control of the connected systems (e.g., a SCADA system). The total number of components (arcs) transmitting physical flow is referred to as N .

For illustration purposes, refer to Figure 2.18 in which the graph of a SoS (top) and its corresponding Hierarchical Graph (bottom) are reported. The SoS in the example is composed by $A = 4$ systems, where the first two, i.e., $S^{(1)}$ and $S^{(2)}$, exchange physical product (solid links in Figure 2.18, top) and the last two, i.e., $S^{(3)}$ and $S^{(4)}$, support system $S^{(1)}$ (dotted links in Figure 2.18, top). The total number of components (arcs) is $N = 8$.

As described above, the Hierarchical Graph depicts the inputs at the bottom of the representation, i.e., in this case, arc $S_I^{(1)}_S S_2^{(1)}$ in Figure 2.18 (bottom); also, it shows the goals (i.e., the demand nodes) at the top: in this case, the demand nodes are represented by all the nodes of systems $S^{(1)}$ and $S^{(2)}$, except $S_I^{(1)}$, which is the source of product. Finally, it organizes the arcs in different hierarchical levels according to the number of demand nodes they supply: for example, in this case arc $S_I^{(2)}_S S_2^{(2)}$ is at hierarchical level 4 since it provides product to four demand nodes, i.e., $S_2^{(2)}$, $S_3^{(2)}$, $S_4^{(2)}$ and $S_5^{(2)}$. The quantity of product required by the demand nodes is referred to as D_{dem} , where the subscript 'dem' is the indicator of a given demand node among the N components.

Notice that the arcs referred to the $(A - q)$ control and information systems (which are not contributing to the flow of product, but influence the state of the other arcs) do not appear in the hierarchical structure: instead, they are reported in a trapezoidal frame under the corresponding arc that they affect.

The squares located between the hierarchical levels mean that the product at that level has to be partitioned among the corresponding demand nodes.

This representation allows highlighting all the paths going from the input sources to the end nodes: for example, in Figure 2.19, the path from input $S_I^{(1)}$ to node $S_3^{(2)}$ is highlighted. In addition, the representation is able to put in evidence the critical arcs as those located at higher hierarchical levels, since their interruption or degradation affects more demand nodes: for example, in Figure 2.19 arc $S_2^{(1)}_S S_I^{(2)}$ is more critical than arc $S_2^{(1)}_S S_3^{(1)}$ since the first one is required to supply five demand nodes (i.e., $S_I^{(2)}$, $S_2^{(2)}$, $S_3^{(2)}$, $S_4^{(2)}$ and $S_5^{(2)}$), whereas the second one is necessary just for node $S_3^{(1)}$.

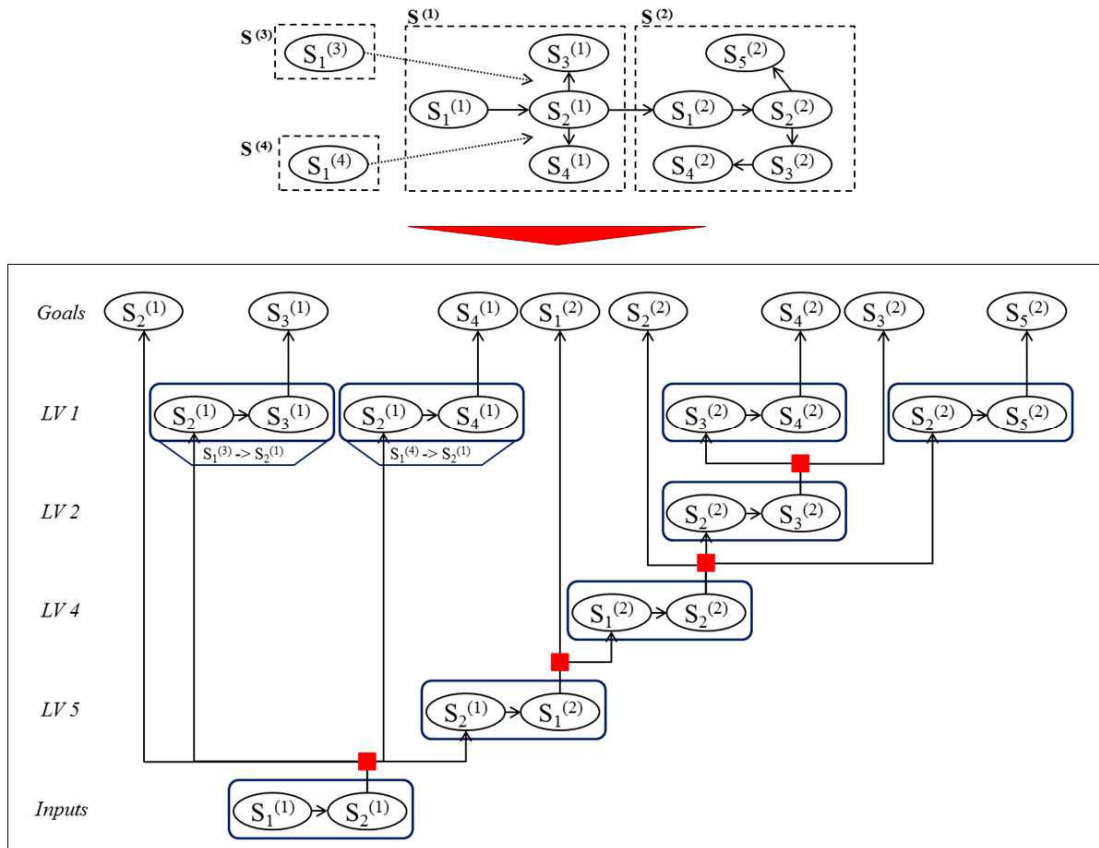


Figure 2.18: Top: graph of the components of the system of systems; the links represent the exchange of physical product (solid lines) and influence/support relationships (dotted lined). Bottom: corresponding Hierarchical Graph; LV: Level.

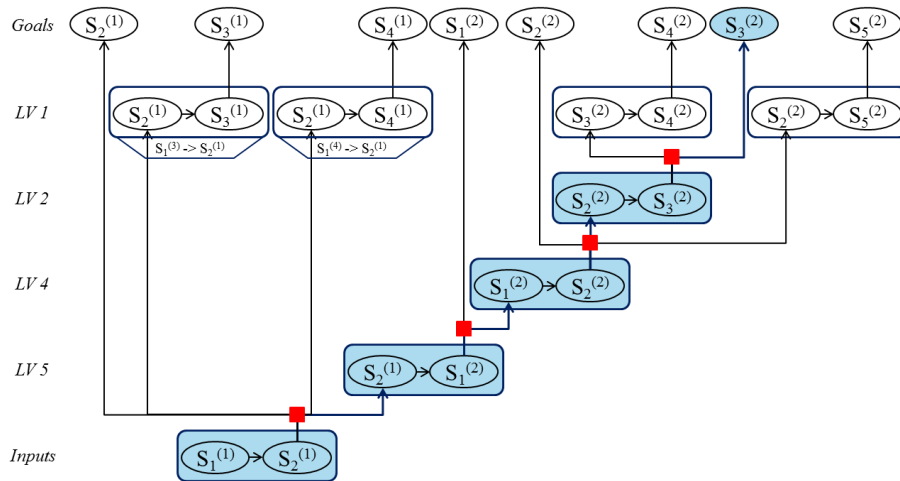


Figure 2.19: Hierarchical Graph of the system of systems in Figure 2.18, highlighting the path from the input to demand node $S_3^{(2)}$; LV: Level.

This representation has been introduced to analyze the robustness of interdependent CIs taking into account possibly different priorities in the partitioning of the product to the

demand nodes according to their importance: for example, in the case of a malfunctioning in the electrical transmission line higher importance may be given to critical buildings, such as hospitals or industries, with respect to common residential areas. However, these importance criteria are not explicitly shown in the representation, which instead is more focused in highlighting the hard, physical constraints that affect the product partitioning. With respect to the analyses carried out with the support of the Hierarchical Graph representation, three different importance criteria are considered, namely sequential (where the product is distributed sequentially on the basis of a chosen “ranking criterion”), proportional (where the product is delivered on the basis of the quantity required by the demand nodes) and equal (where the product is partitioned in equal parts). For further details the reader is referred to paper VI of Part II.

Hierarchical Graphs may be prohibitive when the size of the SoS increases, since all the origin-destination paths have to be identified. Nevertheless, they can be used in combination with clustering algorithms that reduce the systems complexity by “collapsing” many components in few clusters as illustrated in Section 4.5.2.

In Table 2.6, the main advantages and limitations of Hierarchical Graphs are reported.

Table 2.6: Advantages and limitations of Hierarchical Graph.

Hierarchical Graph		
	Advantages	Limitations
Qualitative analysis	Structured representation of the graph of a SoS considering i) inputs, ii) goals and iii) origin-destination paths.	Additional factors (operational, organizational, etc.) are not included.
	Hierarchical levels allow identifying critical arcs in the network and the arcs shared by more than one origin-destination path.	Not flexible, i.e., the addition of a new component may change the entire structure.
	Possible partition of the flow in the SoS according different importance criteria of the demand nodes.	The partition of the flow in the SoS is not clearly shown in the representation, since the algorithm used for its computation is recursive (see Section 4.5.1).
Quantitative analysis	Simulation and possible automatization of the procedure is possible by clustering analysis when the SoS size increases. This bears a reduction in the computational time of the simulation.	

In the present thesis, this representation is adopted for the robustness analysis of CIs: in particular, two case studies are considered: the first one is formed by small-sized interconnected gas and electricity networks and a SCADA system (Section 6.1); the second

one is composed by a moderately large electricity network (Section 6.2). Both are described in details in paper VI of Part II. The operative steps used to simulate the system behavior are given in Section 4.5.

2.7. Comparisons of the representation techniques adopted

Table 2.7 compares FT Analysis, Muir Web, Hierarchical Modeling, GTST-DMLD and Hierarchical Graph on the basis of a list of qualitative and quantitative characteristics.

Table 2.7: Comparison of the Fault Tree Analysis (FT), Muir Web (MW), Hierarchical Modeling (HM), Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD), Hierarchical Graph (HG).

	Description	FT	MW	HM	GTST DMLD	HG
Qualitative analysis	Easy to build		✓	✓	✓	✓
	Well-defined structure	✓				✓
	Flexible		✓			
	System broken in different levels of detail			✓	✓	
	Representation of interdependencies		✓		✓	
	Representation of the strength of the relationships		✓		✓	
	Representation of other factors (organizational/operational)		✓		✓	
	Representation of goals and functions				✓	
	Representation of the flow (energy, material, ...)				✓	✓
	Multi-state modeling				✓	✓
	Rigorous and transparent analysis	✓				
	Possible identification of critical elements	✓		✓	✓	✓
	Conversion into maps		✓			
	Need for further analyses		✓	✓		
	Large quantity of data are required		✓		✓	
Quantitative analysis	Simulation		✓	✓	✓	✓
	Numerical calculation	✓				
	High computational cost			✓		

In synthesis, the FT (Section 2.2) allows identifying critical elements and computing a quantity of interest by a rigorous and transparent analysis, although it lacks of many important characteristics to well represent the behavior of SoS: for example, the dynamic aspect, the representation of dependencies and interdependencies, the flexibility to consider new elements and organizational/operational factors, and the limitation to the Boolean logic.

The Muir Web (Section 2.3) offers a flexible and easy way of representation, with the possibility of managing a large number of nodes and relationships. In addition, extending the analysis to the level of factors (operational, organizational, etc.) that influence the physical elements, it shows the capability of crossing disciplinary boundaries in an integrated representation; then, it can interface straightforwardly with other modeling tools to generate maps representing the spatial localization of the infrastructures, including their interdependences and all related characteristics. However, this flexibility and ease of representation is paid in terms of the large amount of information needed to further characterize the model and the need of further analyses to associate the logic structure of the system and evaluate it in terms of the quantities of interest (which may require costly simulations for large systems).

Hierarchical Modeling (Section 2.4) is easy to build and provides a good understanding of the system; it allows analyzing the SoS at different level of details and identifying critical elements at each level. However, it presents limitation in the lack of flexibility, in the use of Boolean logic, in the difficult representation of dependencies and interdependencies for which it should be supported by further analysis.

The GTST-DMLD (Section 2.5) takes several advantages from the previous representations: it easy to build, it analyzes the system at different levels of detail, it allows representing different types of dependencies and interdependencies, and it allows reasoning on cause-effect relations allowing the identification of critical elements. In addition, it highlights goals and functions, it can integrate other modeling techniques, it supports multi-state modeling, it allows dynamic analysis and it can represent the way in which the flow is partitioned in the networks. However, with respect to this last point, it presents some limitations when large-sized systems are considered and importance criteria of the demand nodes are not geographically sequential.

Finally, the Hierarchical Graph (Section 2.6) is the representation elaborated in this thesis to overcome a limitation of the GTST-DMLD with respect to the possibility of partitioning the

flow in the networks according to different importance and priorities of (some of) the demand nodes. However, the method is still under development and it lacks of some advantages of the previous representations, like the consideration of goal and functions and of other factors (e.g., organizational) that influence the physical elements, the flexibility, the possibility of integrating other modeling approaches, etc.

3. SYSTEM MODELING

In Section 3.1, the issue of system modeling is introduced and an overview of the existing modeling approaches is given; then, in Sections 3.2 and 3.3, the binary and multi-state models adopted in the present thesis are described, respectively.

3.1. Overview on the existing modeling approaches

In reliability analysis two types of models are considered: *binary state* and *multi-state* models. *Binary state* models have two states, perfect functioning and complete failure, that are mutually exclusive. *Multi-state* models have been introduced to obtain more realistic and precise representations of engineering systems than binary state model [Gu and Li, 2012]. In particular, they consider several states to represent different level of performances of a system and its components. If, on one hand, they provide more accurate results than binary state models, on the other hand, they are more complex and present major difficulties in system definition (e.g., in the definition of the various states and their occurrence probabilities) and in the performance evaluation [Gu and Li, 2012; Sallak et al., 2013]. It is worth mentioning that *fuzzy multi-state* systems approach has been considered to handle uncertainties of the state probabilities and the state performance that are represented by fuzzy values [Ding et al., 2010].

A particular type of multi-state model is represented by the wellbeing analysis framework that accounts for the degree of success of any operating system state [Billinton and Karki, 1999a]. Three states are identified: healthy, marginal and at risk. In the *healthy state*, all equipment and operating constraints are within limits and there is sufficient margin such that the loss of any element (specified by some criterion) will not result in a limit being violated; in the *marginal state* the system is operating within the limits, but there is no longer sufficient margin to satisfy the acceptable criterion; and in the *risk state*, equipment constraints are violated and load may be not (or just partially) supplied. Application areas of the wellbeing analysis are represented by generating systems [Billinton and Karki, 1999a], operating reserve assessment [Billinton and Karki, 1999b] and composite generation and transmission systems [da Silva et al., 2004]. In this work, we consider the wellbeing analysis for evaluating the safety of a critical plant.

Practical methods of multi-state system reliability assessment are based on four different approaches [Sallak et al., 2013]: the structure function [Pourret et al., 1999], Monte Carlo (MC) simulation [Marseguerra and Zio, 2002; Zio et al., 2007], the Markov approach [Xue and Yang, 1995] and the Universal Generating Function method [Levitin and Lisnianski, 1999]. In this work, we adopt MC simulation (Chapter 4).

With respect to the applications within the framework of external event risk assessment (Chapter 5), we consider both binary state (Section 3.2) and multi-state (Section 3.3.1) models; in particular, this last one recalls the wellbeing analysis framework. With respect to the applications related to critical infrastructures (CIs) (Chapter 6), we adopt a multi-state model, assuming that the components are described by Markov and semi-Markov processes, which main characteristics are synthetized in Section 3.3.2.

3.2. Binary state models

In binary state models, components can be in two possible states, i.e., functioning or faulty, and, as a consequence, the system performance F^* of interest shows full success or failure at system level [Hu and Modarres, 2000].

Once that the state of each component is defined, the evaluation of the system performance can be carried out with the support of the representation techniques illustrated in Chapter 2, e.g., Fault Tree (FT), Muir Web and Hierarchical Modeling. In this thesis, we adopt a binary state model to evaluate the safety of a critical plant (e.g., a nuclear power plant) exposed to the risk of natural external events (e.g., earthquakes). In this framework, each component is characterized by a fragility function, i.e., the conditional probability to enter in a faulty state given the intensity of the disruptive event, e.g., given the ground motion level with respect to the earthquake occurrence.

3.3. Multi-state models

In this Ph. D. thesis, we consider two types of multi-state models depending on the hazardous event under analysis: natural external events or random failures. In the presence of the first type of hazardous event, we assume that we may have a component/system performance degradation due to an external event and we distinguish between structural damage and

functionality (Section 3.3.1); on the contrary, with respect to random failures, we assume that the component/system degradation occurs stochastically and structural and functional states of damage coincide (Section 3.3.2).

3.3.1. Structural damage and functionality

In the applications involving natural external events, we distinguish two aspects of the multi-state model *at component level*, i.e., structural damage and functionality; instead, we consider only functionality *at system-of-systems* (SoS) level, which is based on the structural and functional states of the constituting components.

Multi-state model at component level: structural damage and functionality

Let us denote as η , $\eta = 1, \dots, N$, a generic component of A systems, $S^{(a)}$, $a = 1, \dots, A$, where N is the total number of components. A disruptive external event can affect both the physical structure and the functional performance of the generic component η , but not necessarily with a one-to-one correspondence. For example, a road can be affected at different levels of damage by an external event: from no damage to slight (few inches), moderate (several inches) or major (few feet) settlements of the ground. When the road is slightly damaged it can still perform its function (of connection) as in normal condition because the damage is negligible: then, the functional performance associated to the structural states “no damage” and “slightly damage” is the same. On the other hand, the correspondence between structural and functional states strongly depends on their definition and on the scope of the application, e.g., in a transportation planning the function of the road can be related to the traffic flow per hour and in this case the performance may be reduced even for slight settlements of the ground due to a decreasing speed of the vehicles, leading to a one-to-one correspondence between structural and functional states.

We define as g_i^η , $i = 1, 2, \dots, G$, and z_j^η , $j = 1, 2, \dots, Z$, the structural and functional states of the generic component η , respectively, where the indices i and j are ordered such that when $i, j = 1$, the component is fully damaged and cannot perform its function (worst condition); when $i = G$ and $j = Z$, the component shows no damage and can fully perform its function (best condition). Relations exist among the structural and functional states: a structural state corresponds to one functional state but one functional state can be associated to one or more structural states (Figure 3.1).

The evaluation of the performance of the system (i.e., the safety of the critical plant) is based on the functional state of the components that in turn depends on their structural state. The analysis of the functional state could be enough for evaluating the safety of the critical plant in the case of one-to-one correspondence between structural and functional states. On the contrary, considering more structural states than functional states allows us taking into account hidden (structural) criticalities that can suddenly turn the functionality of a component into a worse state, e.g., upon occurrence of aftershocks. In fact, a same functional state can be reached from different structural states, i.e., from different degrees of damage: even if functional performance is the same, a component with worse structural state is more fragile if exposed to other external events that can further degrade it structurally and at the same time cause a reduction of its functionality. For example, with respect to Figure 3.1, it can be seen that the functional state $z_j^\eta, j = 3$, can be reached when the component η is in the structural state $g_i^\eta, i = 4, i = 5$ or $i = 6$, but in the case $i = 4$ the component is weaker to withstand subsequent stresses than in the case $i = 6$, and therefore it is more inclined to pass into a lower structural state, i.e., if the structural state is lower than 4 ($g_i^\eta, i < 4$), the functionality will be lower than 3 ($z_j^\eta, j < 3$). With respect to the example of the road above, when the road is slightly damaged it is more exposed to aftershocks than when it is not damaged.

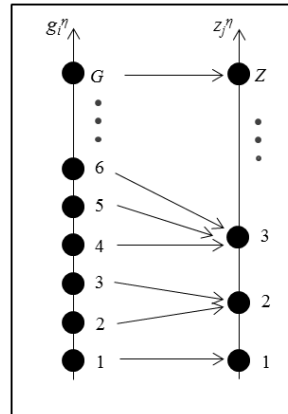


Figure 3.1: Relations between the structural, $g_i^\eta, i = 1, 2, \dots, G$, and functional $z_j^\eta, j = 1, 2, \dots, Z$, states for a component η .

In the case study exemplification of this work (Chapter 5), we consider three structural and functional states, i.e., g_i^η and z_j^η with $i, j = 1, 2, 3$. They represent risk, marginal and healthy conditions, adopting the scheme of well-being analysis [Billinton and Karki, 1999a].

Denoting as $y^{\eta,min}$ the lowest output value that it is requested by a component η to keep a safe state (it represents the risk threshold) and $y^{\eta,opt}$ the optimal output value that should be provided by the component η to keep a safe state with a safety margin, sm , ($sm = y^{\eta,opt} - y^{\eta,min}$), we define:

1. Risk state:

- Structural ($g_i^\eta, i = 1$): the component η is strongly damaged by the external event.
- Functional ($z_j^\eta, j = 1$): the component η cannot fulfill its function; its output y^η is lower than the minimal requested $y^{\eta,min}$, i.e., $y^\eta < y^{\eta,min}$.

2. Marginal state:

- Structural ($g_i^\eta, i = 2$): the component η is slightly damaged by the external event.
- Functional ($z_j^\eta, j = 2$): the component η can fulfill its function, providing an output y^η that is lower than the optimal output $y^{\eta,opt}$, but higher than the minimal requested, i.e., $y^{\eta,min} \leq y^\eta < y^{\eta,opt}$, the safety margin is not satisfied.

3. Healthy state:

- Structural ($g_i^\eta, i = 3$): the component is not damaged by the external event.
- Functional ($z_j^\eta, j = 3$): the component can fulfill its function, providing an output y^η that is equal or higher than the optimal output $y^{\eta,opt}$, i.e., $y^\eta \geq y^{\eta,opt}$.

The relations between structural and functional states depend on the scope of the application, as exemplified above, but also on the intrinsic characteristics of the components. The combinations considered for the case study of Chapter 5 are illustrated in Figure 3.2 for a generic component η . The relations among three structural and functional states (Figure 3.2.a) are typical of elements of the water system since their functional performance is associated to their flow: a reduction of the water flow due to a structural damage means a reduction of their functional performance, e.g., a leak in a pipe reduces the flow capacity. The combinations among three structural states and two functional states (Figure 3.2.b) occur when a component not damaged ($g_i^\eta, i = 3$) or slightly damaged ($g_i^\eta, i = 2$) can perform totally its function ($z_j^\eta, j = 3$), i.e., the structural damage of state 2 has no effects on the functional performance. The components characterized by these relations are, for example, the road accesses, as shown above, and the elements of the power system, e.g., the power pole that can fulfill its function

to carry the power line even if its structure presents some damage. Finally, binary components (Figure 3.2.c), present two structural and functional states: no degrees of damage are considered since also a slight damage lead a component to loose completely its functionality (e.g., in the case of a valve).

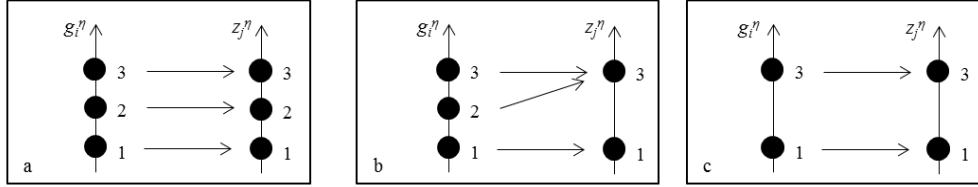


Figure 3.2: Three types of relations between the structural, g_i^η , $i = 1, 2, \dots, G$, and functional z_j^η , $j = 1, 2, \dots, Z$, states of a component η .

In this case, each component is characterized by different fragility functions that express the conditional probability of exceeding a level of damage given the intensity of the disruptive event.

Multi-state model at system-of-systems level: functionality

At SoS level, the focus of the analysis is on the degree of fulfillment of the goal function F^* (in this case, the degree of safety of the critical plant). To obtain a functional state at system-of-system level, we combine the systems $S^{(a)}$, $a = 1, \dots, A$, into K alternative (or redundant) logic paths, ξ_k^F , $k = 1, \dots, K$, that attain the same function F^* , as illustrated in Figure 3.3 for four systems, $S^{(a)}$, $a = 1, \dots, 4$.

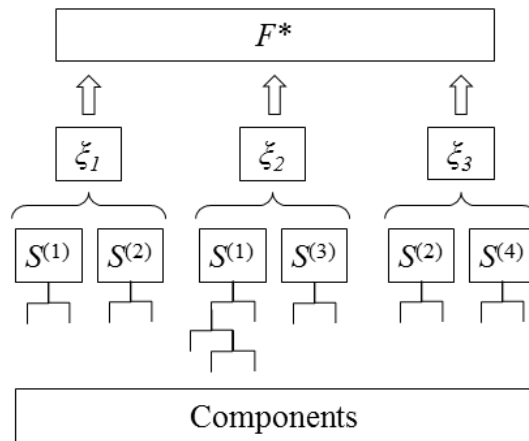


Figure 3.3: Exemplification of the combination of $S^{(a)}$, $a = 1, \dots, 4$, systems into 3 redundant logic paths ξ_k^F , $k = 1, \dots, 3$, that attain the same function F^* .

The functionality of the $S^{(a)}$, $a = 1, \dots, A$, systems is based on the functional performance and

on the structural state of the components η , $\eta = 1, \dots, N$: then, we can identify a healthy, marginal and risk state for these systems on the basis of the states of their components. The functional state of the logic paths, ξ_k^F , $k = 1, \dots, K$, is in turn obtained from the states and the reciprocal relationships of the $S^{(a)}$, $a = 1, \dots, A$, systems. Finally, the functional performance at SoS level is determined on the basis of i) how many and which logic paths, ξ_k^F , $k = 1, \dots, K$, are available and ii) their functional state. The evaluation of the function F^* is different case by case, depending on the characteristics of the SoS and on the expert judgment. In the present work, we still consider three functional states, z_j^H , $j = 1, 2, 3$, i.e., risk, marginal and healthy, respectively, for the critical plant. In all generality, we assume that both the healthy and marginal states assure the system performance (i.e., the safety of the critical plant). While the first one can provide inputs to the critical plant by different available ξ_k^F , $k = 1, \dots, K$, alternative logic paths, i.e., safety margin is satisfied, the second one can assure inputs by only one of the redundant logic paths without possibility of replacing it in case of its accidental interruptions, i.e., a safety margin is not satisfied. Further details are given in paper IV of part II with respect to the case study of interest.

3.3.2. Markov and semi-Markov processes

A finite discrete Markov chain has a finite number NS of possible states $\{0, 1, 2, \dots, NS\}$, at each step $n = 1, 2, 3, \dots$, in the process [Buckley, 2004]. Let us denote by $X(n)$ the random variable indicating the state of the system at the step n . The fundamental assumption characterizing a Markov process is that the future state of the system depends solely on its present state, thus, $p_{ij} = \text{Prob}(X(n+1) = j \mid X(n) = i)$ for $i, j \in \{0, 1, 2, \dots, NS\}$, $n = 1, 2, 3, \dots$; p_{ij} is the one-step transition probability from state i to state j , which do not depend on n . These probabilities can be arranged in a $((NS+1) \times (NS+1))$ transition probability matrix $\underline{\underline{P}}$:

$$\underline{\underline{P}} = \begin{array}{c|cccc} i/j & 0 & 1 & \dots & NS \\ \hline 0 & p_{00} & p_{01} & \dots & p_{0NS} \\ 1 & p_{10} & p_{11} & \dots & p_{1NS} \\ \dots & \dots & \dots & \dots & \dots \\ NS & p_{NS0} & p_{NS1} & \dots & p_{NSNS} \end{array}$$

with the following properties: i) $0 \leq p_{ij} \leq 1$, for each $i, j \in \{0, 1, 2, \dots, NS\}$, since all the

matrix elements are probabilities, and ii) $\sum_{j=0}^S p_{ij} = 1$, since the states are assumed exhaustive [Zio, 2009].

For ergodic systems, the steady-state probability vector, $\underline{\Pi}$, composed by the steady-state probabilities Π_i , $i = 0, 1, \dots, NS$, of the system being in state i asymptotically is determined as [Zio, 2009]:

$$\underline{\Pi} = \underline{\Pi} \cdot \underline{P} \quad (3.1)$$

A semi-Markov process can be considered as an extension of an ordinary Markov process where waiting time distributions are made explicit [Foucher et al., 2005]: for example, the holding time t_{ij} the system remains in state i before performing a transition into state j could be described as normally distributed with given mean μ_{ij} and variance σ_{ij} : $t_{ij} \sim N(\mu_{ij}, \sigma_{ij})$ [Nozick et al., 2005]. In this case, the computation of the steady-state probabilities, ξ_i , $i = 0, 1, \dots, NS$, is made by weighting eq. 3.1 with the expected holding time τ_i as [Barry and Nelson, 1995]:

$$\xi_i = \Pi_i \cdot \tau_i / \sum_{j=0}^{NS} \Pi_j \cdot \tau_j \text{ for } i = 0, 1, \dots, NS \quad (3.2)$$

where $\tau_i < \infty$ is the expected holding time in state i .

Notice that we have presented only discrete-time, discrete-state Markov processes since we have adopted only them in the applications of the present thesis. However, it is worth mentioning that continuous time, discrete-state Markov processes exist and they are fully described by the matrix of the transition rates between the states of the system [Zio, 2009a].

Finally, a family (class) of Markov process includes the piecewise deterministic Markov processes that involve deterministic motion punctuated by random jumps: the process follows a deterministic trajectory until the first jump which occurs either spontaneously in a random manner, or when the trajectory hits the boundary of the state-space, e.g., when a physical parameter reaches a critical value so that the mode of operation changes [Zhang et al., 2008; Azaïs et al., 2014].

4. SYSTEM SIMULATION AND UNCERTAINTY PROPAGATION

In this Chapter, the simulation and uncertainty propagation methods considered in the present Ph. D. work are illustrated. First, in Section 4.1 a brief overview of the existing techniques is given. Then, the operative steps applied to evaluate the system-of-systems (SoS) behavior are detailed. In particular, the Monte Carlo (MC) simulation for Seismic Probabilistic Risk Assessment (SPRA) considering binary (in Section 4.2) and multi-state (in Section 4.3) SoS frameworks is described with respect to the SoS representations adopted, i.e., Fault Tree (FT), Muir Web, Hierarchical Modeling and Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) (Sections 2.2 – 2.5). In Section 4.4, the procedure of MC simulation combined with interval analysis and considering the GTST-DMLD representation is shown. Finally, in Section 4.5 the operative steps of MC simulation adopting the Hierarchical Graph representation (Section 2.6) are given.

4.1. Overview on the existing simulation and uncertainty propagation techniques

The choice of simulation and uncertainty propagation methods depends on the nature and representation of the input variables. When the input variables are considered probabilistic and their uncertainty is represented by probability distributions, a purely probabilistic approach is considered. This is based on the MC sampling of possible values of all the input variables from the corresponding probability distributions and the subsequent computation of the model output in correspondence of the input values sampled [Kalos and Whitlock, 1986; Marseguerra and Zio, 2002]. This procedure is repeated a large number of times to collect different values of the model output in correspondence of different values of the input variables. These random realizations can be used to calculate quantities of interest, e.g., the empirical cumulative distribution function (CDF) of the model output.

When, instead, not all the input variables are probabilistic, e.g., those associated with epistemic uncertainty that can be described by fuzzy set theory, evidence theory, possibility theory and interval analysis (see Section 1.6), the so called "hybrid" approach is adopted that consists in the joint propagation of aleatory and epistemic uncertainty by combining the MC technique [Kalos and Whitlock, 1986] with the other alternative theories like for example

interval analysis (considered in this work) by means of the following two main steps:

- i. repeated MC sampling of the random variables to process aleatory uncertainty;
- ii. interval analysis to process epistemic uncertainty.

The main drawback of MCS is that it requires considerable and often prohibitive computational efforts. The reason is twofold. First, a large number of MC evaluations must generally be carried out for an accurate uncertainty propagation and functional failure probability estimation (the number of simulations required to obtain a given accuracy depends on the magnitude of the failure probability to be estimated: the lower is the functional failure probability, the higher is the number of simulations needed) [Schueller, 2007]. Second, long calculations are typically necessary for each run of a detailed model (one code run is required for each sample of values drawn from the uncertainty distributions) [Pourgol-Mohamad et al., 2010].

Thus, efficient simulation techniques have been sought to perform robust functional failure probability estimations and uncertainty propagation while reducing as much as possible the number of the code simulations and the associated computational time.

Two conceptual definitions of computational efficiency are considered in literature [Zio and Pedroni, 2011]: efficient MC simulation techniques to performing robust estimations based on a limited number of samples drawn (e.g., Line Sampling and Subset Simulation) and fast running, surrogate regression models (metamodels) in replacement of the long-running model codes.

4.2. Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a binary system-of-systems framework

Within a binary SoS analysis framework, we wish to evaluate the safety of a critical plant H (a nuclear power plant) exposed to the risk from earthquakes occurrence, accounting not only for the direct effects of the earthquake on H , but also for the structural and functional responses of the connected systems inside and outside the plant by considering the underlying dependency structure. To do this, we represent the SoS by means of one of the methods illustrated in Sections 2.2 – 2.4 (i.e., FT, Muir Web and Hierarchical Modeling) and we adopt MC simulation to carry out a quantitative SPRA [Huang et al., 2011]; the SPRA steps are not reported here for brevity: the reader is referred to papers I – IV for details. In addition, we

wish to determine the recovery capacity of the SoS, evaluating the period necessary to restore the safety of the critical plant; this is done by adopting the Hierarchical Modeling (Section 2.4).

In the following, we illustrate the operative steps of the simulation methods, differentiating between the procedure adopted in the presence of the FT and Muir Web representations (Section 4.2.1) and the one adopted in the presence of the Hierarchical Modeling framework (Section 4.2.2); the reader is referred to Section 2.3 (Muir Web) and Section 2.4 (Hierarchical Modeling), respectively, for the notation employed.

4.2.1. Operative simulation steps considering Fault Tree or Muir Web system-of-systems representations

The simulation procedure consists of the following operative steps:

1. Represent the system; if the representation chosen is the FT, then build a FT considering as top event “unsafe state of the critical plant H” and identify the minimal cut sets M_1, M_2, \dots, M_{mcs} ; otherwise, build the Muir Web;
2. sample an earthquake magnitude value from its probability distribution (i.e., double truncated exponential distribution, for details the reader is referred to equation 1 in paper II of Part II);
3. compute the ground motion value, e.g., the peak ground acceleration, at each of the $N_{c_i}, i = 1, \dots, N_S$, components of the systems $S_i, i = 1, \dots, N_S$, (i.e., at each of the primary failures with respect to the FT) by the ground motion attenuation relationship that measures the decrease in severity (or amplitude) of ground shaking with increasing distance from the earthquake source (equation 2 in paper II of Part II). The peak ground acceleration is a parameter representing the maximum value of acceleration displayed on an accelerogram [USNRC, 1997];
4. compute the fragility, f , i.e., the conditional probability of failure for any given ground motion level, for all the components $N_{c_i}, i = 1, \dots, N_S$, of the systems $S_i, i = 1, \dots, N_S$, (i.e., for all the primary failures with respect to the FT) by the fragility model (equation 3 in paper II of Part II); f is a vector of N values corresponding to the N components of the system;

5. sample a matrix $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N$, where N_T is the number of simulations, of uniform random numbers in $[0,1)$;
6. determine the fault state matrix $\{g_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N$, by comparing the fragility, f , with the matrix $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N$: if $u_{j,k} < f_k$, $g_{j,k} = 1$; otherwise $g_{j,k} = 0$ for $j = 1, \dots, N_T$ and $k = 1, \dots, N$. When $\{g_{j,k}\}$ assumes value 1, the k -th component is affected by the earthquake, i.e., it enters a faulty state; otherwise, it survives. Each row of the matrix g represents the states of the N system components;
7. determine the state of the critical plant H , considering:
 - a. the impact of the earthquake on H , i.e., taking into account the interconnected systems S_i , $i = 1, \dots, q$, inside the plant. In the present work, this has been performed by considering just the Muir Web.
 - b. the impact of the earthquake both on H , i.e., taking into account the interconnected systems S_i , $i = 1, \dots, q$, inside the plant, and on the interconnected systems S_i , $i = q + 1, \dots, N_S$, outside the plant. In the present work, this has been performed by considering both the FT and the Muir Web.

The state of H is identified by the analysis of the states of the N_{in} components of the systems S_i , $i = 1, \dots, q$, for the case a., and of the N components of the systems S_i , $i = 1, \dots, N_S$, for the case b., together with the analysis of the dependence of H from the services provided by the systems, as represented in the FT and Muir Web model.

In particular, the state of H (in case b.) is assessed for each system configuration sampled at step 6, by evaluating the system structure function $h_j = \Phi(g_{j,1}, \dots, g_{j,N}) = 1 - (1 - M_1)(1 - M_2) \dots (1 - M_{mcs})$, $j = 1, \dots, N_T$. A vector $\{h_j\}$, is thus obtained, whose elements assume value 1 when the critical plant H is in an unsafe state and 0 otherwise.

Adopting the Muir Web, instead, a further analysis is needed to identify the functional logic relations among the components within each system (intra-system links) and among different systems (inter-system links). Knowledge of these

relations allows identifying the state of the critical plant H on the basis of the states of the components of its connected systems and their logic links: trivially, if two components of a system are connected in series (Figure 4.1, left), they should be both in an operational state to guarantee its functioning; on the contrary, if they are connected in parallel (Figure 4.1, right), at least one of them should work.



Figure 4.1: Example of series (left) and parallel (right) configurations between two components.

The state of H is then evaluated through the analysis of the logic connections between the components, as explained above, for each row of the matrix $\{g_{j,k}\}$, i.e., for all the k states determined at step 6, where $k = 1, \dots, N_{in}$ and $k = 1, \dots, N$ for the case a. and b. above, respectively, and for all the simulations j , $j = 1, \dots, N_T$. A vector $\{h_j\}$, $j = 1, \dots, N_T$, is then recorded, whose element assumes value 1 when the critical plant H is in an unsafe state and 0 otherwise;

8. estimate the probability of the critical plant H of being unsafe by computing the sample average of the values of the vector $\{h_j\}$, $j = 1, \dots, N_T$.

The procedure above is repeated a large number of times for different values of earthquake magnitude.

4.2.2. Operative simulation steps considering Hierarchical Modeling system-of-systems representation

The procedure is performed for a given/chosen magnitude value and not for a sampled value as before.

The simulation consists in the following operative steps:

1. choose a value of magnitude with respect to which the analysis is performed;
2. compute the ground acceleration value at each of the $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, $L = N_L$, elements of the SoS, by equation 4; $N_S^{(N_L)}$ is the number of elements at the last level of the hierarchy, i.e., in our case, the number of individual components;

3. compute the fragility, f , for all the components $S_i^{(N_L)}$, $i = 1, \dots, N_S^{(N_L)}$, of the SoS by equation 6; f is a vector of $N_S^{(N_L)}$ values, one for each individual component in the system;
4. sample a matrix of uniform random numbers in $[0,1)$ $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N_S^{(N_L)}$, where N_T is the number of simulations;
5. determine the fault state matrix $\{g_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N_S^{(N_L)}$, by comparing the fragility, f , with the matrix $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N_S^{(N_L)}$: if $u_{j,k} < f_k$, set $g_{j,k} = 1$; otherwise set $g_{j,k} = 0$ for $j = 1, \dots, N_T$ and $k = 1, \dots, N_S^{(N_L)}$. When $g_{j,k}$ assumes value 1, it means that in the j -th simulation the k -th component is hit by the earthquake, i.e., it enters a faulty state; otherwise, it survives. Each row of the matrix g represents the states of the $N_S^{(N_L)}$ system components in the j -th simulation;
6. determine the state of the critical plant H . This is done by propagating bottom-up through the hierarchy the faulty states of the components: the states of the $S_i^{(N_L)}$ components and the state matrix at the level $N_L - 1$ of the hierarchy are used to determine the states of the $S_i^{(N_L-1)}$ systems at the upper hierarchical level, $L = N_L - 1$, and the evaluation is repeated for the states of the systems of the level $N_L - 2$ and so on until the top level of the hierarchy, $L = 1$.
In doing so, the state of H is evaluated for each row of the matrix $\{g_{j,k}\}$, i.e., for each configuration of the system sampled. A vector $\{h_j\}$ is then recorded, whose element h_j , $j = 1, \dots, N_T$, assumes value 1 when the critical plant H is in an unsafe state and 0 otherwise;
7. estimate the probability of the critical plant H of being unsafe by computing the sample average of the values of the elements of the N_T -dimensional vector $\{h_j\}$, $j = 1, \dots, N_T$.
8. for each configuration of the system sampled that turns the critical plant H in an unsafe state, evaluate the recovery time (RT) by the following steps:
 - a. sample a matrix $\{R_{T_r,k}\}$, $r = 1, \dots, N_{R_T}$, $k = 1, \dots, N_S^{(N_L)}$, where N_{R_T} is the number of recovery time simulations of the $S_i^{(N_L)}$, $i = 1, \dots, N_S^{(N_L)}$, elements of

the SoS that are in a faulty state; for each element the sampling is done from the respective recovery time distribution;

- b. determine the recovery time of the critical plant H , computing the recovery times at each hierarchical level accounting for the configurations of the systems $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, $L = N_L, \dots, 1$, from bottom to top of the hierarchy as shown in Section 2.4.

Notice that it is assumed that infinite resources (e.g., repair teams and material) are available for the restoration process so that the recovery can be performed at the same time on all components in need. This assumption is made considering that in emergency situations all the possible means, resources and actions are deployed to keep or restore the critical plant safety. In any case, extension to the situation of limited resources does not pose significant difficulties in both the modeling and its quantification.

Finally, the components are considered with binary states: fully operative or completely damaged; also the critical plant can assume only two states: fully operative or totally failed (Section 3.2).

4.3. Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a multi-state system-of-systems framework

Within a multi-state SoS analysis framework, we wish to evaluate the safety of a critical plant H (a nuclear power plant) exposed to the risk from earthquakes and aftershocks occurrence (see Appendix B of paper IV of Part II), accounting for the structural and functional responses of the systems inside and outside the plant, i.e., main inputs, internal barriers, external supports and recovery supporting elements. In addition, we wish to determine the physical resilience of the SoS, evaluated in terms of the time of recovery of safety states 2 and 3 (marginal and healthy, respectively, see Section 3.3.1) of the critical plant. To do this, we adopt the GTST-DMLD representation of the SoS and MC simulation for the quantitative SPRA [Huang et al., 2011]. For brevity sake, we do not recall here the SPRA steps described in papers I – IV of Part II. The simulation procedure is hereafter illustrated with respect to the notation introduced in Sections 2.5.1 (GTST-DMLD) and 3.3.1 (multi-state model):

1. choose a value of earthquake magnitude and epicenter coordinates with respect to which the analysis is performed;
2. compute by the ground motion attenuation relationship (eq. B.3 in Appendix B of paper IV of Part II) the ground acceleration value at each of the η , $\eta = 1, \dots, N$, components in the last levels of the physical hierarchies of the systems $S^{(a)}$, $a = 1, \dots, A$; N is the total number of components of the SoS³;
3. compute the fragilities, $\{f\}$, for all the components of the SoS by the fragility model (eq. B.4 in Appendix B of paper IV of Part II); $\{f\}$ is a matrix of $2 \times N$ values (two for each component), representing the conditional probability of exceeding a marginal ($f_{1,\eta}$, $\eta = 1, \dots, N$) and risk ($f_{2,\eta}$, $\eta = 1, \dots, N$) threshold;
4. sample a matrix of uniform random numbers in $[0,1)$ $\{u_v^\eta\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, N$, where N_T is the number of simulations;
5. determine the structural multi-state matrix $\{g_{j,v}^\eta\}$, $j \in \{1, 2, 3\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, N$, where j represents the structural state index, by comparing the matrix $\{u_v^\eta\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, N$ with the fragility $\{f\}$: if $u_v^\eta > f_{1,\eta}$, set $\{g_{j,v}^\eta: j = 3\}$; if $f_{2,\eta} < u_v^\eta < f_{1,\eta}$ set $\{g_{j,v}^\eta: j = 2\}$; otherwise if $u_v^\eta < f_{2,\eta}$, set $\{g_{j,v}^\eta: j = 1\}$ for $v = 1, \dots, N_T$ and $\eta = 1, \dots, N$. When $\{g_{j,v}^\eta: j = 1\}$, it means that in the v -th simulation the η -th component is strongly hit by the earthquake, i.e., it enters in a risk state; when $\{g_{j,v}^\eta: j = 2\}$, it means that in the v -th simulation the η -th component is slightly hit by the earthquake, i.e., it enters in a marginal state; otherwise, when $\{g_{j,v}^\eta: j = 3\}$, in the v -th simulation the η -th component survives the earthquake, i.e., it remains in a healthy structural state. Each row of the matrix g represents the states of the N system components in the v -th simulation;
6. determine the functional multi-state matrix $\{z_{i,v}^\eta\}$, $i \in \{1, 2, 3\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, N$, where i represents the functional state index, on the basis of the relationships between the structural and functional states of component η ;
7. determine the state of the critical plant H by propagating through the GTST-DMLD the functional states at component level to the functional states at SoS level. In doing so, the state of H is evaluated for each row of the matrix $\{g_{j,v}^\eta\}$, $j \in \{1, 2, 3\}$, $v = 1, \dots, N_T$.

³ Notice that the total number of components is referred here to “ N ” instead of “ L ” as in paper IV, for the sake of coherence with respect to the external event risk assessment applications and representations in this part I.

- ..., N_T , $\eta = 1, \dots, N$, i.e., for each configuration of the system sampled. A vector $\{h_v\}$ is then recorded, whose element h_v , $v = 1, \dots, N_T$, assumes value 1, 2 or 3 when the critical plant H is in a risk, marginal or healthy state, respectively;
8. estimate the probability of the critical plant H of being in a risk, marginal or healthy state by computing the sample average of the values of the elements of the N_T –dimensional vector $\{h_v\}$, $v = 1, \dots, N_T$;
 9. for each v -th simulation of the system sampled that turns the critical plant H in an unsafe or marginal state, evaluate the recovery time (RT_H) by the following steps:
 - a. set the current time, t^{curr} , equal to zero in correspondence of the earthquake occurrence and initialize the counter q equal to 1;
 - b. initialize the vectors of the time, t^H , and the functional state, z^H_i , of the critical plant H as $t^H(q) = t^{curr}$ and $\{z^H_i(q): i = h_v\}$, respectively;
 - c. compute the number of aftershocks, n_{max}^{af} , that will occur with a magnitude higher than a given threshold, m_{min}^{af} , and lower than the maximum possible m_{max}^{af} (eq. B.5 in Appendix B of paper IV of Part II) by the Gutenberg-Richter law (eq. B.1 in Appendix B of paper IV of Part II); sample their magnitude, m^{af} (eq. B.2 in Appendix B of paper IV of Part II) and their time of occurrence from the CDF of Figure B.1 (Appendix B of paper IV of Part II);
 - d. sample a vector RT_η , $\eta = 1, \dots, N$, of recovery times of the components that are in state 1 or 2, from the respective probability density functions (PDFs) and set to infinity (i.e., to a very large value) the recovery time of the components in state 3. If component $\eta = 1, \dots, N$, is in state 1, then it can reach both state 2 and state 3. In this case, sample the two recovery times and choose the lower. Save then a vector $g^{next}_j, j \in \{1, 2, 3\}$, $\eta = 1, \dots, N$, of structural states in which the components will enter if the recovery is carried out.
 - e. while the critical plant H does not turn into a healthy state $\{z^H_i(q): i = 3\}$, perform the following steps:
 - i. evaluate the vector RT^{sum}_η , $\eta = 1, \dots, N$, that is equal to RT_η , $\eta = 1, \dots, N$, when the functional state of the recovery supporting elements (e.g., road accesses) to component η in state 1 and 2 is in a state 3, i.e., the recovery supporting elements are available; on the contrary, it is the

sum of the recovery times of the recovery supporting elements and of the component, when the recovery supporting elements are not available;

- ii. identify the minimum recovery time, RT^{min} , of the vector RT^{sum}_η , $\eta = 1, \dots, N$;
- iii. evaluate if aftershocks have occurred in the interval $t^{int} = [t^{curr}, t^{curr} + RT^{min}]$. If no, go to the following step iv.; otherwise, go to step v.;
- iv. update the structural state vector $g^\eta_j, j \in \{1, 2, 3\}, \eta = 1, \dots, N$, for the component η that has performed the transition with the corresponding index j of the vector $g^{next}_j, j \in \{1, 2, 3\}, \eta = 1, \dots, N$. If the component η enters in a state 2, sample a new recovery time for η and update that value in the vector RT_η . For all other components, reduce the recovery time of the quantity equal to RT^{min} since the recovery of all the components proceeds at the same time. Then, update the functional state vector $\{z^\eta_i\}, i \in \{1, 2, 3\}, \eta = 1, \dots, N$, and evaluate the state of the critical plant H as in step 7., identifying the value $h^{new}, h^{new} \in \{1, 2, 3\}$. Set $q = q+1, t^H(q) = RT^{min}$ and $\{z^H_i(q): i = h^{new}\}$; Return to step e.
- v. consider the first aftershock that occurs in the interval t^{int} and evaluate its impact on the structural states of the components $\eta, \eta = 1, \dots, N$, by steps 4. and 5. for the first row of the matrix u , i.e., for one simulation;
 - if the aftershock changes the state of one or more components, consider the new vectors of structural and functional state, $\{g^\eta_{j,v}\}$ and $\{z^\eta_{i,v}\}$, respectively, and update the vector RT_η , sampling the recovery time of the components η that have changed structural state. Update the vector $g^{next}_j, j \in \{1, 2, 3\}, \eta = 1, \dots, N$, with the new structural state in which the components will enter if their recovery is carried out. Set $q = q+1, t^H(q) = t^{af} - t^{curr}$ and set $t^{curr} = t^{af}$. Return to step e.i.;
 - otherwise, perform again step e.v., evaluating the impact of the following aftershock that occurs in the interval t^{int} ; if there are no other aftershocks in the interval t^{int} , the recovery of the component

- η associated with the minimum recovery time RT^{min} (step e.ii.) is carried out. Return to step e.iv.;
- f. if the critical plant H was in state 1 ($h_v = 1$), save the time needed to recover the safety from state 1 to state 2 ($RT_{H(1)}^{(1 \rightarrow 2)}$), from state 2 to state 3 ($RT_{H(1)}^{(2 \rightarrow 3)}$) and from state 1 to state 3 ($RT_{H(1)}^{(1 \rightarrow 3)}$); if the critical plant H was in state 2, save the time needed to recover the safety from state 2 to state 3 ($RT_{H(2)}^{(2 \rightarrow 3)}$);
 - g. repeat the steps 9.a. – 9.g. N_{RT} number of times (e.g., $N_{RT} = 4000$);
10. save the recovery time for all the configurations from states 1 and 2, and obtain the empirical PDFs and corresponding CDFs.

4.4. Monte Carlo simulation and interval analysis within a multi-state system-of-systems framework

We consider a multi-state SoS analysis framework where the components state transition probabilities and the mean of the state holding time distributions are affected by epistemic uncertainties and are represented by intervals. In this context, we wish to evaluate the performance of the system illustrated in Section 6.1 and consisting of interdependent gas and electricity networks and a SCADA system; the evaluation is made in terms of i) *robustness*, measured by the steady-state probability distributions of the product delivered at the demand nodes (Section 4.4.1) and ii) *recovery capacity*, measured by the time needed to recover the SoS from the worst scenario (Section 4.4.2). The assessment is carried out by adopting the GTST-DMLD presented in Section 2.5.2.

In presence of epistemic uncertainty, the transition probability matrix \underline{P}_{comp} ($comp = 1, \dots, N$, where N is the total number of system components and the subscript “*comp*” indicates the component of interest) of Markov and semi-Markov processes (see Section 3.3.2), is composed by probability intervals $[\underline{p}_{ij}, \bar{p}_{ij}]$, $i, j = 1, \dots, NS_{comp}$, instead of fixed constant values, where NS_{comp} the number of states of the component *comp*. The corresponding component steady-state probabilities are also affected by epistemic uncertainty and represented by intervals of possible values, $[\Pi_{min}^{comp,i}, \Pi_{max}^{comp,i}]$, $i = 1, \dots, NS_{comp}$. As a consequence, a set of CDFs corresponding to the set of possible steady-state probabilities

within the intervals $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, \dots, NS_{comp}$, is obtained for each demand node. For the same reason, i.e., for the presence of the epistemic uncertainty in the state transition probabilities and also in the mean of the holding time distributions, a set of CDFs corresponding to the set of possible state transition probabilities is obtained for the evaluation of the recovery capacity.

4.4.1. Operative simulation steps considering GTST-DMLD system-of-systems representation: evaluation of robustness

To compute the steady-state probability distributions of the product delivered at the demand nodes the following three main steps are carried out:

1. Processing the epistemic uncertainties by interval analysis: this step leads to the evaluation of the intervals of the steady-state probabilities, $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, 2, \dots, NS_{comp}$, for the states of each component ($comp = 1, 2, \dots, N$) of the SoS.
2. Evaluation of the SoS performance (i.e, robustness) by MC simulation: this step leads to the determination of a set of CDFs of the product delivered at each demand node at steady state, one for each possible combination of steady-state probabilities ranging within the intervals $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, 2, \dots, NS_{comp}$, (found at step 1. above).
3. Post-processing the results obtained at the previous step 2: this step leads to the identification of two extreme upper and lower CDFs that bound the set of CDFs produced at step 2. above.

In more details:

1. Solve the following optimization problems for the lower (resp., upper) bounds $\Pi_{\min}^{comp,i}$ (resp., $\Pi_{\max}^{comp,i}$), $i = 1, 2, \dots, NS_{comp}$, for all the N components of the SoS:

$$\begin{aligned} \Pi_{\min}^{comp,i} &= \min_{p_{ij}, i, j=1,2,\dots,NS_{comp}} \{\Pi^{comp,i}\}, \forall i = 1, 2, \dots, NS_{comp}, comp = 1, 2, \dots, N \\ \Pi_{\max}^{comp,i} &= \max_{p_{ij}, i, j=1,2,\dots,NS_{comp}} \{\Pi^{comp,i}\}, \forall i = 1, 2, \dots, NS_{comp}, comp = 1, 2, \dots, N \end{aligned} \quad (4.1)$$

such that:

$$p_{ij} \in [\underline{p}_{ij}, \bar{p}_{ij}] \quad (4.2)$$

$$\sum_{j=1}^{NS_{comp}} p_{ij} = 1 \quad (4.3)$$

$$\underline{\Pi}^{comp} = \underline{\Pi}^{comp} \cdot \underline{P}_{comp} \quad (4.4)$$

The constraint of eq. (4.2) means that the transition probability from state i to state j is not known precisely and can take values in the interval of probabilities $[\underline{p}_{ij}, \bar{p}_{ij}]$ [Buckley, 2004]; the constraint of eq. (4.3) refers to a fundamental property of Markov and semi-Markov processes, i.e., that the states for each component are assumed exhaustive [Zio, 2009a]; finally, eq. (4.4) reports the definition of steady-state probability for a Markov process [Zio, 2009a]. In the case of a semi-Markov process, eq. (4.4) is weighted by the expected time of residence, τ^i , in a given state, i , before performing a transition [Barry, 1995]: $\xi^{comp,i} = \Pi^{comp,i} \cdot \tau^i / \sum_{j=1}^{NS_{comp}} \Pi^{comp,j} \cdot \tau^j$ for $i = 1, \dots, NS_{comp}$.

Notice that the optimization problems (4.1) can be solved by performing an exhaustive greedy search within the probability intervals $[\underline{p}_{ij}, \bar{p}_{ij}]$, if the dimensions of the corresponding transition probability matrices are relatively small (e.g., below 4 x 4), otherwise, alternative intelligent techniques should be sought, e.g., meta-heuristic methods like Genetic Algorithms (GAs) [Buckley, 2004]. In this work, we resort to GAs for arcs a_b, b_c, c_d, d_e (whose transition probability matrices are 7 x 7), whereas we perform an exhaustive search for all the other arcs.

2. Identify the CDFs of the product delivered at each demand node at steady state for all the possible combinations of components steady-state probabilities found at step 1. above:

- a. For each component $comp$, let the steady-state probabilities, $\Pi^{comp,i}$, $i = 1, 2, \dots, NS_{comp}$, range within the corresponding interval $[\Pi_{min}^{comp,i}, \Pi_{max}^{comp,i}]$, $i = 1, 2, \dots, NS_{comp}$, to obtain a set of Q_{comp} vectors of steady-state probabilities, $\{\underline{\Pi}^{comp,1}, \underline{\Pi}^{comp,2}, \dots, \underline{\Pi}^{comp,q}, \dots, \underline{\Pi}^{comp,Q_{comp}} : q = 1, \dots, Q_{comp}\}$, such that $\sum_{i=1}^{NS_{comp}} \Pi^{comp,q,i} = 1$, $q = 1, \dots, Q_{comp}$. Notice that this gives rise to $\prod_{comp=1}^N Q_{comp} = N_{tot}$ possible combinations of steady-state probability vectors of the system components, i.e., to N_{tot} steady-state probability vectors for the entire system.
- b. For all the N components, select one steady-state probability vector among the set $\underline{\Pi}^{comp,q}$, $q \in \{1, \dots, Q_{comp}\}$ (generated at step a. above); in other words, this amounts

to selecting one of the $N_{\text{tot}} = \prod_{comp=1}^N Q_{comp}$ steady-state probability vectors for the entire SoS.

- c. Fixing the SoS steady-state probability vector selected in b., randomly sample the states $\zeta_{comp,i}$ (i.e., the capacities), $i \in \{1, \dots, NS_{comp}\}$, of all the components of the system (i.e., arcs). Then, compute the product delivered at the demand nodes propagating the flow in each component of the SoS through the GTST-DMLD (see Section 2.5.2).
- d. Repeat step c. a large number of times (e.g., 1000 in this work) and obtain the CDF for the product delivered at each demand node.
- e. Repeat steps c.-d. for another combination of the steady-state probability vectors, $\underline{\Pi}^{comp,q}$, $q \in \{1, \dots, Q_{comp}\}$, of all the N components, until all the N_{tot} possible combinations of the steady-state probability vectors of the SoS are explored.

At the end of steps a.-e., an ensemble of CDFs for each demand nodes is obtained, one for each of the N_{tot} possible combinations of steady-state probabilities of the entire SoS.

3. Identify the extreme minimum and maximum CDFs of the product delivered at the demand nodes that bound the set of CDFs produced at step 2. above.

4.4.2. Operative simulation steps considering GTST-DMLD system-of-systems representation: evaluation of recovery capacity

The time needed to recover the SoS from the worst scenario (i.e., the one characterized by components in the worst state) to a level in which all the demand nodes are satisfied, is carried out by three main steps:

1. Processing the epistemic uncertainties by interval analysis: this step leads to the identification of K_{comp} transition probability matrices \underline{P}_{comp}^k , $k = 1, 2, \dots, K_{comp}$, for each component ($comp = 1, 2, \dots, N$) of the SoS.
2. Evaluation of the SoS performance (i.e., recovery capacity) by MC simulation: this step leads to the determination of a set of CDFs of the time needed to recover the SoS, one for each possible combination of state probability matrices sampled.

3. Post-processing the results obtained at the previous step 2: this step leads to the identification of two extreme upper and lower CDFs that bound the set of CDFs produced at step 2. above.

In more details, step 1. is performed as follows:

- a. Select a component $comp$ and a row i of the matrix $\underline{\underline{P}}_{comp}$ and let the probability p_{ij} , $j = 1, 2, \dots, NS_{comp}$, vary within the corresponding interval $[\underline{p}_{ij}, \bar{p}_{ij}]$, in order to identify $C_{comp,i}$ combinations of probabilities such that $\sum_{j=1}^{NS_{comp}} p_{ij} = 1$ (by the assumption that the states are exhaustive, as for the previous eq. 4.3). If the component $comp$ is described by a semi-Markov process, select also a row i of the matrix $\underline{\underline{T}}_{comp}$ and let the mean, μ_{ij} , $j = 1, 2, \dots, NS_{comp}$, of the holding time distributions vary within the corresponding interval $[\underline{\mu}_{ij}, \bar{\mu}_{ij}]$ to obtain $M_{comp,i}$ vectors of combinations of mean values for the row i . Repeat this step 1. a. for all the rows $i = 1, 2, \dots, NS_{comp}$, of the matrices $\underline{\underline{P}}_{comp}$ and $\underline{\underline{T}}_{comp}$.
- b. Combine the $\sum_{i=1}^{NS_{comp}} C_{comp,i}$, vectors of probabilities for all the components ($comp = 1, 2, \dots, N$) to obtain K_{comp} transition probability matrices $\underline{\underline{P}}_{comp}^k$, $k = 1, 2, \dots, K_{comp}$, for each component. If the component $comp$ is described by a semi-Markov process, combine also the $\sum_{i=1}^{S_{comp}} M_{comp,i}$ vectors of mean values to obtain H_{comp} matrices $\underline{\underline{Mu}}_{comp}^h$, $h = 1, 2, \dots, H_{comp}$, of the mean values of the holding time distribution.
- c. Repeat steps a.-b. for each component ($comp = 1, 2, \dots, N$) of the SoS. All the N components are, then, associated with a set of transition probabilities matrices $\underline{\underline{P}}_{comp}^k$, $k = 1, 2, \dots, K_{comp}$; in addition, those components described by a semi-Markov process (i.e., N_{compSM} components) are also associated with a set of matrices, $\underline{\underline{Mu}}_{comp}^h$, $h = 1, 2, \dots, H_{comp}$, containing the mean values of the corresponding holding time distributions.

Step 2. is carried out as follows:

- a. Randomly select N matrices $\underline{\underline{P}}_k^{comp}$, $k \in \{1, 2, \dots, K_{comp}\}$, $comp = 1, 2, \dots, N$, for all the components of the SoS and N_{compSM} matrices $\underline{\underline{Mu}}_h^{comp}$, $h \in \{1, 2, \dots, H_{comp}\}$ for the components described by a semi-Markov process.
- b. Set $u = 1$ (counter of the number of simulations).
- c. Initialize the state of the components at the worst state ($\zeta_{comp,i}$, $i = 1, comp = 1, 2, \dots, N$): in this state configuration of the SoS, the product delivered to the demand nodes is lower than the optimum required.
- d. Initialize the following time variables: system simulation time $t = 0$, starting time of the simulation: this variable represent the current simulation time and is needed to compute the recovery time of the SoS; set $t^{comp} = \Delta t$, $comp = 1, 2, \dots, N$, where Δt is the time step of the simulation ($\Delta t = 1$ in arbitrary units in this work): these time variables are needed to determine if the component $comp$ can perform a state transition at a given time step (they are set to 1 since at this time step all the components perform the first state transition).
- e. Set $t = t + \Delta t$: if $t = t^{comp}$, then the component $comp$ performs a state transition: then, randomly sample its new state from the matrix $\underline{\underline{P}}_{comp}^k$ selected at step 2. a. and update the variable t^{comp} as follows:
 - ✓ If $comp$ is described by a Markov process, $t^{comp} = t^{comp} + \Delta t$, since a state transition occurs at each time step.
 - ✓ If $comp$ is described by a semi-Markov process, $t^{comp} = t^{comp} + t^*$, where t^* is the time of next transition that is sampled from the corresponding holding time distribution with mean value taken from the matrix $\underline{\underline{Mu}}_h^{comp}$ selected at the previous step 2. a. The sampled value t^* is rounded to the nearest integer except when it is zero; in this case, the value is rounded to 1.
- f. Evaluate the product delivered to the demand nodes by adopting the GTST-DMLD (see Section 2.5.2).

- g. Repeat steps e.-f. until the product delivered to the demand nodes is equal to, or higher than, the optimum required: the corresponding value of recovery time (t_{RT}^u) is then recorded for the simulation u .
- h. Set $u = u + 1$ and repeat steps c.-g. a large number of times (e.g., 1000 in this work).
- i. A CDF of the recovery time of the SoS is identified for a combination of state probability matrices \underline{P}_{comp}^k , $k \in \{1, 2, \dots, K_{comp}\}$, selected at step 2. a.
- j. Repeat the entire procedure (steps a.-i.) a large number of times (e.g., 10000 in this work) to explore many different combinations of probability matrices \underline{P}_{comp}^k , $k \in \{1, 2, \dots, K_{comp}\}$.

At the end of the procedure, a set of CDFs of the recovery time of the performance of the SoS is obtained.

The results are processed at step 3., where the minimum and maximum CDFs of the recovery time that bound the set of CDFs obtained at step 2. above are identified and the 99th percentiles of the distributions are computed as a measure of the recovery time.

4.5. Monte Carlo simulation and Hierarchical Graph within a multi-state system-of-systems framework

Within a multi-state SoS analysis framework, we wish to evaluate the performance of critical infrastructures (CIs) in terms of *robustness*, measured by the steady-state probability distribution of the product delivered at the demand nodes of the system. The quantitative evaluation is carried out by combining the Hierarchical Graph representation of Section 2.6 and MC simulation.

In Section 4.5.1, the operative steps of the basic procedure are presented. Then, a modification of the basic procedure is proposed in Section 4.5.2 to deal with CIs of large size: in particular, a clustering algorithm is adopted to pre-process the CIs in order to make its size manageable and reduce the computational burden associated to the analysis; details about the unsupervised spectral clustering technique adopted are given in Appendix of paper VI of Part II.

4.5.1. Operative simulation steps combining Monte Carlo method and Hierarchical Graph system-of-systems representations for robustness evaluation

We generically denote the state of a component of the CIs (i.e., the capacity of the arcs) as $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, $comp = 1, \dots, N$, where N is the total number of components in the SoS, the subscript ‘ $comp$ ’ indicates the component of interest, identified by its name or by an integer number from 1 to N , NS_{comp} is the total number of states for component $comp$, and i is the state identification number (when $i = 1$, the component is in the worst state, whereas when $i = NS_{comp}$, it is in the best state). For example, supposing that component $S_I^{(2)}_S_2^{(2)}$ can enter three possible states, namely 0, 10 and 20, we denote the total number of states for the component as $NS_{S_I^{(2)}_S_2^{(2)}} = 3$, and the corresponding states as $\zeta_{S_I^{(2)}_S_2^{(2)},1} = 0$, $\zeta_{S_I^{(2)}_S_2^{(2)},2} = 10$, and $\zeta_{S_I^{(2)}_S_2^{(2)},3} = 20$.

The quantity of product requested by the demand nodes is indicated by the vector $\{D_{dem}\}$, $dem \in \{1, \dots, N\}$, where the subscript ‘ dem ’ identifies the demand nodes.

In what follows, we describe an algorithm combining the MC method and Hierarchical Graph representations for the evaluation of the robustness of CIs within a multi-state SoS framework; as mentioned before, the robustness is quantified in terms of the steady-state probability distribution of the product delivered to the demand nodes.

In extreme synthesis, the algorithm requires as inputs:

- the Hierarchical Graph that allows representing the origin-destination paths and the corresponding arcs in hierarchical levels (see Section 2.6);
- the steady-state probabilities of transition between the different arc states (i.e., capacities) $\zeta_{comp,i}$, $i = \{1, 2, \dots, NS_{comp}\}$, $comp = \{1, 2, \dots, N\}$;
- the vector $\{D_{dem}\}$, $dem \in \{1, \dots, N\}$, of product required by demand nodes;
- the importance of the demand nodes (see Section 2.6).

The output of the algorithm is represented by the steady-state probability distributions of the product delivered to the demand nodes. For clarification purposes, we describe the procedure with reference to the simple example of Figure 4.2, where two interconnected systems, $S^{(1)}$ and $S^{(2)}$, are shown. The $N = 5$ components are: $S_I^{(1)}_S_2^{(1)}$, $S_2^{(1)}_S_3^{(1)}$, $S_2^{(1)}_S_I^{(2)}$, $S_I^{(2)}_S_2^{(2)}$ and

$S_2^{(2)}_S S_3^{(2)}$. The input component is arc $S_I^{(1)}_S S_2^{(1)}$ that serves five demand nodes (i.e., the goals), $S_2^{(1)}$, $S_3^{(1)}$, $S_I^{(2)}$, $S_2^{(2)}$ and $S_3^{(2)}$, explicitly represented at the top of the diagram.

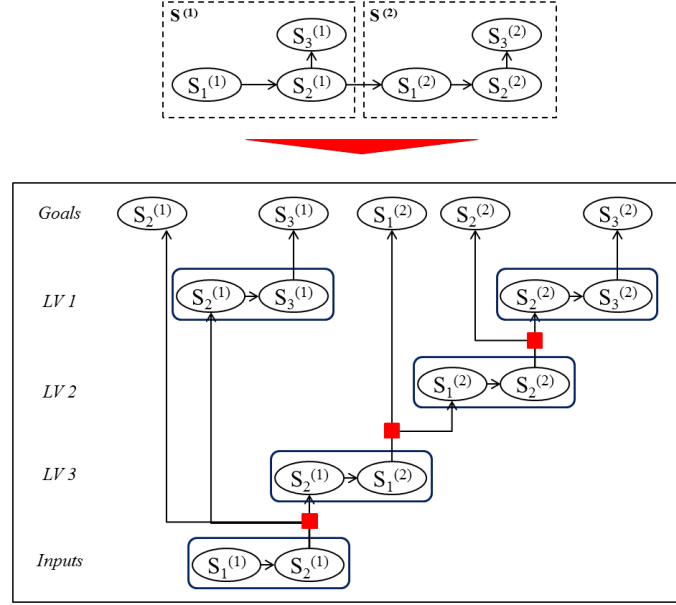


Figure 4.2: Hierarchical Graph of a generic example taken as reference to illustrate the algorithm; LV: Level.

The evaluation is carried out from the bottom to the top of the hierarchy and consists of the following steps:

- 1) Determine one possible system configuration by sampling the capacity of the arcs, $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, $comp = \{1, \dots, N\}$, from the corresponding steady-state probability distributions;
- 2) Identify the minimum arc capacity ($mpath_{dem}$, $dem \in \{1, \dots, N\}$) for each origin-destination path: this capacity corresponds to the maximum product that can be delivered to the corresponding demand node dem , $dem \in \{1, \dots, N\}$; for example, in Figure 4.3 the minimum arc capacity for the path from $S_I^{(1)}$ to $S_3^{(1)}$ is the minimum among the capacities of arcs $S_I^{(1)}_S S_2^{(1)}$ and $S_2^{(1)}_S S_3^{(1)}$, connecting $S_I^{(1)}$ and $S_3^{(1)}$;
- 3) Set the input (inp) to the network equal to the capacity of the input arc, i.e., $inp = \zeta_{comp,i}$, where $i \in \{1, 2, \dots, NS_{comp}\}$ and $comp$ is the index of the input arc (in the example of Figure 4.2, the input arc is $S_I^{(1)}_S S_2^{(1)}$);

- 4) If the input is zero ($inp = 0$), no product can be delivered to the demand nodes: $EP_{dem} = 0$ for all $dem \in \{1, \dots, N\}$; otherwise, estimate the optimal flows $\{EP_{dem}\}$ that can be delivered to the demand nodes by the following steps:
- a. Estimate the vector $\{EP_{dem}\}$ of optimal flows to the demand nodes taking into account i) the importance of the demand nodes and ii) the minimum capacity of each path ($mpath_{dem}$, $dem \in \{1, \dots, N\}$) that limits the quantity of product that can be delivered to the demand nodes (Figure 4.4, top).
 - b. Initialize an auxiliary variable $surp$ to zero (i.e., $surp = 0$). This variable is used to quantify the surplus, i.e., the amount of product that cannot be allocated in the network due to arc capacity constraints (i.e., due to the bottlenecks of the infrastructure).
 - c. Check if the capacities of the links, $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, can support the sum of the estimated optimal products to the corresponding demand nodes, (dem) computed at the previous step 4 a. Such evaluation is performed from the bottom to the top of the diagram. If the sum of the estimated optimal product to the nodes served by a link is higher than its capacity, save the exceeding amount (Δ) in the auxiliary variable $surp$ (i.e., $surp = surp + \Delta$) and compute the optimal partition just for the nodes that are supplied by that link, considering as input the corresponding arc capacity $inp = \zeta_{comp,i}$, where $i \in \{1, 2, \dots, NS_{comp}\}$ and $comp$ is the link under analysis (Figure 4.4, middle).
 - d. Create a "new" graph, where the "new" capacities of all the arcs are updated on the basis of the quantity of product, $\{EP_{dem}\}$, that has been effectively allocated at step 4 c. In particular, the arc capacities are reduced by the total quantity of product that they have already supplied to the corresponding demand nodes (Figure 4.4, bottom).
 - e. Compute again the minimum arc capacity for each path of the "new" graph (as in step 2) to evaluate the new maximum product that can reach the corresponding demand nodes (Figure 4.4, bottom).
 - f. Update the demands $\{D_{dem}\}$, $dem \in \{1, \dots, N\}$, reducing them by the quantity $\{EP_{dem}\}$ that has been already allocated at step 4.c (Figure 4.4, bottom).

- g. Set the input inp equal to the auxiliary variable $surp$ ($inp = surp$) and repeat step 4 until $surp > 0$ and the minimum ("new") arc capacity for at least one path is not zero. When one of these conditions is verified, the final vector $\{EP_{dem}\}$ of the optimal product that can be delivered to the demand nodes is determined (Figure 4.4, bottom).

The procedure above is repeated a large number of times (e.g., 10000) for many different MC-sampled values of the arc capacities and the probability distribution of the product delivered at steady state to each demand node is obtained.

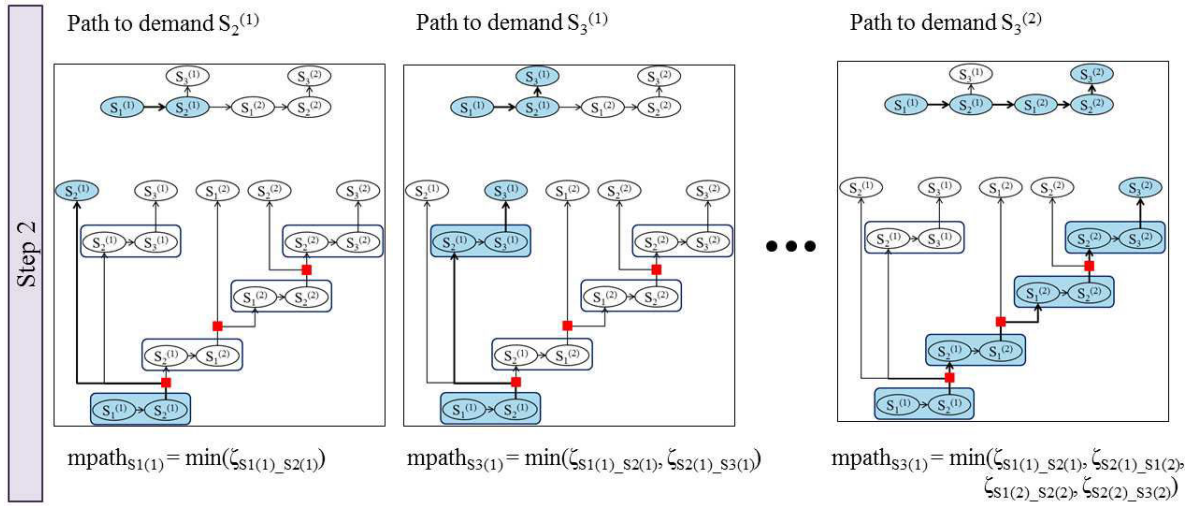


Figure 4.3: Exemplification of step 2 of the algorithm with respect to the example proposed in Figure 4.2.

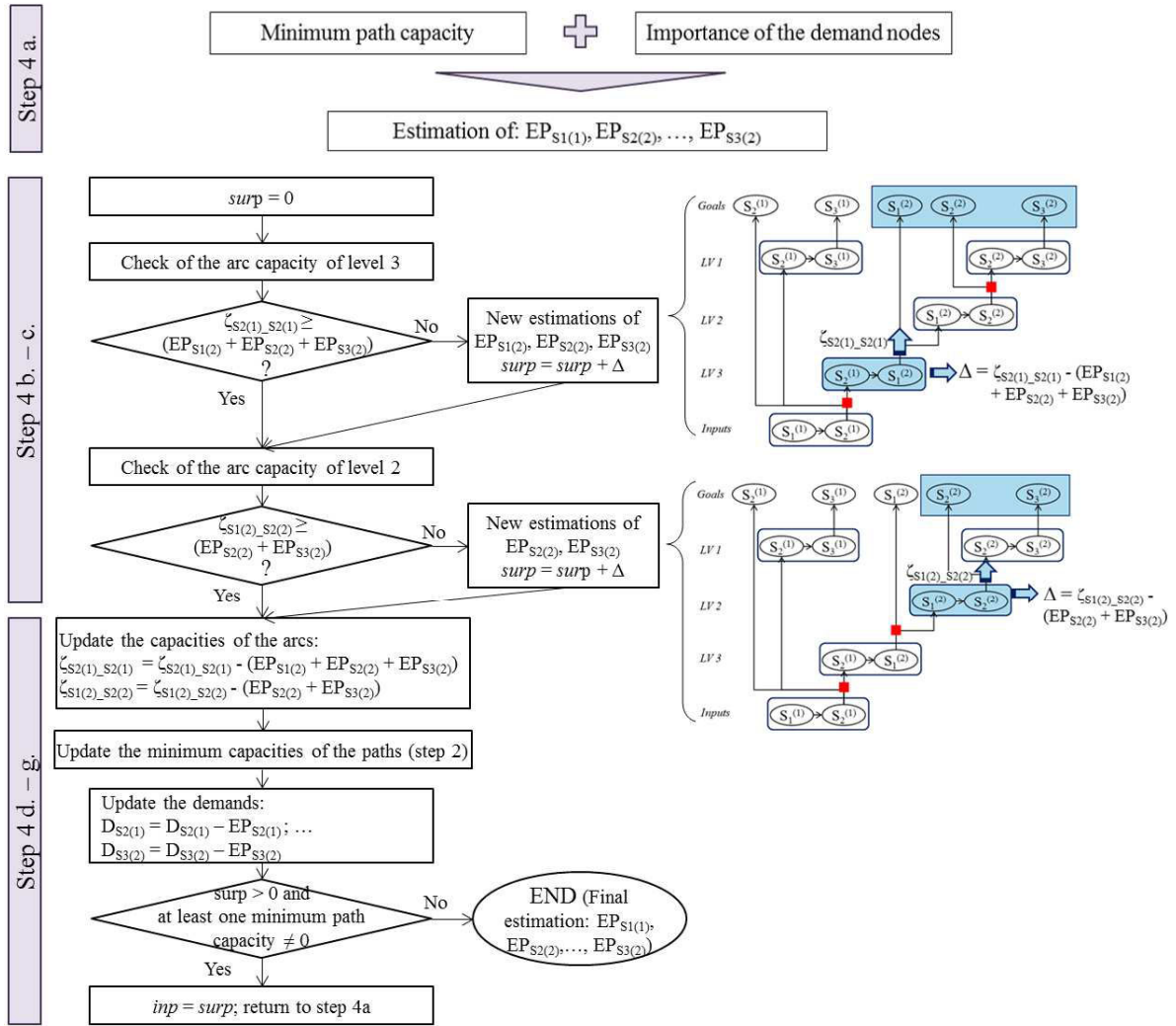


Figure 4.4: Exemplification of step 4 of the algorithm with respect to the example proposed in Figure 4.2.

It is worth noting that the procedure proposed is based on several iterative estimations of the vector $\{EP_{dem}\}$, obtained by repeating steps 4 a. – g. from the bottom to the top of the hierarchy: in the very first iteration, the system configuration is the one sampled at step 1. and the input product corresponds to the capacity of the input arc; then, at each loop a “new” graph is considered where (i) the new product input value is represented by the surplus ($surp$), i.e., the amount of product that has not been allocated in the network at the previous iteration, (ii) the “new arc capacities are reduced by the total amount of product they have already supplied at the previous iteration and (iii) the “new” demands are scaled by the quantity already allocated in the previous iteration.

Finally, it is worth mentioning that a drawback of the Hierarchical Graph representation proposed may be represented by its difficult applicability to large networks since *all* the origin-destination paths have to be identified and the bottlenecks of each path have to be spotted out. To overcome this limitation, we propose to pre-process the infrastructure system by means of a clustering algorithm to reduce the systems dimension by “collapsing” many components in few representative clusters and then apply the Hierarchical Graph to the “clustered” infrastructure (details about the particular clustering technique adopted in the present work, namely unsupervised spectral clustering algorithm, are reported in the Appendix of paper VI of part II). The general concepts underlying the pre-processing phase based on clustering is discussed in the following Section 4.5.2.

4.5.2. Combination of the Hierarchical Graph representation and a clustering algorithm for managing large-sized critical infrastructures

In order to manage large-sized CIs, it may be useful to resort to clustering techniques to reduce the complexity and dimension of the system. For illustration purpose, refer to the simple example of Figure 4.5, left, where the original components of a network, namely $S_1^{(I)}$, $S_2^{(I)}$, ..., $S_{16}^{(I)}$, are reported. According to some features of interest (e.g., proximity), such components can be clustered in groups of “similar characteristics”: in the example proposed, four clusters, C_1 , ..., C_4 , are identified (dotted oval shape in Figure 4.5, left). Then, a less complex analysis can be performed on the new fictitious, artificial (i.e., clustered) network, composed just by the identified clusters (Figure 4.5, right).

The cluster analysis can be carried out at different level of details: an artificial network with a high number of clusters is closer to the original one and, thus, it is more detailed (i.e., it carries more information) than one with a small number of clusters. The system can be clustered at different levels of details, which allows building a hierarchical⁴ clustering representation where the different hierarchical levels correspond to the different levels of detail of the analysis.

⁴ Notice that in this case the term “hierarchical” refers to the level of detail of the clustering and not to the levels of the Hierarchical Graph representation that instead correspond to the number of demands served by a given arc of the network.

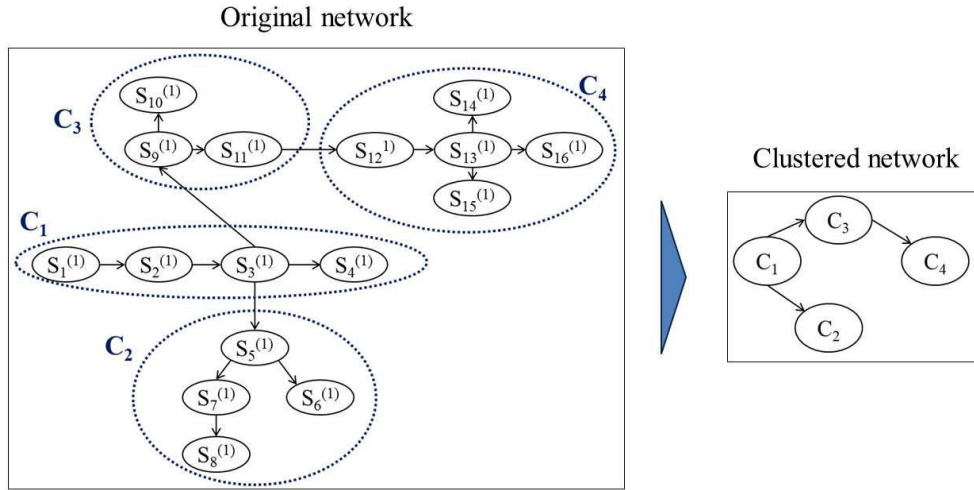


Figure 4.5: Exemplification of the clustering procedure.

In order to reduce the size of the infrastructure under analysis, the SoS is clustered (possibly at different hierarchical levels of detail): an artificial (fictitious) network composed by k_L clusters, $C_1^{(L)}, \dots, C_{k_L}^{(L)}$, is produced at each (clustering) hierarchical level L . Notice that the last level of the clustering hierarchy coincides with the real SoS, i.e., the corresponding clusters coincide with the actual/original/real nodes of the SoS. The clustering is performed on the entire network except for the input nodes that are left out (only one generation node is considered in the application of the present thesis). For illustration purposes, Figure 4.6 depicts a sketch of the decomposition in five (clustering) hierarchical levels of a SoS with one input node, $S_I^{(1)}$; level 1 of the hierarchy is then composed by two nodes: the input, $S_I^{(1)}$, and the rest of the system “condensed” in cluster $C_1^{(1)}$. The clustering algorithm allows a new analysis at hierarchical level 2 and it decompose cluster $C_1^{(1)}$ of hierarchical level 1 into two clusters $C_1^{(2)}$ and $C_2^{(2)}$. At this point, if we want to increase the level of refinement of the analysis we can use the algorithm to further split clusters $C_1^{(2)}$ and $C_2^{(2)}$. In the example of Figure 4.6, this results in the decomposition of cluster $C_1^{(2)}$ into three clusters ($C_1^{(3)}$, $C_2^{(3)}$ and $C_3^{(3)}$) and cluster $C_2^{(2)}$ into five clusters ($C_4^{(3)}$, $C_5^{(3)}$, $C_6^{(3)}$, $C_7^{(3)}$ and $C_8^{(3)}$). The Hierarchical Graph representation of the decomposed system at level 3 is also shown on the right.

A cluster k is characterized by its demand D_k that is the sum of the demands of the real nodes at its inside: for example, cluster $C_1^{(4)}$ of Figure 4.6 has demand equal to the sum of the demands of nodes $S_3^{(1)}$, $S_6^{(1)}$ and $S_7^{(1)}$.

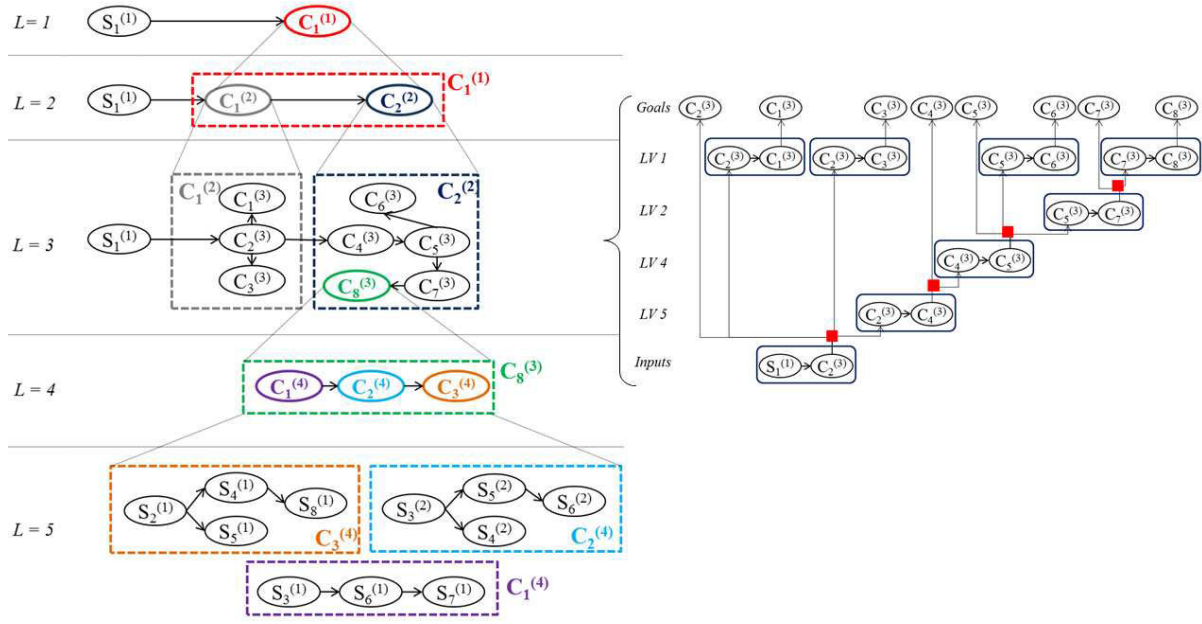


Figure 4.6: Left: sketch of the decomposition of a system in five hierarchical levels (L) where the last one ($L = 5$) coincides with the actual nodes of the system; right: Hierarchical Graph of the corresponding hierarchical level 3; LV: Level of the Hierarchical Graph.

For a given clustering hierarchical level L the quantitative evaluation of the performance of the "artificial" clustered system is carried out as illustrated in the previous Section 4.5.1 by means of an indicator that represent the "global" state of the clusters $C_1^{(L)}, \dots, C_{k_L}^{(L)}$ of level L as a "synthesis" of the real capacity $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, $comp = 1, \dots, N$, of the arcs contained in the cluster itself. Actually, a measure of the cluster state is needed to approximately estimate the quantity of product that a cluster can receive and deliver to other clusters.

To represent the state of a cluster k (i.e., its performance) we consider an indicator id_k based on the ratio of the expected capacity of cluster k at current and at nominal (optimal) conditions as follows:

$$id_k = \frac{\sum_{comp=1}^{n_k} w_{comp} * \zeta_{comp,i}}{\sum_{comp=1}^{n_k} w_{comp} * \zeta_{comp,NS_{comp}}}, \quad (4.5)$$

where $comp$ indicates the component (arc) of the original network, n_k is the number of arcs inside cluster k , $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, is the current (i.e., actual / sampled) state of the component $comp$, $\zeta_{comp,NS_{comp}}$ is the maximum capacity of the arc $comp$, and w_i is the weight associated to the capacity of the arc $comp$. The weight w_{comp} is computed as the ratio between the capacity of the arc $comp$ and the sum of the maximum capacities of all the arcs of

the network, i.e., $w_{comp} = \zeta_{comp,i} / \sum_{comp=1}^N \zeta_{comp,NS_{comp}}$, $i \in \{1, 2, \dots, NS_{comp}\}$ and gives an idea of the weight of the arc in the entire network.

Notice that the state of a cluster affects the cluster itself and the connected clusters, since the cluster is both a *fictitious load node* (which should provide itself with the required amount of product) and a *fictitious transmission node* (which should transmit the product to the other connected clusters). The top of Figure 4.7 shows two clusters, C_1 and C_2 , supplied by the input source $S_1^{(I)}$: cluster C_2 is both a load and a transmission node, since on one side it contains five demand nodes ($S_2^{(I)}$, $S_3^{(I)}$, $S_4^{(I)}$, $S_5^{(I)}$ and $S_6^{(I)}$ in Figure 4.7, bottom) and on the other side it is required to transmit the product to cluster C_2 . In particular, the product from input source $S_1^{(I)}$ has to pass through two arcs ($S_2^{(I)} \rightarrow S_5^{(I)}$ and $S_5^{(I)} \rightarrow S_6^{(I)}$) contained in C_1 to reach cluster C_2 : if their capacities decrease, then the flow to nodes $S_5^{(I)}$ and $S_6^{(I)}$ (i.e., to the cluster C_1) and to nodes $S_7^{(I)}$, $S_8^{(I)}$ and $S_9^{(I)}$ (i.e., to the cluster C_2) is reduced.

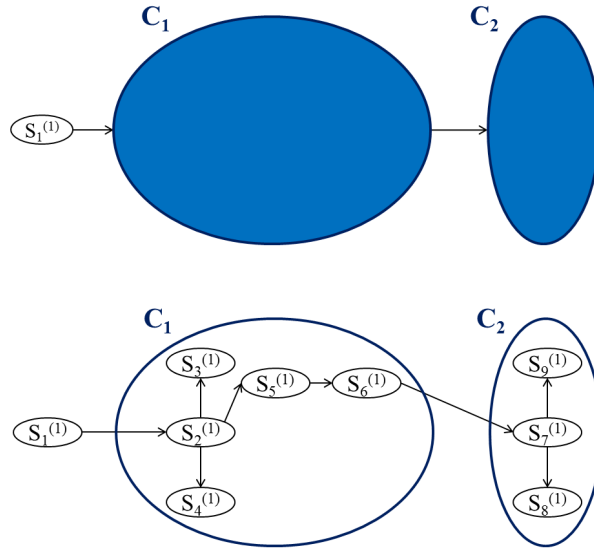


Figure 4.7: Top: artificial system composed by two clusters C_1 and C_2 supplied by one input node $S_1^{(I)}$. Bottom: illustration of the real nodes inside the fictitious clusters: two arcs of C_1 are needed to supply C_2 .

Thus, when the "capacity" of a cluster decreases, the consequence is twofold: the cluster cannot satisfy itself (i.e., the demand nodes at its inside) and also the connected clusters at best. In order to take into account the "twofold" reduction of performance, we "artificially reduce" the amount of product that can be given to the cluster itself and that can be delivered to the connected clusters by multiplying i) the maximum demand D_k that it requires and ii) the maximum capacities, $\zeta_{comp,NS_{comp}}$, of the arcs (*comp*) that link the output clusters, respectively, by the indicator of the state of the cluster, id_k .

5. APPLICATION 1: EXTERNAL EVENT RISK ASSESSMENT

In this Chapter, the case studies considered within the framework of external event risk assessment are briefly illustrated (Section 5.1), the corresponding system representations adopted are shown and the main results are provided (Sections 5.2 and 5.3). For further details the interested reader is referred to the corresponding papers I - IV of Part II.

Two case studies (hereafter referred to as “A” and “B”) are taken into account. They deal with a critical plant, i.e., a nuclear power plant (NPP), exposed to risk of natural external events, i.e., earthquakes. Internal emergency devices have been designed to provide safety for the plant upon occurrence of the hazardous event, i.e., even if the infrastructure services are not available. However, since the internal emergency devices can fail too, the boundaries of the study are extended to the analysis of the responses of interconnected systems (i.e., the power and water distribution and transportation networks) that provide inputs necessary to keep or restore the plant in the safe state. The NPP is considered in a safe condition if it does not cause health problems and environmental damages, i.e., if it does not release radioactive material to the environment. To maintain this state it must be provided with energy and water flow inputs to absorb the heat that it generates. Case study B is an extension of case study A in what concerns the technical aspects inside the plant: while case study A is provided with an internal water system, case study B distinguishes between main feedwater (MFW) system, high pressure coolant injection (HPCI) system and low pressure coolant injection (LPCI) system that is associated with the automatic depressurization system (ADS).

Two quantities are used to characterize the loss of functionality of the various components of the system of systems (SoS) embedding the critical plant, upon the occurrence of a damaging external event:

- i. from the safety viewpoint, the probability that the critical plant remains in a safe condition given the possible failure configuration of the components;
- ii. from the physical resilience viewpoint, the time needed to restore the safe state of the critical plant, i.e., the duration of the recovery actions needed to bring the components back to the level of functionality required to restore the safe condition of the plant, eventually facing also the occurrence of subsequent aftershocks.

Both quantities are computed in this work. In particular, the evaluation of the safety (point i. above) is carried out within a binary state modeling by considering Fault Tree (FT) and Muir Web representation; both the evaluations of safety and physical resilience (points i. and ii. above) are performed within a binary state and multi-state modeling, by adopting Hierarchical Modeling and Goal Tree Success Tree – Dynamic Master Logic Diagram representation (GTST-DMLD), respectively.

5.1. Case studies A and B: description

The system under analysis is composed by a nuclear power plant, a water system that provides coolant useful to absorb the heat generated in the nuclear power plant, a power system that provides electrical energy for the running of the nuclear power plant and the water system, and a road network relevant to the power and water systems for the transport of material and/or plant operators. For illustration purpose, the physical representation of the case study B is reported in Figure 5.1, referring to a Cartesian plan (x, y) with origin in a river.

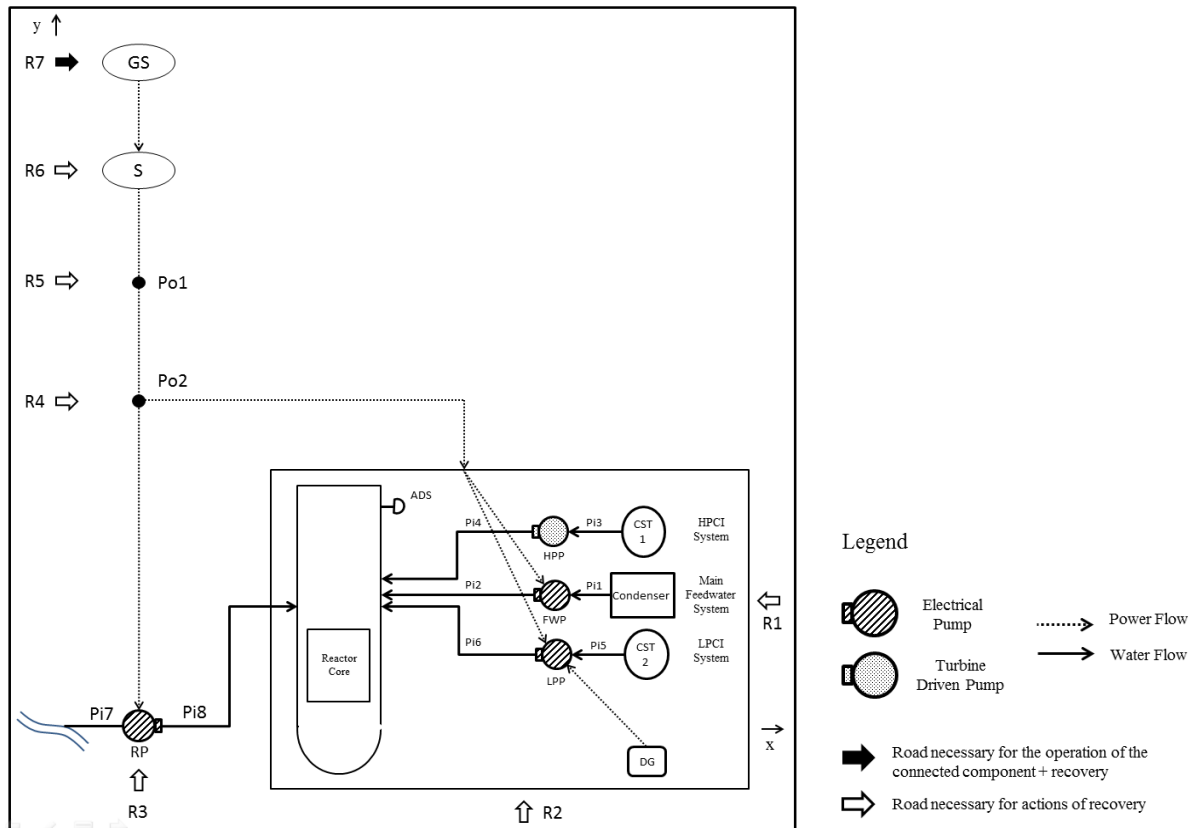


Figure 5.1: Physical representation of the system. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, RP: River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator, R: Road access.

The nuclear reactor is the element of the nuclear power plant that must be provided with the necessary inputs to assure the safe state of the entire plant.

The common components to both case studies A and B are the following:

- an offsite power system (EE) composed by a generation station (GS) that produces the electrical energy, a substation (S) that transforms the voltage from high to low, power lines and poles (Po1 and Po2) to support them;
- an internal emergency power system (IE) represented by the emergency diesel generator (DG);
- an external water (EW) system constituted by the river (i.e., the source of water) a pump (RP) that receives electrical power from the offsite power system and pipes (Pi7 and Pi8) that carry the water;
- a road transportation system identified by seven road accesses to the components of the SoS. The state of the roads is important for access of materials (e.g., fuel) and/or operators for operation and/or maintenance. Distinctions are made between road accesses needed just for the recovery actions (e.g., R1 – R7) and road accesses needed also for the operation of the connected elements (e.g., R7). Notice that when the recovery time is not evaluated the road access assumes the function of “reserve component”, so that elements that fail can be immediately repaired/replaced if the access to it through the road system does not fail; otherwise the road accesses increase the time to recover the connecting components if they are in a failure state.

Peculiar features of case study A are the following:

- the water and power systems are subdivided into two independent parts, external and internal to the plant; the latter one represents the emergency system of the plant which needs to obviate at the absence of input from the main external system;
- the internal water system is composed by the same elements of the external water system, except for the source of water that is an artificial reservoir (e.g., a tank or a pool).

On the contrary, the main feature of case study B is that distinctions are made between main inputs, internal barriers, external support and recovery supporting elements, according to the

safety levels introduced in Section 2.5.1. In particular:

- main inputs include the main feedwater (MFW) system that provides coolant useful to absorb the heat generated; it is constituted by a condenser where the unused steam coming from a turbine is condensed into water that is pumped to the reactor vessel by the feedwater pump (FWP) and pipes (Pi1 and Pi2);
- internal barriers consider High Pressure Coolant Injection (HPCI) and Low Pressure Coolant Injection (LPCI) systems that provide water to cool the reactor, an automatic depressurization system (ADS) that reduces the pressure in the reactor vessel and a diesel generator (DG) that can provide the LPCI system with power (see Figure 5.1). In case of accident damaging the MFW system function, the HPCI and LPCI systems need to provide the necessary function. Both systems are composed by a condensate storage tank (CST1 and CST2, respectively), a pump (HPP and LPP, respectively) and pipes (Pi3, Pi4 and Pi5, Pi6, respectively). To operate, the LPCI system needs the automatic depressurization system (ADS) to reduce the pressure inside the vessel;
- external supports are the external water system (EW) and off-site power system (EE) (as for the case study A);
- recovery supporting elements include the road accesses (as for the case study A).

5.2. Case study A: system-of-systems representations and main results

The case study A has been represented by the following system representations: Fault Tree (FT) (Section 5.2.1), Muir Web (Section 5.2.2) and Hierarchical Modeling (Section 5.2.3).

5.2.1. Fault Tree representation

Figure 5.2 shows the primary levels of the FT built for the analysis. It is used to derive the main causes of occurrence of the event that the nuclear power plant (NPP) is in an unsafe state (top event), which are the lack of energy and/or water supply by both the internal and external systems. The triangular elements in the tree indicate that the corresponding events are not further “expanded” here into the corresponding causes; a more detailed FT representation is given instead in paper I of Part II.

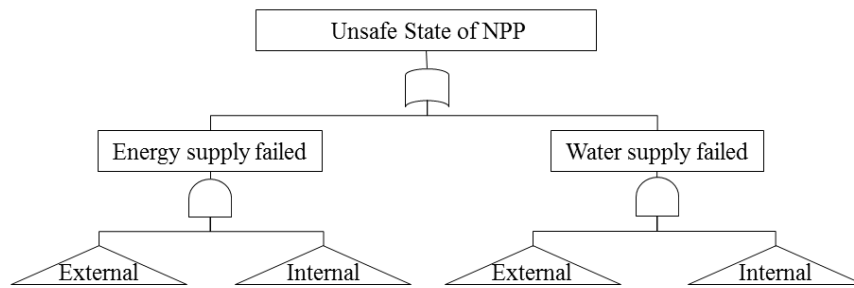


Figure 5.2: Fault tree of the system of systems of interest: upper levels. The elements in the triangular shape are not detailed. NPP: Nuclear Power Plant.

The following analyses have been carried out:

- a. a comparison between the probabilities that the nuclear power plant reaches an unsafe state after an earthquake of a given magnitude, with respect to different site-to-source distances: as expected, the higher the distance, the lower the probability to get to an unsafe state;
- b. a comparison of the previous probabilities (a.), obtained in the case of dependence of the nuclear power plant on the interconnected infrastructure systems, with those obtained in the case of independence, i.e., considering the nuclear power plant as an isolated system provided only by its internal emergency devices: the results show that the probability of reaching an unsafe state is higher in this latter case; in particular, the “resilience” contribution of the interdependent systems to the safety of the nuclear power plant is significant for low magnitudes when the source-to-site distance is small, and for high magnitudes when the source-to-site distance is big.

5.2.2. Muir Web representation

In Figure 5.3, the SoS representation is given by the Muir Web that shows the physical components of the infrastructure systems and the factors they depend on. In this representation the connections among the elements are depicted without expliciting the types of dependencies introduced in Section 2.3: such dependencies are illustrated in Figure 5.4 only with respect to the physical components of the SoS. A more thorough description is given in paper II of Part II.

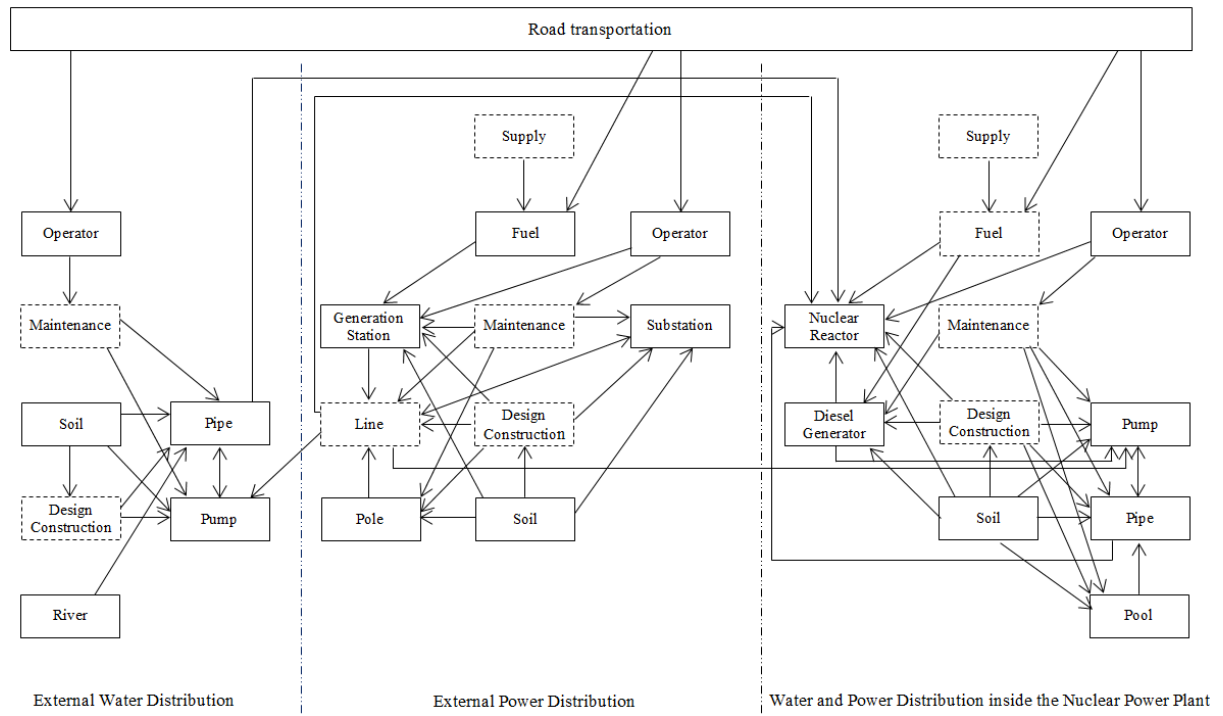


Figure 5.3: Muir Web of the system of systems of interest: the elements in the dashed box are not considered in the present study.

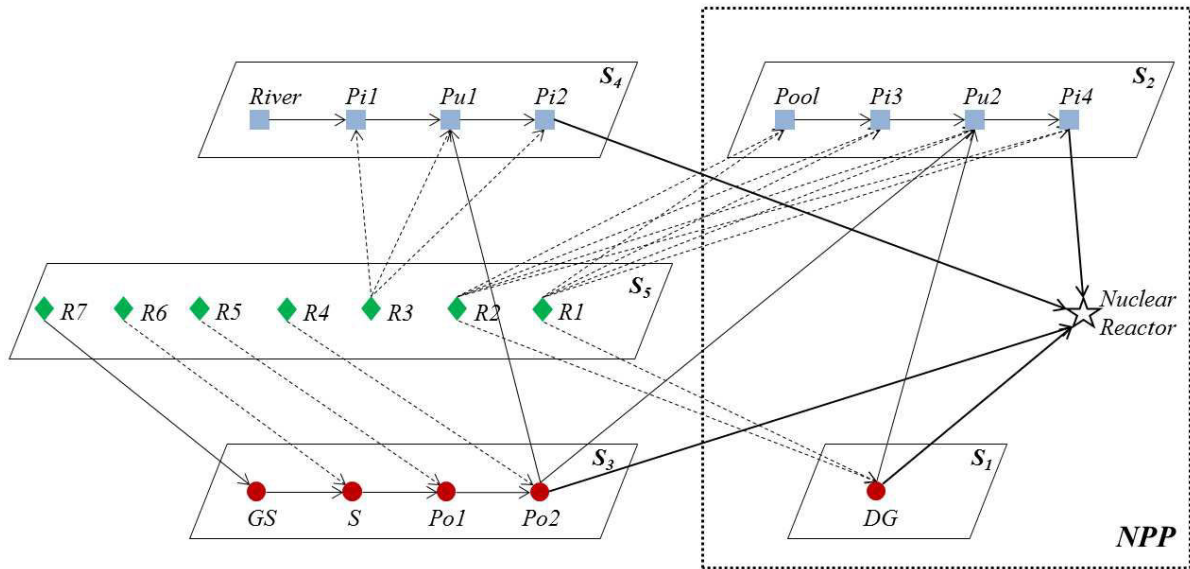


Figure 5.4: Representation of the physical components of the Muir Web of Figure 5.3, highlighting the different types of dependencies. The interconnected systems S_i , $i=1, \dots, 5$, can provide services relevant to the safe state of the nuclear power plant (NPP). The links represent the direct dependencies (solid lines), the support dependencies (dashed lines) and the dependencies of the nuclear reactor (star) on its interconnected systems (bold lines). GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, S_1 : internal power system, S_2 : internal water system, S_3 : external power system, S_4 : external water system, S_5 : Road transportation. The name of the components may differ with respect to the physical representation of Figure 5.1 since this representation is referred to the case study A.

The following analyses have been carried out (in addition to the comparisons of point a. and b. presented with respect to the FT results) (see paper II of Part II):

- a. a comparison of the probability that the NPP reaches an unsafe state after an earthquake of a given magnitude in the case of dependence of the NPP on the interconnected infrastructure systems, with the same probability obtained under the hypothesis of isolated infrastructure systems (i.e., removing all the inter-system links): this comparison allows highlighting the impact of the interdependencies in the safety of the NPP. The results show that the probability to reach an unsafe state is higher in the latter case (i.e., in the case of isolated infrastructure systems), due to the particular “redundancy” role of the road accesses under the assumption of immediate recovery of the components;
- b. the same comparison as in a., but considering, in the “isolated” case a dependence between the road accesses and the corresponding components and maintaining the independence among the other systems: the results show that in this case the probability to reach an unsafe state is lower than before; this means that the inter-system links among the power and water systems increase the probability of failure of the SoS and, thus, of the NPP being in an unsafe state.

5.2.3. Hierarchical Modeling representation

The SoS is structured hierarchically as in Figure 5.5.

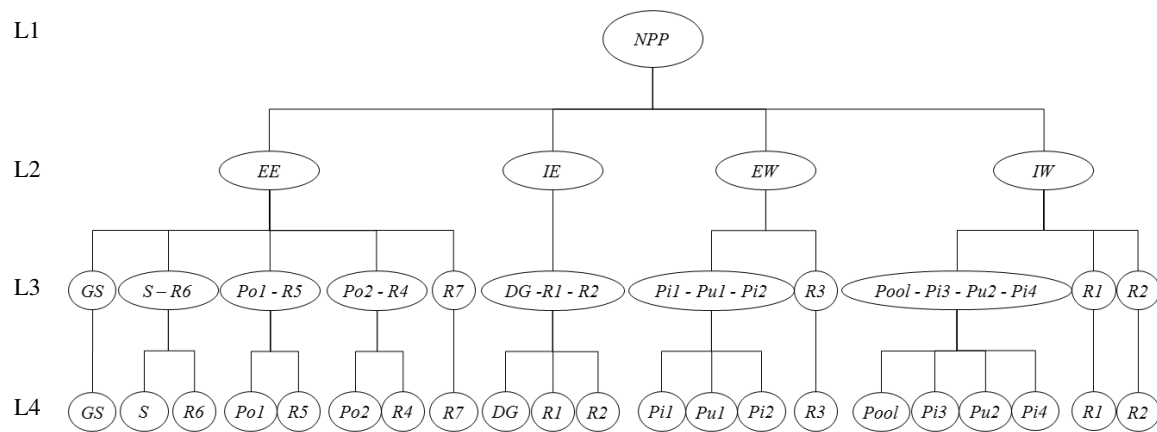


Figure 5.5: Hierarchical representation of the system of systems. NPP: Nuclear Power Plant, EE: External Energy system, EW: External Water system, IE: Internal Energy system, IW: Internal Water system, GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, L: Level. The name of the components may differ with respect to the physical representation of Figure 5.1 since this representation is referred to the case study A.

The nuclear power plant is at the top (level 1) of the hierarchy. Its safety is guaranteed by the power and water systems that are partitioned, at the level 2, into external and internal parts: external energy (EE), internal energy (IE), external water (EW) and internal water (IW). The road accesses represent the recovery supporting elements and, as explained in Section 2.4, they belong to the systems to which they provide support: in this case they belong to the corresponding EE, IE, EW and IW systems. Level 3 is, then, composed by single individual components or road accesses or a combination of them, whereas level 4, the most specified level, is formed by the individual elements (components and road accesses) of the SoS. Notice that only the recovery supporting elements may belong to different systems (or groups): for example, R1 and R2 are parts of both the IE and IW systems, whereas the other components appear in just one system (e.g., the pole Po2 belongs to the EE system).

A more thorough description is given in paper III of Part II.

The following analyses have been carried out:

- the estimation of i) the probability that the NPP reaches an unsafe state upon occurrence of an earthquake of a given magnitude and ii) the conditional probabilities of failure of the corresponding external energy, external water, internal energy and internal water systems, given that the NPP has entered an unsafe state (safety analysis);
- the estimation of the capacity of recovering the SoS starting from the top level of the hierarchy (recovery of the NPP) and proceeding downward to the lower hierarchical levels to identify the root causes and major contributors to the system recovery. The criticality importance measure [Zio, 2009a] for a component (or group) at a given level of the hierarchy at time t is used to guide the analysis through the hierarchical model: such measure is defined as the probability that the component (or group) at a given level of the hierarchy is critical for the system and failed at time t , given that the system is failed at time t (recovery capacity analysis).

5.3. Case study B: system-of-systems representations and main results

The case study B, differently from the case study A, has been represented by only one representation technique: the GTST-DMLD.

Figure 5.6 shows the GTST-DMLD of the SoS depicted following the scheme of Figure 2.8 in Section 2.5.1. The goal function is the safety of the nuclear power plant assured by water inputs (i.e., the principal function) that can be provided through four different alternative paths (ξ_k^{Water} , $k = 1, \dots, 4$): the main feedwater system (ξ_1^{Water}), the high pressure coolant injection system (ξ_2^{Water}), the combination of low pressure coolant injection and depressurization systems (ξ_3^{Water}), the external water system (ξ_4^{Water}). The power coming from outside (Ext) or inside (Int) the plant is an auxiliary function to support the operation of most of the water systems. For the explanation of the logic gates, of dot- and square- dependencies, the reader is referred to Section 2.5.1 and to paper IV of Part II.

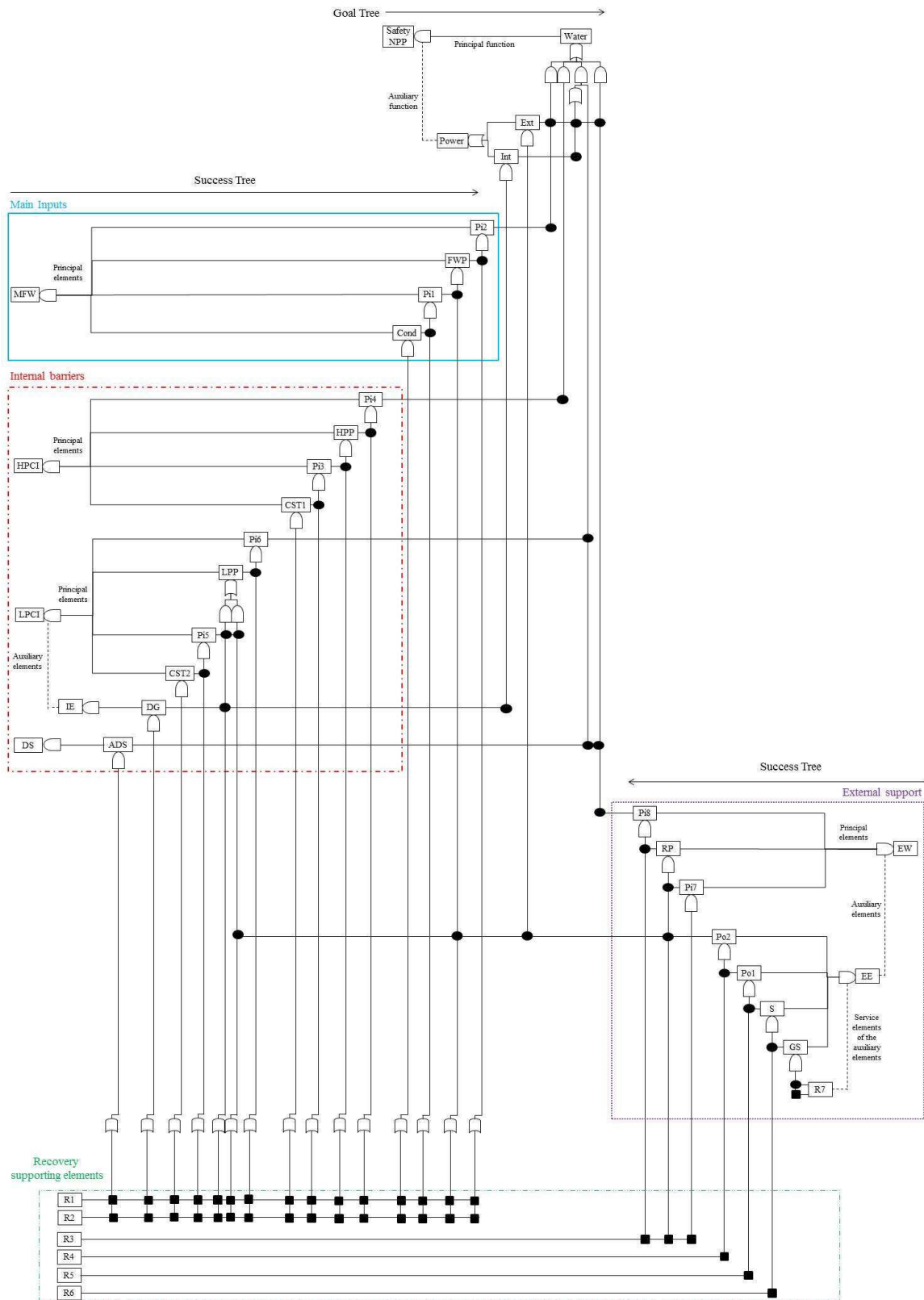


Figure 5.6: GTST-DMLD of the system of systems of Figure 5.1. MFW: Main Feedwater System; HPCI: High Pressure Coolant Injection System; LPCI: Low Pressure Coolant Injection System; IE: Internal Energy System; DS: Depressurization System; EW: External Water System; EE: Offsite power system; R: Road access; GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, Cond: Condenser; RP:

River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator.

The following analyses have been carried out:

- a. a comparison between the probabilities that the nuclear power plant enters risk, marginal and healthy states, calculated by multi-state and binary state models: as expected, the probability to enter a risk state is the same for both models; on the contrary, the probability of being in a healthy state is lower for the multi-state model that identifies (marginal) configurations of the SoS that present criticalities because they do not satisfy given safety margins (Figure 5.7, left);
- b. a comparison of the previous probabilities (a.) obtained taking into account also sequences of aftershocks that could further degrade the safety of the nuclear power plant. The multi-state model evidences a higher probability that the nuclear power plant enters a risk state with respect to the binary state model (Figure 5.7, right). This because the multi-state model can capture the impact of the aftershocks that instead are almost neglected by the binary state model since the structural healthy state of the components is characterized by fragilities that are not much sensitive to the small ground motion levels produced by aftershocks. Actually, the increased probability of the risk state is mainly due to the degradation of the marginal state that is more exposed to aftershocks than the healthy state;
- c. a comparison between the probability density function (PDF) of the time necessary to restore the healthy state of the nuclear power plant, given that the plant has entered a marginal and risk state, and the PDF of the recovery time of the marginal state given that the plant has entered in risk state using both the i) binary state model and ii) multi-state model, without considering the occurrence of the aftershocks:
 - i) from the first comparison, it can be seen that the binary state model is less conservative than the multi-state model in the sense that it identifies a mean time to recover the healthy state which is lower than the one identified by the multi-state model (Figure 5.8, right), but higher than the one needed to recover a marginal state (Figure 5.8, left). On the contrary, the multi-state model is capable of capturing the fact that a faster recovery to a safe condition is possible, but this safe condition is still marginal, i.e., it is far from the optimal

functioning condition, with respect to the safety margins and a longer time is needed to arrive to a completely safe state;

- ii) from the second comparison, important differences cannot be seen in the recovery time distribution from risk to marginal state (i.e., for fast recovery from risk states), since, in this work, a component in risk state cannot further degrade into a worse state. On the contrary, the impact of aftershocks is evident in the recovery from a marginal state to a healthy state since components in state 2 can degrade to state 1 more than once during the total recovery process. As a consequence, the time needed for the restoration of the healthy state increases when the occurrence of aftershocks is considered.

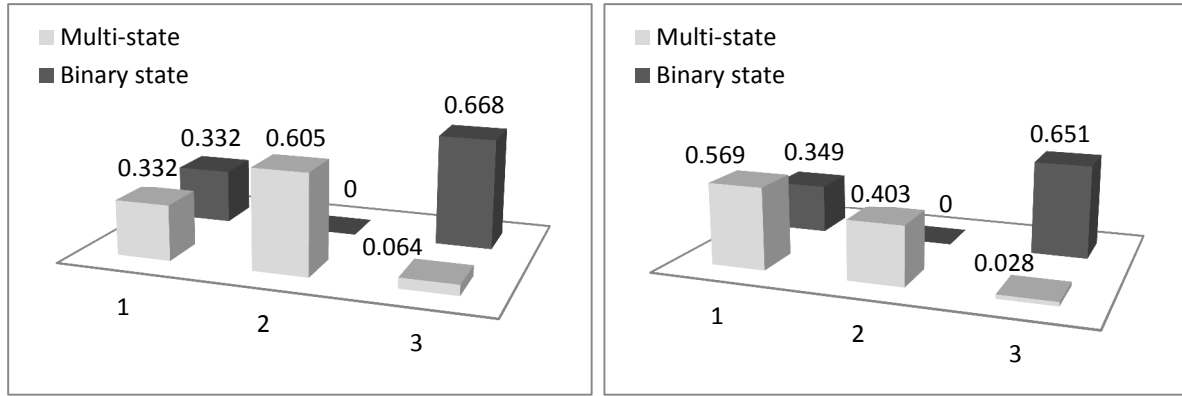


Figure 5.7: Left: estimate of the probability that the nuclear power plant reaches a risk (1), marginal (2) and healthy (3) state upon occurrence of an earthquake of moment magnitude equal to 5.5, in the case of multi-state (grey) and binary state (black) models. Right: same as Figure on the left, but considering also occurrence of subsequent aftershocks, in the case of multi-state (grey) and binary state (black) models.

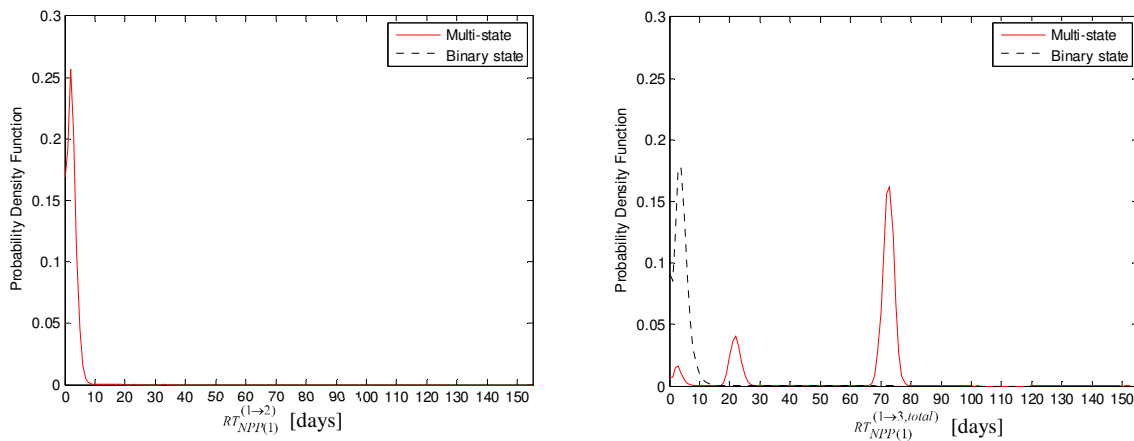


Figure 5.8: Left: probability density function (PDF) of the time (RT) necessary to restore the marginal state (2) of the nuclear power plant (NPP) from a risk state (1). Right: comparison of the probability density function (PDF) of the time (RT) necessary to restore the healthy state (3) of the nuclear power plant (NPP) from a risk state (1), in the case of a multi-state (solid line) and binary state (dashed line) model.

6. APPLICATION 2: CRITICAL INFRASTRUCTURES

In this Chapter, the case studies considered with respect to critical infrastructures (CIs) risk analysis under a system-of-systems (SoS) viewpoint are briefly illustrated, the corresponding system representations adopted are shown and the main results are provided. For further details the reader is referred to the corresponding papers V and VI of Part II.

Two case studies (hereafter referred to as “A” and “B”) are taken into account; case study A (Section 6.1) consists of two interdependent infrastructures (gas and electric power networks) and a supervisory control and data acquisition (SCADA) system connected to the gas network; case study B (Section 6.2) considers an electric power distribution network adapted from the IEEE 123 node test feeders. In both cases, we adopt a multi-state model to account for different degrees of damage of the components and we describe state transitions (random failures) by Markov and semi-Markov processes (Section 3.3.2). In addition, we include the epistemic uncertainty in the transition probability between different components states and in the mean of the holding time distributions by means of probability intervals.

Two quantities are used to evaluate the performance of CIs exposed to random failures:

- iii. from the robustness viewpoint, the probability of the product delivered to the demand nodes at steady state (steady state behavior of the SoS);
- iv. from the recovery viewpoint, the time needed to restore the system from the worst scenario (transient behavior of the SoS).

Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) is adopted to evaluate both quantities for the SoS A in presence of epistemic uncertainties, considering mainly a sequential importance of the demand nodes. Then, Hierarchical Graph is introduced to account for different priorities in the partitioning of the product; it is applied to both the case studies A and B to evaluate their robustness.

6.1. Case study A: interconnected gas and electricity networks

The case study is taken from [Nozick et al., 2005] and deals with two interconnected infrastructures, i.e., a natural gas distribution network and an electricity

arcs c_d and d_e); ii) the SCADA is always provided with electric power [Nozick et al., 2005].

This case study has been represented by the following system representations: GTST-DMLD (Section 6.1.1 and paper V of Part II) and Hierarchical Graph (Section 6.1.2 and paper VI of Part II).

6.1.1. GTST-DMLD representation and main results of case study A

In Figure 6.2, the GTST-DMLD of the case study of Figure 6.1 is shown.

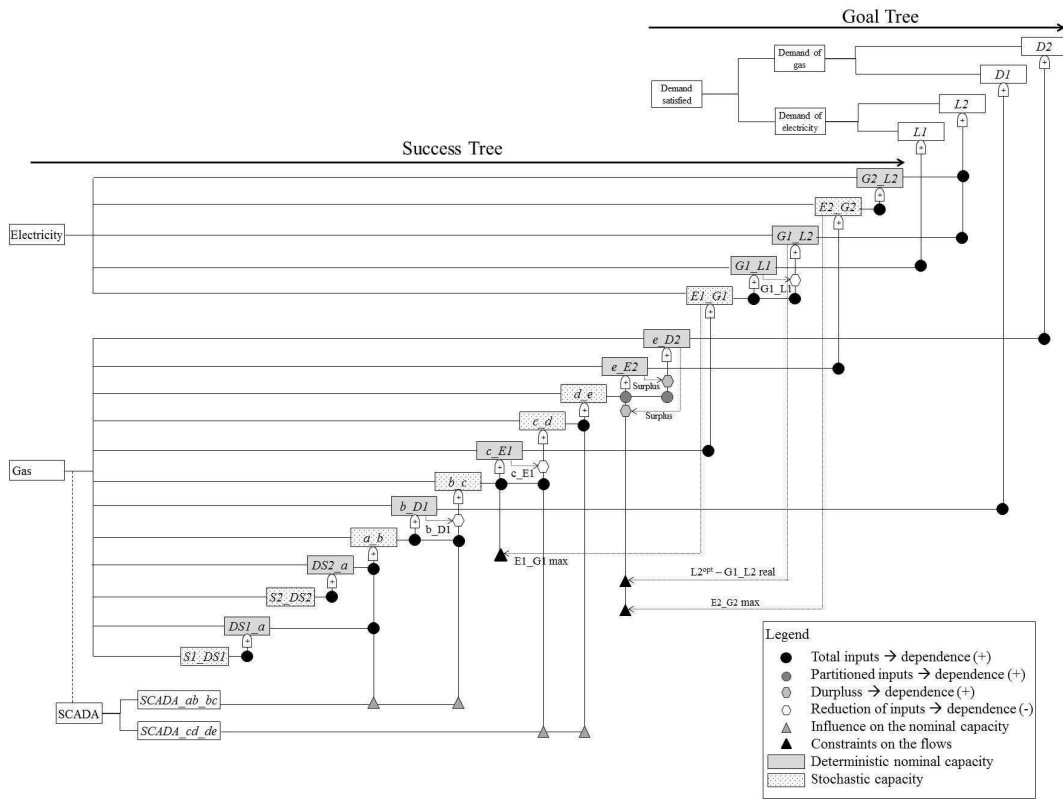


Figure 6.2: GTST-DMLD of the system of systems of Figure 6.1 (case study A).

The GT on the top represents the main goal of the SoS, related to the supply of the demands of gas and electricity: the objective is achieved if the corresponding nodes $D1$, $D2$, $L1$ and $L2$ receive the required amount of gas and electricity, respectively. In the present case study, we limit the analysis to the last level of the GT, i.e., we analyze the performance of each demand,

without investigating a global indicator of the SoS.

The ST is composed by the main hierarchies of the gas and electricity networks (that directly provide the demand nodes with gas and electricity to achieve the goal function) and by the support hierarchy of the SCADA system (that is needed for the control of the gas network and, therefore, it is not directly involved in the achievement of the goal function); given its support role, it is represented in a parallel dashed branch connected to the gas hierarchy.

The DMLD is represented by the relationships between objects of the ST or between objects of the ST and functions of the GT. It allows determining the goal function by the evaluation of all the dependencies from the bottom to the top of the diagram, following the rules explained in Section 2.5.2 for the dot-, hexagon- and triangle- dependencies.

Notice that, in this analysis by GTST-DMLD, the demand nodes are not given the same importance: in particular, DI is more important than LI ; on its turn, LI is more important than both $D2$ and $L2$ (which instead are equally important). These assumptions are made to illustrate and motivate the repartition of electricity and gas flows in the network and its representation in the GTST-DMLD, as introduced in Section 2.5.2.

The following analyses have been carried out: the evaluation of the SoS robustness and recovery properties, considering epistemic uncertainty in the transition probability between different components states and in the mean of the holding time distributions by means of probability intervals. The reader is referred to paper V of Part II for the results obtained.

6.1.2. Hierarchical Graph representation and main results of case study A

In Figure 6.3, the Hierarchical Graph of the case study of Figure 6.1 is illustrated.

The injection of product (i.e., gas) in the SoS is made through arcs $S1_DS1$ and $S2_DS2$, thus, they are located at the bottom of the diagram (Section 2.6). Since both arcs carry the product to node a , also the following links $DS1_a$ and $DS2_a$ are considered part of the inputs and reported at the bottom of the hierarchy. Four demand nodes, i.e., DI , $D2$ (gas) and LI , $L2$ (electricity), represent the goals of the analysis and they are explicitly located at the top of the diagram. The graph presents four hierarchical levels: in level 4, arc a_b , is reported since it supplies all the four demand nodes; in level 3, arc b_c is depicted, since it serves three demand nodes (i.e., LI , $L2$ and $D2$); in level 2, arcs c_E1 , $E1_G1$, c_d and d_e are

considered, since they supply two demand nodes: in particular, arcs c_E1 and $E1_G1$ supply $L1$ and $L2$, whereas arcs c_d and d_e serve $L2$ and $D2$; in level 1, there are the remaining arcs that are related just to one demand node: for example, e_E2 serves only node $L2$. The influence of the SCADA subsystems SUB1 on the arcs a_b and b_c and of the SCADA subsystem SUB2 on the arcs c_d and d_e is illustrated in the trapezoidal frames under the corresponding arcs.

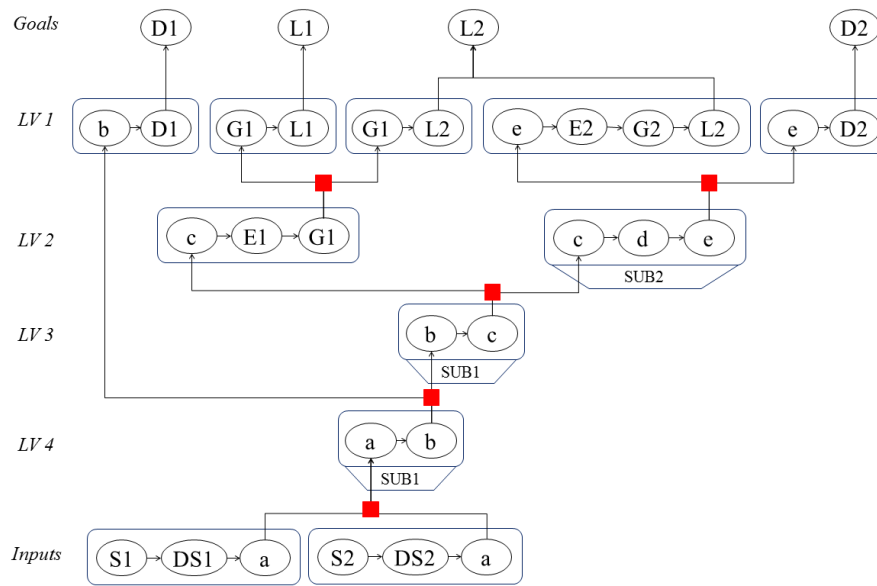


Figure 6.3: Hierarchical Graph of the system of systems depicted in Figure 6.1 (case study A); LV: Level.

The analyses carried out consist in the evaluation of the steady-state probabilities of the product delivered to the demand nodes ($D1$, $L1$, $L2$ and $D2$), considering different priorities in the partition of the product on the basis of sequential, proportional and equal importance of the demand nodes (see Section 2.6). For illustration purpose, Table 6.1 reports the steady-state probabilities of (i) delivering the (maximum, optimal) required product to the demand nodes (top) and (ii) delivering a quantity of product exceeding the 90% of the corresponding demands (bottom).

Table 6.1: Steady-state probabilities of (i) delivering the (optimal) required product to the demand nodes (top) and (ii) delivering a quantity of product exceeding the 90% of the corresponding demands (bottom).

Importance criterion	P(D1 = 100 [1000 cu. ft.])	P(L1 = 50 [10 MWh])	P(L2 = 40 [10 MWh])	P(D2 = 80 [1000 cu. ft.])
Sequential	0.9927	0.9867	0.9723	0.7526
Proportional	0.8195	0.7563	0.7563	0.7568
Equal	0.8205	0.9306	0.9678	0.9063

Importance criterion	P(D1 > 90 [1000 cu. ft.])	P(L1 > 45 [10 MWh])	P(L2 > 36 [10 MWh])	P(D2 > 72 [1000 cu. ft.])
Sequential	0.9927	0.9867	0.9723	0.7526
Proportional	0.9778	0.9155	0.9507	0.9160
Equal	0.9717	0.9482	0.9816	0.9063

In extreme synthesis, the results show that adopting different criterion, the product is partitioned in different ways in the network. Thus, the Hierarchical Graph representation can allow robustness analysis of SoS by considering different priorities of the demand nodes. The reader is referred to paper VI for a complete analysis of these results.

6.2. Case study B: electric power distribution network

Figure 6.4 shows the electric power distribution network here considered: it is adapted from the IEEE 123 nodes test feeder [IEEE, 2000] in the sense that regulators, capacitors, switches and feeders with length equals to zero are neglected. With these simplifications, the network is composed of 114 nodes: 1 generation point (node 115) and 113 load/transmission nodes. Node 61 of the original IEEE 123 node test feeders is missing here, since after the removal of switches and transformers it turns out to be an end node with load equal to zero. The arcs (i.e., the feeders) connect different nodes and distribute the power through the network.

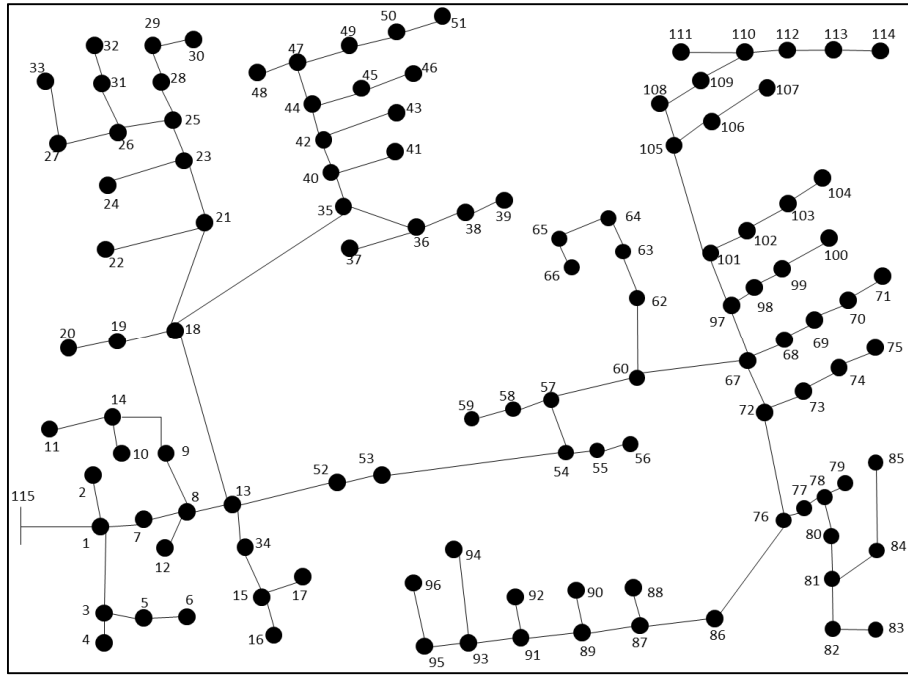


Figure 6.4: IEEE 123 node test feeders adapted to the purposes of the present analysis

Due to the large size of the electric power distribution network of interest, an unsupervised spectral clustering algorithm (see paper VI of Part II) is applied; it identifies five hierarchical clustering levels. The evaluation of the robustness of the system is performed at a given clustering level in combination with the Hierarchical Graph representation techniques (Section 2.6). The outcome of the procedure is represented by the steady-state probability distribution of the electricity delivered at the demand nodes (or clusters). In this case we consider a proportional importance of the clusters/demand nodes (see Section 2.6). A thorough explanation of the results can be found in paper VI of Part II.

7. CONCLUSIONS

Continuous advancements and innovations in technology are making our systems more and more efficient. This, however, comes at the “expenses” of increased complexity through the interconnectedness of technical, economic, societal, political, and cultural aspects. In spite of the fact that solutions to manage this complexity are sought, and mostly found, new risks arise making our systems vulnerable to component failures, natural and men-made hazards. To discover such risks, prevent them and protect from their consequences, performing reliability, risk, vulnerability and resilience analyses on one single system may not suffice given the interconnections with other systems that can generate cascading effects and cause failures and damages well beyond the single-system impact zone. For this reason, a system-of-systems (SoS) perspective is strongly advocated.

The risk analysis of safety-critical industrial installations and critical infrastructures (CIs) has been considered in this Ph. D. thesis, within a SoS framework. The three fundamental steps of the analysis have been tackled systematically, to understand and analyze the features of SoS relevant to their vulnerability and physical resilience:

- 1) system representation;
- 2) system modeling;
- 3) system behavior quantification (by simulation).

Proper attention in the modeling has been devoted to the presence of uncertainties due to randomness (aleatory uncertainty) and incomplete knowledge (epistemic uncertainty).

The representation step is aimed at capturing the SoS essential characteristics, with particular attention to dependences and interdependences, thus providing the support for the next steps of modeling and quantification.

Various representation techniques of literature, i.e., Fault Tree (FT), Muir Web, Hierarchical Modeling, Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD), have been investigated in this thesis, and in some cases extended in original ways to fit the purpose of the analysis; one representation method, namely the Hierarchical Graph, has been

CONCLUSIONS

developed ex-novo. The advantages and limitations of these techniques have been identified and discussed, on the basis of the analyses carried out and of the results obtained.

The modeling step is aimed at building a mathematical model of the system behavior. In the SoS framework here embraced, among the inputs of the model are the system components states, with respect to which we have considered both binary state and multi-state settings at the component level, that can result in binary and multi-state performances at SoS level.

The quantification of the model has been carried out by simulation of the system behavior for computing the output indicators of SoS performance. The Monte Carlo simulation method has been adopted to treat the aleatory uncertainty related to the random transitions of the system components among their reachable states and to the natural external events; interval analysis has been used to account for the epistemic uncertainty in the system model parameters, e.g., the component transition probabilities.

The SoS framework has been developed methodologically and analyses have been carried out with reference to two applications:

- plant external event risk assessment;
- CIs risk analysis.

The first application concerns the evaluation of the safety and physical resilience of critical installations (in particular a nuclear power plant, in the case of this thesis) subject to earthquakes and subsequent aftershocks. The critical installation is regarded within its supporting technical environment, i.e., embedded within the infrastructures that can provide support to keep or recover the plant safe condition upon the occurrence of the earthquake. The SoS (critical installation plus infrastructures) has been represented by FT, Muir Web, Hierarchical Modeling and GTST-DMLD. Both binary and multi-state models have been considered at the components level, distinguishing between structural states and functional states.

The results obtained have shown that the interdependent infrastructure services (e.g., the road transportation network) may play a role for the safety of a critical plant (as witnessed in recent accident occurrences, like Fukushima), and it is, thus, advisable to include them in the

analysis framework. In fact, they can provide additional support to the safety of the critical plant providing the service inputs required for its safe operation.

The systematic analysis through the levels of the Hierarchical Model has shown the possibility of identifying the contributions of the SoS individual elements (measured by their criticality importance measure) to the safety recovery time: this is useful information to point at which systems should be recovered early in the accident.

The multi-state model developed through the GTST-DMLD representation within the scheme of well-being analysis has allowed identifying marginal conditions of safety of the critical plant that may turn into a risk state under an additional triggering event. The knowledge acquired by such an analysis is relevant for the decision making process: a marginal condition may degrade to a risky one, whereas a complete safe state can mainly degrade to a marginal state. On the contrary, a binary state model does not allow these considerations since it does not distinguish different safety levels. .

The findings of this type of analysis can help to improve the structural/functional responses of the critical elements of the system, for improving the global physical resilience of the SoS so as to increase the safety of the critical plant. One may even imagine considering the optimization of some controllable characteristics of the SoS with the objective of increasing the safety of the critical plant. The multi-state model is a valid support for achieving these goals, provided that the definition of the structural and functional limit states is carefully addressed.

The second application relates to the evaluation of the robustness and recovery properties of CIs, and is illustrated by means of two case studies. The first one includes small-sized interconnected gas and electricity networks, and a supervisory control and data acquisition (SCADA) system, whereas the second one deals with a moderately large-size electric power distribution network. The GTST-DMLD and the Hierarchical Graph are adopted as system representation techniques. Multi-state models have been considered to describe variations in the network link capacities that occur stochastically according to Markov and semi-Markov processes. The epistemic uncertainty in the state transition probabilities has been propagated by combining Monte Carlo simulation and interval analysis.

The outcomes of the analyses are represented by the steady state behavior (referring to the robustness assessment) and the transient behavior (referring to the analysis of the recovery

CONCLUSIONS

capacity) of the SoS, including the treatment of the epistemic uncertainty. Both analyses are fundamental for giving insights about the performance of the SoS and provide suggestions about i) improving those arcs that more easily turns into damage states or ii) developing a more redundant network that allows the supply of the product from different paths. In addition, the possibility of giving different priorities to the different demand nodes has been explored to highlight how the product can be best partitioned in the network according to the characteristics and relative importances of the nodes (i.e., of the end-users). This is relevant when failures in the SoS occur and decisions have to be made with respect to the partitioning of product.

In general, the work carried out in the present Ph. D. thesis has investigated representation, modeling and simulation techniques for the vulnerability and physical resilience assessment of critical industrial plants and CIs within a SoS framework. Actually, the idea behind the consideration of the approaches here adopted is one of simplification of reality to a level of abstraction sufficient to understand the logic of functioning/dysfunctioning of the complex system and identify preliminary, relevant criticalities: “it is not always necessary for a model to *quantitatively* mimic the real system, but sometimes it is only necessary that the model *qualitatively* mimic it” [Pepyne et al., 2001]. In particular, the correct performance order of competing alternatives is sufficient to draw conclusions on design instead of obtaining the correct performance value [Pepyne et al., 2001].

In conclusion, the present work allows identifying preliminary vulnerabilities of critical installations and infrastructures, by providing information on their (physical) resilient capacity and guiding further analyses, e.g., a cost/benefit analysis that can rationally direct the investments of efforts and resources for improving the component responses within a comprehensive SoS approach.

Extensions of the present Ph. D. work can be carried out with respect to the three steps of system representation, modeling and simulation and uncertainty propagation.

System representation and modeling can be extended to include the role of humans on the system response to hazardous events. Actually, human aspects are part of every system and their lack of consideration may affect the overall system risk estimate.

CONCLUSIONS

Improvements in system simulation and uncertainty propagation can arrive from the consideration of efficient simulation techniques to perform robust functional failure probability estimations and uncertainty propagation, while reducing as much as possible the number of code simulations and the associated computational time.

REFERENCES

- Albert, R., Jeong, H., and Barabasi, A. L. (2000). "Error and attack tolerance of complex networks." *Nature*, 406(6794), 378-382.
- Alessandri, A., and Filippini, R. (2013). "Evaluation of Resilience of Interconnected Systems Based on Stability Analysis." *Critical Information Infrastructures Security - 7th International Workshop, CRITIS 2012*, B. M. Hämmerli, N. K. Svendsen, and J. Lopez, eds., Springer Berlin Heidelberg, 180-190.
- Apostolakis, G. (1990). "The Concept of Probability in Safety Assessments of Technological Systems." *Science*, 250(4986), 1359-1364.
- Apostolakis, G. E., and Lemon, D. M. (2005). "A screening methodology for the identification and ranking of infrastructure vulnerabilities due to terrorism." *Risk Analysis*, 25(2), 361-376.
- Aven, T. (2003). *Foundations of Risk Analysis*, Wiley.
- Aven, T. (2011). "On Some Recent Definitions and Analysis Frameworks for Risk, Vulnerability, and Resilience Response." *Risk Analysis*, 31(5), 693-697.
- Aven, T., and Zio, E. (2011). "Some considerations on the treatment of uncertainties in risk assessment for practical decision making." *Reliability Engineering & System Safety*, 96(1), 64-74.
- Aven, T., Zio, E., Baraldi, P., and Flage, R. (2014). *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*, Wiley.
- Azaïs, R., Bardet, J.-B., Génadot, A., Krell, N., and Zitt, P. A. (2014). "Piecewise deterministic Markov process - recent results." *ESAIM: Proceedings, EDP Sciences*, 276-290.
- Barlow, R. (1998). *Engineering Reliability*, ASA-SIAM Series on Statistics and Applied Probability.
- Barry, L. N. (1995). *Stochastic modeling: analysis and simulation*, McGraw-Hill, New York.
- Bedford, T., and Cooke, R. (2001). *Probabilistic Risk Analysis*, Cambridge University Press.
- Béjar, R., Latre, M. A., Nogueras-Iso, J., Muro-Medrano, P. R., and Zarazaga-Soria, F. J. (2009). "Systems of Systems as a Conceptual Framework for Spatial Data Infrastructures." *International Journal of Spatial Data Infrastructures Research*, 4, 201-217.
- Bhasin, K. B., and Hayden, J. L. (2009). "Communication and Navigation Networks in Space System of Systems." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 348-408.
- Billinton, R., and Karki, R. (1999a). "Application of Monte Carlo simulation to generating system well-being analysis." *IEEE Transactions on Power Systems*, 14(3), 1172-1177.
- Billinton, R., and Karki, R. (1999b). "Capacity reserve assessment using system well-being analysis." *IEEE Transactions on Power Systems*, 14(2), 433-438.
- Boardman, J., Pallas, S., Sauser, B. J., and Verma, D. (2006). "Report on system of systems engineering." Stevens Institute of Technology, Hoboken, NJ.
- Bobbio, A., Bonanni, G., Ciancamerla, E., Clemente, R., Iacomini, A., Minichino, M., Scarlatti, A., Terruggia, R., and Zendri, E. (2010). "Unavailability of critical SCADA

- communication links interconnecting a power grid and a Telco network." *Reliability Engineering & System Safety*, 95(12), 1345-1357.
- Borshchev, A., and Filippov, A. "From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools." *The 22nd International Conference of the System Dynamics Society*, Oxford, England.
- Bouchon, S. (2006). "The Vulnerability of interdependent Critical Infrastructures Systems: Epistemological and Conceptual State-of-the-Art." European Commission, Directorate-General Joint Research Centre, Institute for the Protection and Security of the Citizen, Ispra, Italy.
- Brissaud, F., Barros, A., Berenguer, C., and Charpentier, D. (2011). "Reliability analysis for new technology-based transmitters." *Reliability Engineering & System Safety*, 96(2), 299-313.
- Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and von Winterfeldt, D. (2003). "A framework to quantitatively assess and enhance the seismic resilience of communities." *Earthquake Spectra*, 19(4), 733-752.
- Buckley, J. J. (2004). "Fuzzy Markov Chains." Fuzzy probabilities and fuzzy sets for web planning, Springer, Berlin, 35-43.
- Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). "Catastrophic cascade of failures in interdependent networks." *Nature*, 464(7291), 1025-1028.
- Casalicchio, E., Bologna, S., Brasca, L., Buschi, S., Ciapessoni, E., D'Agostino, G., Fioriti, V., and Morabito, F. (2011). "Inter-dependency Assessment in the ICT-PS Network: The MIA Project Results." Critical Information Infrastructures Security, C. Xenakis and S. Wolthusen, eds., Springer Berlin Heidelberg, 1-12.
- Chou, C. C., and Tseng, S. M. (2010). "Collection and Analysis of Critical Infrastructure Interdependency Relationships." *Journal of Computing in Civil Engineering*, 24(6), 539-547.
- Cimellaro, G. P., Reinhorn, A. M., and Bruneau, M. (2010). "Framework for analytical quantification of disaster resilience." *Engineering Structures*, 32(11), 3639-3649.
- COM. (2004). "Communication from the Commission to the Council and the European Parliament: Critical Infrastructure Protection in the fight against terrorism." Commission of the European Community, Brussels, Belgium.
- Courtois, P. J. (1985). "On Time and Space Decomposition of Complex Structures." *Communications of the Acm*, 28(6), 590-603.
- D'Agostino, G., Bologna, S., Fioriti, V., Casalicchio, E., Brasca, L., Ciapessoni, E., and Buschi, S. "Methodologies for inter-dependency assessment." *Critical Infrastructure (CRIS), 2010 5th International Conference on*, 1-7.
- da Silva, A. M. L., de Resende, L. C., da Fonseca Manso, L. A., and Billinton, R. (2004). "Well-being analysis for composite generation and transmission systems." *IEEE Transactions on Power Systems*, 19(4), 1763-1770.
- Dahmann, J. S. (2009). "Systems Engineering for Department of Defense Systems of Systems." System of Systems Engineering: Innovations for the 21st Century, M. Jamshidi, ed., Wiley, 218-231.
- DeLaurentis, D. (2007). "Role of humans in complexity of a system-of-systems." Digital Human Modeling, V. G. Duffy, ed., Springer Berlin Heidelberg, 363-371.
- DeLaurentis, D. (2009). "Understanding Transportation as a System of Systems Problem." System of Systems Engineering: Innovations for the 21st Century, M. Jamshidi, ed., Wiley, 520-541.

REFERENCES

- Ding, Y., Zuo, M. J., Lisnianski, A., and Li, W. (2010). "A Framework for Reliability Approximation of Multi-State Weighted k-out-of-n Systems." *IEEE Transactions on Reliability*, 59(2), 297-308.
- Dueñas-Osorio, L., Craig, J. I., and Goodno, B. J. (2007a). "Seismic response of critical interdependent networks." *Earthquake Engineering & Structural Dynamics*, 36(2), 285-306.
- Dueñas-Osorio, L., Craig, J. I., Goodno, B. J., and Bostrom, A. (2007b). "Interdependent Response of Networked Systems." *Journal of Infrastructure Systems*, 13(3), 185-194.
- Dupuy, G. (1985). "Systèmes, réseaux et territoires." *Presses de l'École nationale des Ponts et Chaussées*.
- ESRI. (2014). "GIS dictionary." Environmental Systems Research Institute, <http://support.esri.com/en/knowledgebase/GISDictionary/term/thematic%20map>.
- Eusgeld, I., Nan, C., and Dietz, S. (2011). "System-of-systems" approach for interdependent critical infrastructures." *Reliability Engineering & System Safety*, 96(6), 679-686.
- Fang, Y., Pedroni, N., and Zio, E. (2014). "Optimization of Cascade-Resilient Electrical Infrastructures and its Validation by Power Flow Modelling." *Accepted for publication in Risk Analysis*.
- Ferson, S., and Ginzburg, L. R. (1996). "Different methods are needed to propagate ignorance and variability." *Reliability Engineering & System Safety*, 54(2-3), 133-144.
- Fioriti, V., D'Agostino, G., and Bologna, S. (2010). "On Modeling and Measuring Interdependencies among Critical Infrastructures." *2010 Complexity in Engineering: Compeng 2010, Proceedings*, 85-87.
- Fisher, D. A. (2006). "An Emergent Perspective on Interoperation in Systems of Systems." Carnegie Mellon Software Engineering Institute, Pittsburgh, PA.
- Flammini, F., Gagliione, A., Mazzocca, N., and Pragliola, C. (2009). "Quantitative Security Risk Assessment and Management for Railway Transportation Infrastructures." *Critical Information Infrastructures Security*, 5508, 180-189.
- Gheorghe, A. V., and Schlapfer, M. (2006). "Ubiquity of digitalization and risks of interdependent critical infrastructures." *2006 IEEE International Conference on Systems, Man, and Cybernetics, Vols 1-6, Proceedings*, 580-584.
- Gómez, C., Sánchez-Silva, M., and Dueñas-Osorio, L. (2011). "Clustering methods for risk assessment of infrastructure network systems." *Applications of Statistics and Probability in Civil Engineering - Proceedings of the 11th International Conference on Applications of Statistics and Probability in Civil Engineering*, M. Faber, J. Köhler, and K. Nishijima, eds., Taylor and Francis Group, London, 1389-1397.
- Gómez, C., Sanchez-Silva, M., Dueñas-Osorio, L., and Rosowsky, D. (2013). "Hierarchical infrastructure network representation methods for risk-based decision-making." *Structure and Infrastructure Engineering*, 9(3), 260-274.
- Gu, Y. K., and Li, J. (2012). "Multi-State System Reliability: A New and Systematic Review." *2012 International Workshop on Information and Electronics Engineering*, 29, 531-536.
- Guckenheimer, J., and Ottino, J. (2008). "Foundations for Complex Systems Research in the Physical Sciences and Engineering." Report from an NSF Workshop in September 2008.
- Haimes, Y. Y. (2009a). "On the Complex Definition of Risk: A Systems-Based Approach." *Risk Analysis*, 29(12), 1647-1654.
- Haimes, Y. Y. (2009b). "On the Definition of Resilience in Systems." *Risk Analysis*, 29(4), 498-501.

REFERENCES

- Haimes, Y. Y. (2012). "Modeling complex systems of systems with Phantom System Models." *Systems Engineering*, 15(3), 333-346.
- Helton, J. C., and Oberkampf, W. L. (2004). "Alternative representations of epistemic uncertainty." *Reliability Engineering & System Safety*, 85(1-3), 1-10.
- Henry, D., and Ramirez-Marquez, J. E. (2012). "Generic metrics and quantitative approaches for system resilience as a function of time." *Reliability Engineering & System Safety*, 99, 114-122.
- Hipel, K. W., Obeidi, A., Fang, L., and Kilgour, D. M. (2009). "Sustainable Environmental Management from a System of Systems Engineering Perspective." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 443-481.
- Hollnagel, E., Woods, D., and Levenson, N. (2006). *Resilience engineering: concepts and precepts*, Ashgate Publishing Limited, Abingdon, Oxon, GBR.
- Hu, Y. S., and Modarres, M. (1999). "Evaluating system behavior through Dynamic Master Logic Diagram (DMLD) modeling." *Reliability Engineering & System Safety*, 64(2), 241-269.
- Hu, Y. S., and Modarres, M. (2000). "Logic-based hierarchies for modeling behavior of complex dynamic systems with applications." *Fuzzy systems and soft computing in nuclear engineering*, D. Ruan, ed., Springer-Verlag, Berlin Heidelberg.
- Huang, Y. N., Whittaker, A. S., and Loco, N. (2011). "A probabilistic seismic risk assessment procedure for nuclear power plants: (I) Methodology." *Nuclear Engineering and Design*, 241(9), 3996-4003.
- Huzurbazar, A. V. (2005). "Flowgraph models: a Bayesian case study in construction engineering." *Journal of Statistical Planning and Inference*, 129(1-2), 181-193.
- IEEE. (2000). "IEEE power and energy society. Distribution test feeders." <http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html>.
- Jamshidi, M. (2009). "Introduction to System of Systems." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 1-20.
- Johansson, J., and Hassel, H. (2010). "An approach for modelling interdependent infrastructures in the context of vulnerability analysis." *Reliability Engineering & System Safety*, 95(12), 1335-1344.
- Jolly, S. D., and Muirhead, B. K. (2009). "System of Systems Engineering in Space Exploration." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 317-347.
- Kalos, M. H., and Whitlock, P. A. (1986). *Monte Carlo Methods Volume 1: Basics*, John Wiley & Sons, Inc., New York.
- Katina, P. F., and Keating, C. B. (2014). "Critical Infrastructures: Systems of systems as an integrating worldview." *Submitted to International Journal of Critical Infrastructures*.
- Klir, G. J., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River, NJ.
- Koonce, A. M., Apostolakis, G. E., and Cook, B. K. (2008). "Bulk power risk analysis: Ranking infrastructure elements according to their risk significance." *International Journal of Electrical Power & Energy Systems*, 30(3), 169-183.
- Kotov, V. (1997). "System of Systems as Communicating Structures." *Hewlett Packard Computer Systems Laboratory*, 1-14.
- Kröger, W. (2008). "Critical infrastructures at risk: A need for a new conceptual approach and extended analytical tools." *Reliability Engineering & System Safety*, 93(12), 1781-1787.
- Kröger, W., and Zio, E. (2011). *Vulnerable Systems*, Springer, London.

REFERENCES

- La Rocca, S., Guikema, S. D., Cole, J., and Sanderson, E. (2011). "Broadening the discourse on infrastructure interdependence by modeling the "Ecology" of infrastructure systems." *Applications of Statistics and Probability in Civil Engineering*, M. Faber, J. Köhler, and K. Nishijima, eds., London, 1905–1912.
- Laprie, J. C., Kanoun, K., and Kaaniche, M. (2007). "Modelling interdependencies between the electricity and information infrastructures." *Computer Safety, Reliability, and Security, Proceedings*, 4680, 54–67.
- Larsson, J. E. (1992). "Knowledge-based methods for control systems ", Lund Institute of Technology.
- Lee, E. E., Mitchell, J. E., and Wallace, W. A. (2007). "Restoration of services in interdependent infrastructure systems: A network flows approach." *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, 37(6), 1303–1317.
- Levitin, G., and Lisnianski, A. (1999). "Importance and sensitivity analysis of multi-state systems using the universal generating function method." *Reliability Engineering & System Safety*, 65(3), 271–282.
- Lind, M. (2011a). "An introduction to multilevel flow modeling." *Nuclear safety and simulation*, 2(1), 22–32.
- Lind, M. (2011b). "Reasoning about causes and consequences in Multilevel Flow Models." *Advances in Safety, Reliability and Risk Management*, C. Guedes Soares, ed., CRC Press, 2359–2367.
- Maier, M. W. (1996). "Architecting Principles for Systems-of-Systems." *Proceedings of the Sixth International Symposium of the International Council on Systems Engineering*, Boston, Massachusetts, 567–574.
- Maier, M. W. (1998). "Architecting principles for systems-of-systems." *Systems Engineering*, 1(4), 267–284.
- Manyena, S. B. (2006). "The Concept of Resilience Revisited." *Disasters*, 30(4), 434–450.
- Marseguerra, M., and Zio, E. (2002). *Basics of the Monte Carlo Method with Application to System Reliability*, LiLoLe - Verlag GmbH, Hagen, Germany.
- Modarres, M., Kaminskiy, M., and Krivtsov, V. (1999). *Reliability engineering and risk analysis: a practical guide*, CRC press, New York.
- Moore, D. A. (2006). "Application of the API/NPRA SVA methodology to transportation security issues." *Journal of Hazardous Materials*, 130(1–2), 107–121.
- Moteff, J., Copeland, C., and Fischer, J. (2003). "Critical Infrastructures: What Makes an Infrastructure Critical?", Congressional Research Service, The Library of Congress, Washington, DC.
- Muir, J., Limbaugh, R. H., and Lewis, K. E. (1985). "Notebook, July 27, 1869." *The John Muir Papers 1858–1957*, R. H. Limbaugh and K. E. Lewis, eds., Chadwyck-Healey, Alexandria, VA.
- Newman, M. E. J. (2002). "Spread of epidemic disease on networks." *Physical Review E*, 66(1).
- Nozick, L. K., Turnquist, M. A., Jones, D. A., Davis, J. R., and Lawton, C. R. (2005). "Assessing the performance of interdependent infrastructures and optimising investments " *International Journal of Critical Infrastructures*, 1(2–3), 144–154.
- Oliveira, D., Garrett, J. H., and Soibelman, L. (2009). "Spatial Clustering Analysis of Water Main Break Events." *Computing in Civil Engineering*, ASCE, 338–347.
- Ouyang, M. (2014). "Review on modeling and simulation of interdependent critical infrastructure systems." *Reliability Engineering & System Safety*, 121, 43–60.

REFERENCES

- Ouyang, M., Duenas-Osorio, L., and Min, X. (2012). "A three-stage resilience analysis framework for urban infrastructure systems." *Structural Safety*, 36-37, 23-31.
- Ouyang, M., Hong, L., Mao, Z.-J., Yu, M.-H., and Qi, F. (2009). "A methodological approach to analyze vulnerability of interdependent infrastructures." *Simulation Modelling Practice and Theory*, 17(5), 817-828.
- Parshani, R., Buldyrev, S. V., and Havlin, S. (2011). "Critical effect of dependency groups on the function of networks." *Proceedings of the National Academy of Sciences of the United States of America*, 108(3), 1007-1010.
- PCCIP. (1997). "Critical Foundations - Protecting America's Infrastructures." Washington, DC.
- Pederson, P., Dudenhoefter, D., Hartley, S., and Permann, M. (2006). "Critical infrastructure interdependency modeling: a survey of US and international research." *INL/EXT-06-11464*, Idaho National Laboratory, Idaho Falls.
- Pepyne, D. L., Panayiotou, C. G., Cassandras, C. G., and Ho, Y. C. (2001). "Vulnerability assessment and allocation of protection resources in power systems." *Proceedings of the 2001 American Control Conference, Vols 1-6*, 4705-4710.
- Piwowar, J., Chatelet, E., and Laclemece, P. (2009). "An efficient process to reduce infrastructure vulnerabilities facing malevolence." *Reliability Engineering & System Safety*, 94(11), 1869-1877.
- Pourgol-Mohamad, M., Mosleh, A., and Modarres, M. (2010). "Methodology for the use of experimental data to enhance model output uncertainty assessment in thermal hydraulics codes." *Reliability Engineering & System Safety*, 95(2), 77-86.
- Pourret, O., Collet, J., and Bon, J. L. (1999). "Evaluation of the unavailability of a multistate-component system using a binary model." *Reliability Engineering & System Safety*, 64(1), 13-17.
- Reed, D. A., Kapur, K. C., and Christie, R. D. (2009). "Methodology for Assessing the Resilience of Networked Infrastructure." *IEEE Systems Journal*, 3(2), 174-180.
- Rinaldi, S. A., Peerenboom, J. P., and Kelly, T. K. (2001). "Identifying, understanding, and analyzing critical infrastructure interdependencies." *IEEE Control Systems Magazine*, 21(6), 11-25.
- Ruzzante, S., Castorini, E., Marchei, E., and Fioriti, V. (2010). "A Metric for Measuring the Strength of Inter-dependencies." *Computer Safety, Reliability, and Security*, 6351, 291-302.
- Sage, A. P., and Cuppan, C. D. (2001). "On the Systems Engineering and Management of Systems of Systems and Federations of Systems." *Information-Knowledge-Systems Management*, 2(4), 325-345.
- Sahin, F. (2009). "Robotic Swarms as Aystem of Systems." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 482-519.
- Sallak, M., Schon, W., and Aguirre, F. (2013). "Reliability assessment for multi-state systems under uncertainties based on the DempsterShafer theory." *IIE Transactions*, 45(9), 995-1007.
- Sanderson, E. (2009). *Mannahatta: a natural history of New York City*, Abrams, New York.
- Schueller, G. I. (2007). "On the treatment of uncertainties in structural mechanics and analysis." *Computers & Structures*, 85(5-6), 235-243.
- Shibasaki, R., and Pearlman, J. S. (2009). "System of Systems Engineering of GEOSS." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 551-572.

REFERENCES

- Shih, C. Y., Scown, C. D., Soibelman, L., Matthews, H. S., Garrett, J. H., Dodrill, K., and McSurdy, S. (2009). "Data Management for Geospatial Vulnerability Assessment of Interdependencies in US Power Generation." *Journal of Infrastructure Systems*, 15(3), 179-189.
- Sridhar, P., Madni, A. M., and Jamshidi, J. (2009). "Advances in Wireless Sensor Networks: A case Study in System of Systems Perspective." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 275-292.
- Tolone, W. J., Johnson, E. W., Lee, S. W., Xiang, W. N., Marsh, L., Yeager, C., and Blackwell, J. (2009). "Enabling System of Systems Analysis of Critical Infrastructure Behaviors." *Critical Information Infrastructures Security*, 5508, 24-35.
- Trucco, P., Cagno, E., and De Ambroggi, M. (2012). "Dynamic functional modelling of vulnerability and interoperability of Critical Infrastructures." *Reliability Engineering & System Safety*, 105, 51-63.
- U.S. Department of Homeland Security. (2009). "National infrastructure protection plan - Partnering to enhance protection and resiliency."
- USNRC. (1997). "Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts." *NUREG/CR-6372, UCRL-ID-122160*, Livermore, CA.
- USNRC. (2009). "Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making." NUREG-1855. US Nuclear Regulatory Commission, Washington, DC.
- USNRC. (2013). "Glossary." <http://www.nrc.gov/reading-rm/basic-ref/glossary.html>.
- Wallace, W. A., Mendonça, D., Lee, E., Mitchell, J., and Chow, J. (2003). "Managing Disruptions to Critical Interdependent Infrastructures in the Context of the 2001 World Trade Center Attack." *Beyond September 11th: An Account of Post-disaster Research*, Institute of Behavioral Science, Natural Hazards Research and Applications Information Center, University of Colorado, Boulder, CO, 165-198.
- Wang, S., Hong, L., Chen, X., Zhang, J., and Yan, Y. (2011). "Review of interdependent infrastructure systems vulnerability analysis." *Intelligent Control and Information Processing (ICICIP)*, 2nd International Conference on, 446-451.
- White, G. F. (1974). *Natural hazards, local, national, global*, Oxford University Press.
- Wickramasinghe, N., Chalasani S, Boppana, R. V., and Madni, A. M. (2009). "Health Care System of Systems." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 542-550.
- Wilber, G. F. (2009). "Boeing's SoSE Approach to e-Enabling Commercial Airlines." *System of Systems Engineering: Innovations for the 21st Century*, M. Jamshidi, ed., Wiley, 232-256.
- Xue, J. N., and Yang, K. (1995). "Dynamic Reliability-Analysis of Coherent Multistate Systems." *IEEE Transactions on Reliability*, 44(4), 683-688.
- Yamijala, S., Guikema, S. D., and Brumbelow, K. (2009). "Statistical models for the analysis of water distribution system pipe break data." *Reliability Engineering & System Safety*, 94(2), 282-293.
- Zhang, H., Dufour, F., Dutuit, Y., and Gonzalez, K. (2008). "Piecewise deterministic Markov processes and dynamic reliability." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 222(4), 545-551.
- Zhang, P. C., and Peeta, S. (2011). "A generalized modeling framework to analyze interdependencies among infrastructure systems." *Transportation Research Part B-Methodological*, 45(3), 553-579.

REFERENCES

- Zimmerman, R. (2001). "Social Implications of Infrastructure Network Interactions." *Journal of Urban Technology*, 8(3), 97-119.
- Zio, E. (2007). *An Introduction to the Basics of Reliability and Risk Analysis*, World Scientific Publishing Co. Pte. Ltd.
- Zio, E. (2009a). *Computational methods for reliability and risk analysis*, World Scientific Publishing Co. Pte. Ltd., Singapore.
- Zio, E. (2009b). "Reliability engineering: Old problems and new challenges." *Reliability Engineering & System Safety*, 94(2), 125-141.
- Zio, E. (2013). *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, Springer, London.
- Zio, E. (2014). "Vulnerability and risk analysis of critical infrastructures." Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM2014) and Sixth International Symposium on Uncertainty, Modeling and Analysis (ISUMA2014), University of Liverpool, UK.
- Zio, E., and Aven, T. (2011). "Uncertainties in smart grids behavior and modeling: What are the risks and vulnerabilities? How to analyze them?" *Energy Policy*, 39(10), 6308-6320.
- Zio, E., Marella, M., and Podofillini, L. (2007). "A Monte Carlo simulation approach to the availability assessment of multi-state systems with operational dependencies." *Reliability Engineering & System Safety*, 92(7), 871-882.
- Zio, E., and Pedroni, N. (2011). "How to effectively compute the reliability of a thermal-hydraulic nuclear passive system." *Nuclear Engineering and Design*, 241(1), 310-327.

Part II

Paper I

A system-of-systems framework of Nuclear Power Plant Probabilistic Seismic Hazard Analysis by Fault Tree analysis and Monte Carlo simulation

E. Ferrario and E. Zio

Proceeding of the joint PSAM 11 and ESREL 2012 Conference, 25-29 June 2012, Helsinki.

A system-of-systems framework of Nuclear Power Plant Probabilistic Seismic Hazard Analysis by Fault Tree analysis and Monte Carlo simulation

E. Ferrario^a and E. Zio^{a,b}

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supélec, France

enrico.zio@ecp.fr, enrico.zio@supelec.fr

^bDepartment of Energy, Politecnico di Milano, Italy

enrico.zio@polimi.it

Abstract

We propose a quantitative safety analysis of a critical plant with respect to the occurrence of an earthquake, extending the envelope of the study to the interdependent infrastructures which are connected to it in a “system-of-systems” – like fashion. As a mock-up case study, we consider the impacts produced on a nuclear power plant (the critical plant) embedded in the connected power and water distribution, and transportation networks which support its operation. The Probabilistic Seismic Hazard Analysis of such system of systems is carried out by Fault Tree analysis and Monte Carlo simulation. As outcome of the analysis, the probability that the nuclear power plant reaches an unsafe state is computed for different earthquake’s epicentre distances and the contribution of the interdependent infrastructures to the safety of such critical plant is highlighted.

Keywords: System of systems, Probabilistic Seismic Hazard Analysis, Fault Tree Analysis, Monte Carlo simulation

1. INTRODUCTION

In the present paper, we consider the quantitative safety analysis of a nuclear power plant (NPP) with respect to the occurrence of an earthquake. We assume that internal emergency devices are available to provide safety for the plant upon such disturbances. However, accidental events in the industrial history, e.g., the recent Fukushima disaster (IAEA, 2011), have shown that the post-accident recovery of the full or partial safety of the plant may also depend on the infrastructures connected to it. In this view, the surrounding environment may or may not provide “resilience” properties.

Then, the analysis for the evaluation of the probability that a critical plant remains or not in a safe state must extend to the interdependent infrastructures connected to it, adopting a “system-of-systems” point of view. To this aim, both the intra-system and inter-systems dependencies, i.e., the dependencies between the components of a same infrastructure system and between the components of different infrastructure systems, respectively, are taken into account.

As a mock-up case study for the analysis, we consider the impacts of an earthquake produced on a nuclear power plant, extending the analysis to the power and water distribution, and to the transportation networks (the interdependent infrastructure systems) that can provide services necessary for keeping or restoring its safety. The case study is fictitious and highly simplified, intended only to illustrate the way of analyzing the problem under a “system-of-systems” viewpoint, with the effects of the interdependencies.

The assessment is performed by two main steps: first, a conceptual map is built to understand all the intra-system and inter-system dependencies among the components of the infrastructure systems connected to the nuclear power plant; then, a Fault Tree analysis is applied and the probability that the nuclear power plant enters in an unsafe state is computed by Monte Carlo simulation accounting for the contributions of both the internal emergency devices and the connected infrastructures.

The reminder of the paper is organized as follows. In Section 2, the basic concepts of Probabilistic Seismic Hazard Analysis are illustrated; in Section 3, the Fault Tree analysis by Monte Carlo simulation for Probabilistic Seismic Hazard Analysis is described; in Section 4, the case study and the results of the analysis are presented and discussed; in Section 5, conclusions are provided.

2. PROBABILISTIC SEISMIC HAZARD ANALYSIS

A Probabilistic Seismic Hazard Analysis (PSHA) consists of four procedural steps (EPRI, 2003; NUREG/CR-6372, 1997):

- 1) Earthquake source zones identification and characterization
- 2) Earthquake recurrence relationship definition
- 3) Ground motion attenuation relationship formulation
- 4) Exceedance probability calculation

The first step concerns the identification and characterization of the seismic sources in the proximity of the site of interest. It involves geological, seismological, geophysical data and scientific interpretations; as a consequence it is a critical part of the analysis and it is associated with considerable uncertainty (EPRI, 2003; NUREG/CR-6372, 1997).

The major outputs of the seismic hazard analysis are the seismic map that defines the seismic zones (areas where the earthquake sources have common characteristics like geometry, earthquake activity, earthquake annual recurrence rate), the probability distribution of the source-to-site distance and the identification of the maximum earthquake magnitude, i.e., the largest magnitude that a source can generate (EPRI, 2003; NUREG/CR-6372, 1997).

In the second step, the seismic earthquake recurrence relationship, i.e., the annual frequency of occurrence of a given magnitude event for each source, is defined. Typically, it is described by the Gutenberg-Richter law, $\log(n) = a - bm$ where n is the number of earthquakes with magnitude greater than m and a and b are parameters obtained by regression data analysis (EPRI, 2003; NUREG/CR-6372, 1997). This relation implies that the magnitude is exponentially distributed:

$$F_M(m) = 1 - e^{-\beta m} \quad (1)$$

where $\beta = \log_{10} b \approx 2.303b$ represents the relative frequency of smaller to larger events. Equation 1, however, is an unbounded probability distribution so that the magnitude can assume very high values, which are unrealistic and very low values, which are negligible. Therefore, the distribution is double-truncated by upper and lower bounds, m_{max} and m_{min} , respectively, and it is reformulated as follows (EPRI, 2003):

$$F_M(m) = \frac{1 - e^{-\beta(m-m_{min})}}{1 - e^{-\beta(m_{max}-m_{min})}} \quad (2)$$

The third step identifies the ground motion value at the site of interest, given the source-to-site distance and the magnitude. The higher the distance from the source, the lower is the ground motion value. Typical ground motion parameters are the peak ground acceleration and the spectral acceleration. Many ground motion equations have been defined on the basis of the earthquake and site characteristics (Douglas, 2011). They usually assume this expression (EPRI, 2003):

$$\log z' = C_1 + C_2 m + C_3 m C_4 + C_5 \log[r + C_6 \exp(C_7 m)] + C_8 r + g(source) + g(site) \quad (3)$$

where z' is the mean ground motion parameter, C_i , $i=1, \dots, 8$, are the regression coefficients, r is the source-to-site distance, m is the magnitude and $g(source)$ and $g(site)$ are terms that reflect the characteristics of the source and site, respectively.

For example, the peak ground acceleration is well described by (Ambraseys et al., 2005):

$$\log_{10} z' = C_1 + C_2 m + (C_3 + C_4 m) * \log_{10} \sqrt{r^2 + C_5^2} + C_6 S_S + C_7 S_A + C_8 F_N + C_9 F_T + C_{10} F_0 \quad (4)$$

where S_S and S_A represent the types of soil (soft, stiff or rock, when both variables are set to zero) and F_N , F_T and F_O describe the faulting mechanism (normal, thrust or odd).

In the fourth step, the probability of exceedance of ground motion in any time interval is computed by an analytical integration for each magnitude, distance and ground motion value by the following equation (EPRI, 2003):

$$\nu(z) = \sum_{i=1}^S \lambda_i(m_{\min}) \int_{m_{\min}}^{m_{\max}} \int_{r_{\min}}^{r_{\max}} f_{R_i}(r|m) f_{M_i}(m) P(Z > z|m, r) dm dr \quad (5)$$

where $i = 1, \dots, S$ represents the source zone, $f_{R_i}(r|m)$ and $f_{M_i}(m)$ are the probability density functions of the source-to-site distance and of the magnitude, respectively, $P(Z > z|m, r)$ is the probability of exceedance of the ground motion for each source zone, m_{\min} , m_{\max} , r_{\min} , r_{\max} are the lower and upper bounds of the magnitude and distance considered and $\lambda_i(m_{\min})$ is a rate that removes the contribution of earthquakes with magnitude lower than m_{\min} that is not significant.

A fragility evaluation is then carried out to provide the parameter values (i.e., the median acceleration capacity A_m and the logarithmic standard deviation due to randomness and to uncertainty in the median capacity β_r and β_u , respectively) for the fragility model that assumes this expression (EPRI, 2003):

$$f' = \Phi \left[\frac{\log \left(\frac{z'}{A_m} \right) + \beta_u \Phi^{-1}(Q)}{\beta_r} \right] \quad (6)$$

where f' is the conditional probability of failure for any given ground motion level z' and Q is the subjective probability of not exceeding a fragility f' .

3. FAULT TREE ANALYSIS AND MONTE CARLO SIMULATION FOR THE PROBABILISTIC SEISMIC HAZARD ANALYSIS UNDER A SYSTEM-OF-SYSTEMS FRAMEWORK

Consider a critical plant H (in our case the nuclear power plant) connected to N_S interdependent systems S_i (in our case the power and water distribution networks and the transport network), with N_{S_i} components linked by $K_{S(i,l)}$, $i, l = 1, \dots, N_S$, intra-system and inter-systems dependencies. The overall system is therefore composed by $N = \sum_{i=1}^{N_S} N_{S_i} + 1$ components, where the critical plant object of the analysis has been purposely explicated.

We wish to comprehensively evaluate the safety of the critical plant H with respect to the occurrence of an earthquake. To do this, in addition to the direct effects of the earthquake on H , we evaluate also the structural and functional responses of the N_{S_i} , $i = 1, \dots, N_S$, components and their impacts on the systems S_i , $i = 1, \dots, N_S$, and on the critical plant H . The approach taken is based on Fault Tree (FT) analysis and Monte Carlo (MC) simulation for Probabilistic Seismic Hazard Analysis (PSHA), and consists of the following operative steps:

1. build the fault tree of the top event “unsafe state of critical plant H ”, within a system-of-systems viewpoint that accounts also for the infrastructure connected to H ;
2. identify the minimal cut sets M_1, M_2, \dots, M_{mcs} ;
3. sample a magnitude value from the double truncated exponential distribution (2);
4. compute the ground motion value at each of the N_{S_i} , $i=1, \dots, N_S$, components of the systems S_i , $i=1, \dots, N_S$, and on the critical plant H , by equation 4;
5. compute the fragility, f , for all the components N_{S_i} , $i=1, \dots, N_S$, of the systems S_i , $i=1, \dots, N_S$, and for the critical plant H by equation 6; f is a vector of N values corresponding to the N components of the system;
6. sample a matrix $\{u_{j,k}\}$, $j=1, \dots, N_T$, $k=1, \dots, N$, where N_T is the number of simulations, of uniform random numbers in $[0,1)$;
7. determine the fault state matrix $\{g_{j,k}\}$, by comparing the fragility, f , with the matrix $\{u_{j,k}\}$, $j=1, \dots, N_T$, $k=1, \dots, N$: if $u_{j,k} < f_k$, $g_{j,k} = 1$; otherwise $g_{j,k} = 0$. When $g_{j,k}$ assumes value 1, the k -th component is affected by the earthquake, i.e., it enters a faulty state; otherwise, it survives. Each row of the matrix $\{g_{j,k}\}$ represents the states of the N system components of the system, i.e., its configuration;
8. assess the state of H for each row j of the matrix $\{g_{j,k}\}$ determined at step 7., i.e., for each system configuration sampled, by evaluating the system structure function $X_{H_j} = \Phi(g_{j,1}, \dots, g_{j,N}) = 1 - (1 - M_1)(1 - M_2) \dots (1 - M_{mcs})$, $j=1, \dots, N_T$. A vector $\{X_{H_j}\}$, is thus obtained, whose elements assume value 1 when the critical plant H is in an unsafe state and 0 otherwise;
9. estimate the probability of the critical plant H of being unsafe by computing the sample average of the values of the vector $\{X_{H_j}\}$, $j=1, \dots, N_T$.

The procedure above is repeated a large number of times for different values of earthquake magnitude.

4. CASE STUDY

We consider the evaluation of the safety of a fictitious nuclear power plant in response to earthquakes. We include in the analysis the responses of the interconnected systems that provide services which can aid keeping or restoring its safe state.

In Section 4.1, the description of the specific system studied is given; in Section 4.2, the results of its evaluation are provided, together with some critical considerations.

4.1. Description of the physical system and its view as a system of systems

The system under analysis is composed by a critical plant, i.e., a nuclear power plant, H , the power system, S_1 , that provides electrical energy for the running of the nuclear power plant, the water system, S_2 , that provides coolant useful to absorb the heat generated in the nuclear

power plant, and the road network, S_3 , relevant to the nuclear power plant for the transport of material and plant operators.

The water and power systems are subdivided into two independent parts, external and internal to the plant; the latter one represents the emergency system of the plant which needs to obviate at the absence of input from the main external system.

In Figure 1 the physical representation of the system is reported referring to a Cartesian plan (x, y) with origin in the river. Given the large scale system under analysis, two types of soil are considered, rock and soft soil. Figure 2 represents the spatial localization of the system shown in Figure 1 with reference to the reciprocal position of all the components (Figure 2, left) and to the position of the system, with respect to three earthquake's epicenters, A, B, C , (Figure 2, right). The distances on the axes are expressed in kilometers.

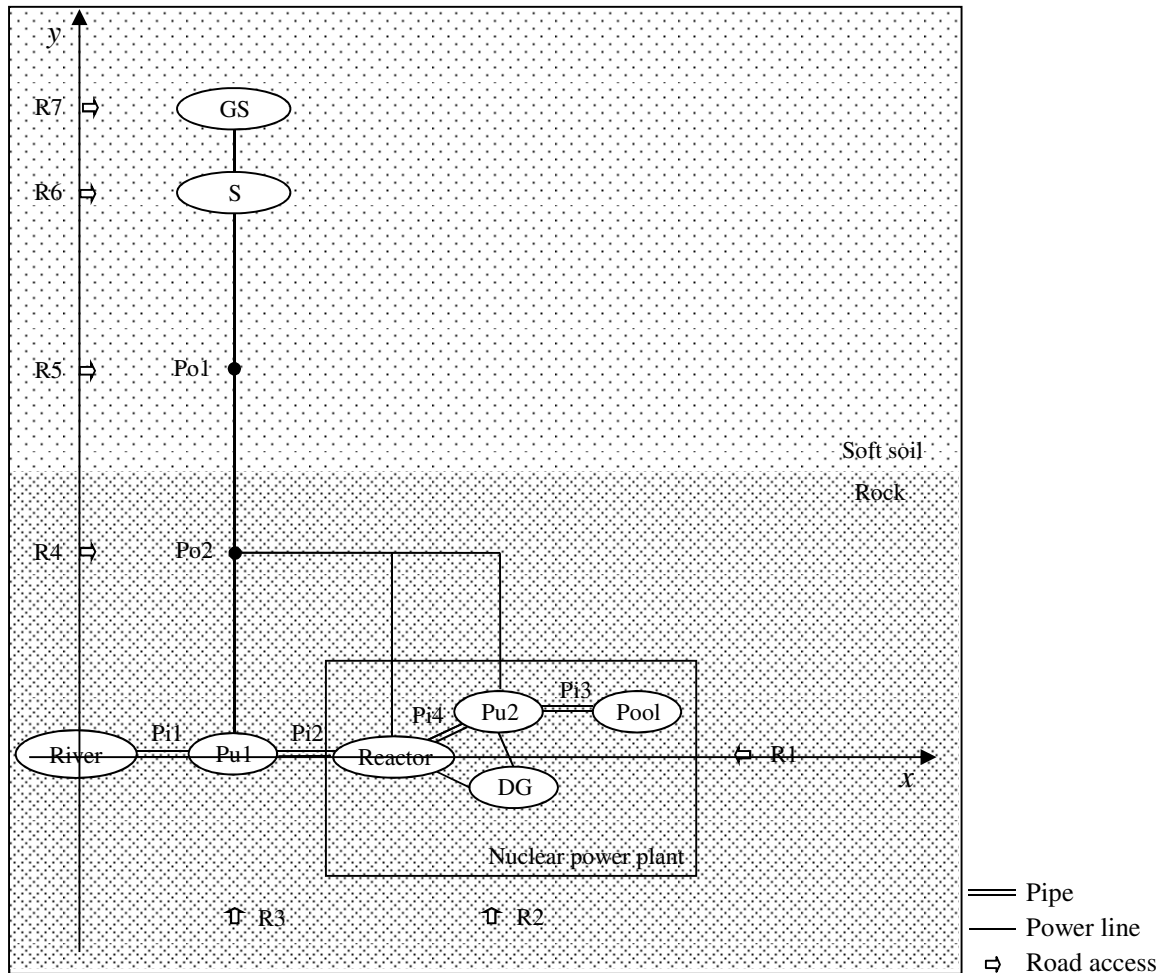


Figure 1: Physical representation of the system. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access.

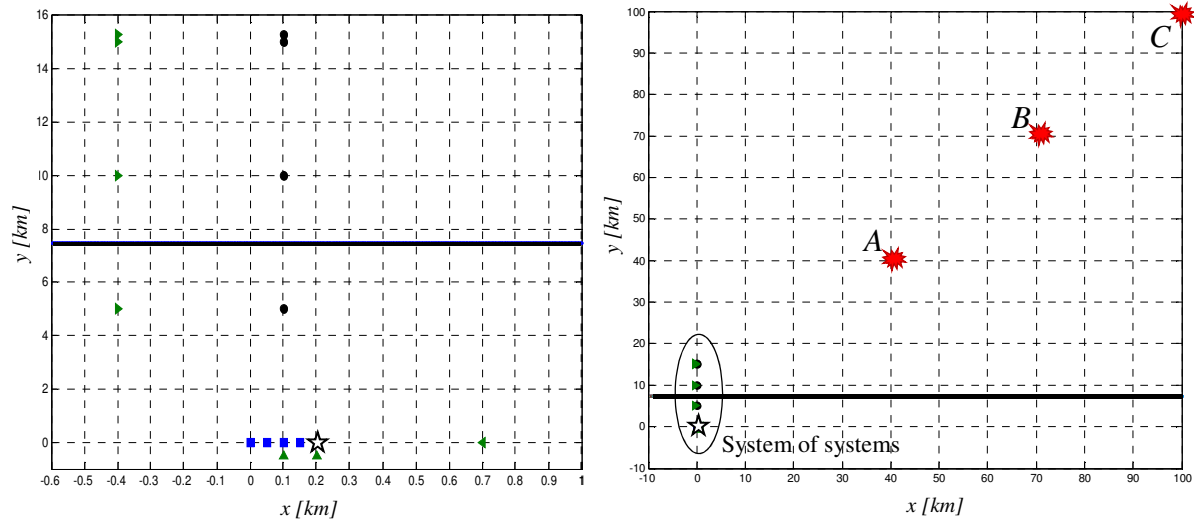


Figure 2. Left: spatial localization of the nuclear power plant (star) with respect to the components of the electric power system (circle, from top to bottom: Generation Station, Substation, Pole 1, Pole 2), water system (square, from left to right: River, Pipe 1, Pump 1, Pipe 2) and road transportation (triangle, from top to bottom and from left to right: R7, R6, R5, R4, R3, R2, R1). Right: spatial localization of the system of systems with respect to three earthquake's epicenters A(40, 40), B(70, 70), C(100, 100). The horizontal bold line in both Figures represents the division between soft soil (above the line) and rock (below the line).

In Figure 3, the system-of-systems representation is given by a conceptual map showing the components of the systems and their relationships, intra- and inter-systems. The intra-system dependencies are represented by the solid lines, the inter-system ones by dashed lines and those with the critical system by the bold lines.

The external water distribution system (Figure 3, left) is formed by a source of water (e.g., a river), a pump and pipes that carry the water. The failure probability of these elements depends on the type of soil and on the design and materials of construction. Operators are in charge of the maintenance of the structural elements and mechanical components.

The external power distribution system (Figure 3, center) is composed by the following elements: a generation station that produces the electrical energy, a substation that transforms the voltage from high to low, power lines and poles to support them, the type of soil on which the infrastructures rest and the operators that run the generation station and provide the maintenance for all its elements and components.

The components of the emergency water and power distribution systems inside the plant are shown in Figure 3 on the right. The first system is composed by the same elements of the correspondent external system except for the source of water that is an artificial reservoir, whereas the power system includes only the emergency diesel generators.

The elements considered for the transportation system are the roads (Figure 3, top). The state of this system is important for access of the materials and operators that are needed to restore the components required for the safe state of the critical plant.

Actually, in view of the methodological character of this work, for the sake of simplicity, the influence of the design construction and materials, the supply of fuel and materials for plant operation, and the maintenance tasks are not included in the analysis. The power lines, being

aerial elements and therefore being not directly affected by an earthquake, are also not considered. Finally, the assumption is made that the river is not perturbed by the earthquake so that it is a source of water always available.

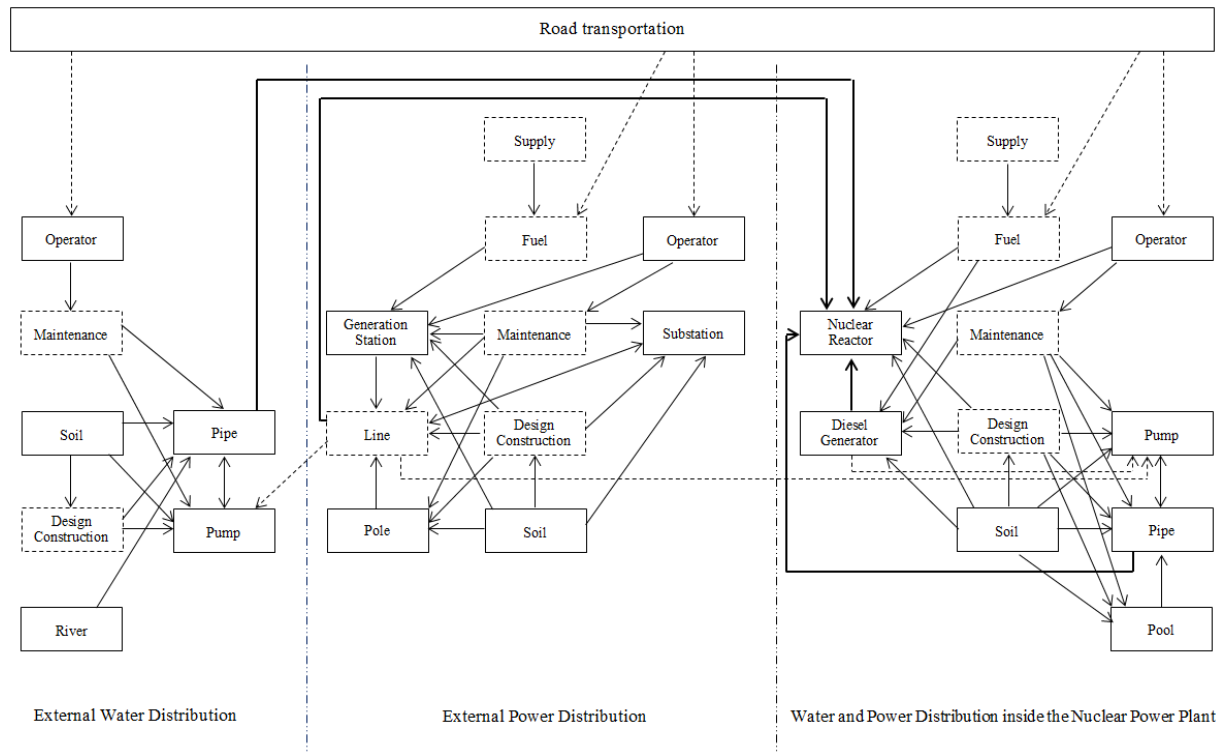


Figure 3. System of systems: the elements in the dashed box are not considered in the present study; the links represent the intra-systems dependencies (solid lines), the inter-systems dependencies (dashed lines) and the dependencies of the nuclear power plant on its interconnected systems (bold lines).

The inter-system dependencies are modeled as links connecting components of the three systems, S_i , $i=1, 2, 3$, (Figure 3, dashed lines); these links are conceptually similar to those linking components of the individual systems (intra-systems dependencies), and are considered bidirectional with respect to the “flow” of dependence between the connected systems. For example, the water system depends on the power system as the pump needs electrical energy to work. This component receives the electrical energy from the external power distribution network; on the contrary, it is assumed that the pump inside the nuclear power plant can obtain energy from both the external and internal power systems.

The road transport network allows access to the components of the power and water systems for transporting material (e.g., fuel) and operators for operation and/or maintenance.

The transport system is composed by seven interdependent road access points to the components of the power and water systems. One access is provided for the components outside the nuclear power plant, whereas two accesses are provided for the elements inside (Figure 1).

Note that, in the present study, the road assumes the function of “reserve component”, since we assume that elements that fail can be immediately repaired/replaced if the access to it through the road system does not fail (recovery times are not considered).

Figure 4 shows the primary levels of the fault tree built for the analysis. It depicts the main causes of occurrence of the top event, i.e., critical plant H is in an unsafe state, which are the lack of energy and/or water supply by both the internal and external systems. For space limitation, the triangular elements in the tree are not detailed in the Figure. By way of example, the fault tree of the pump of the external water distribution is reported in Figure 5. This component is unable to provide its service if 1) the component itself fails and at the same time there is no road access to repair it or 2) it does not receive the necessary inputs of electrical energy and water. The external energy system fails if one or more of its elements fail, i.e., the generation station, the substation or the poles, whereas the external water system cannot provide water for the pump if a rupture of the pipe occurs.

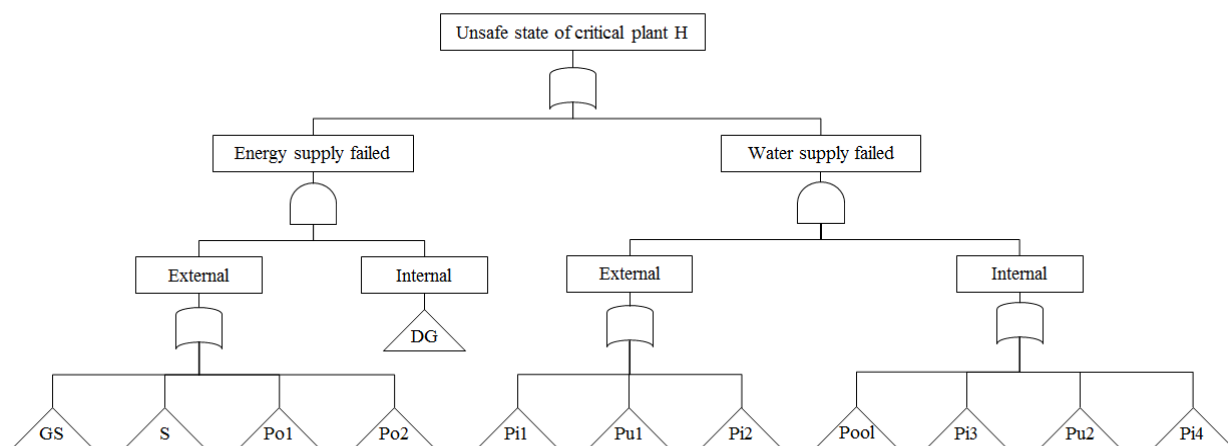


Figure 4. Fault tree of the system of systems: upper levels. The elements in the triangular shape are not detailed. NPP: Nuclear Power Plant, GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator.

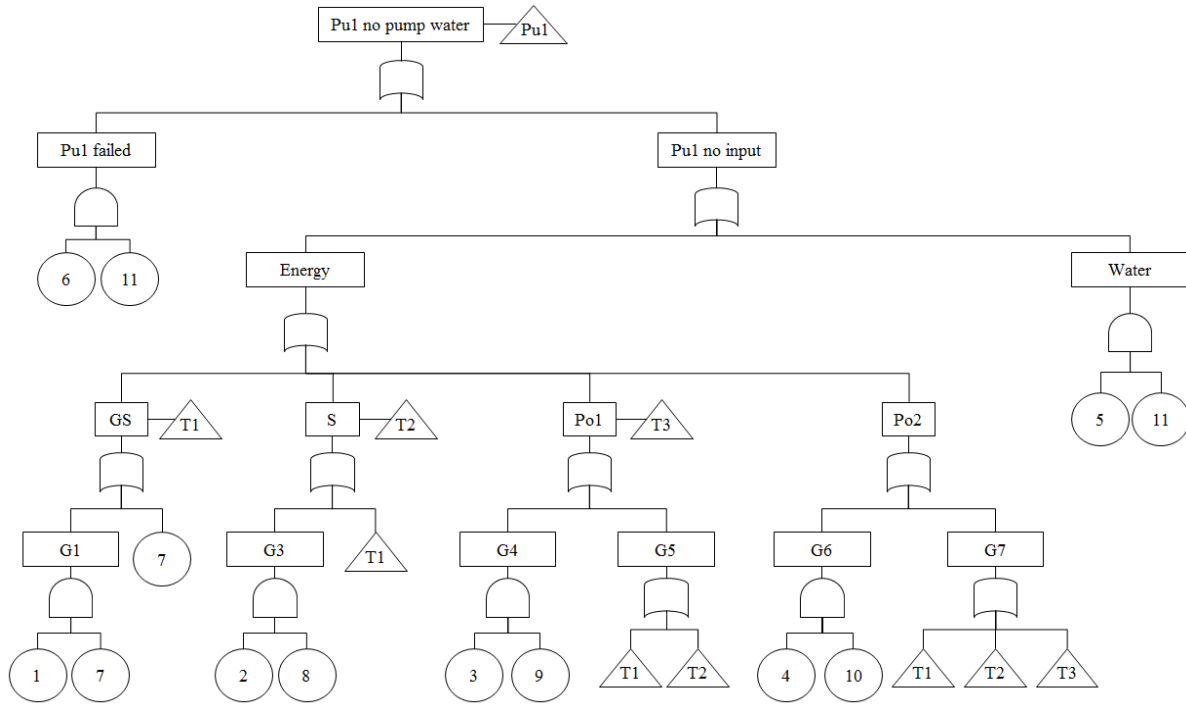


Figure 5. Fault tree details for the failure event of the component “pump” of the external water distribution system. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, R: Road access. The numbers in the circles are referred to the failure of the components: 1 → GS, 2 → S, 3 → Po1, 4 → Po2, 5 → Pi1, 6 → Pu1, 7 → R7, 8 → R6, 9 → R5, 10 → R4, 11 → R3.

4.2. Results and limitations

Figure 6 shows the results of the evaluation by Monte Carlo simulation of the fault tree presented in the previous Section, within the Probabilistic Seismic Hazard Analysis procedure introduced in Section 3. For each magnitude level sampled from a truncated exponential probability distribution (2) with lower threshold $m_{min}=5$ and upper bound $m_{max}=7$, the estimate of the probability of the nuclear power plant to reach an unsafe condition, is computed. The analysis is carried out for the three earthquake’s epicenters, A, B, C, shown in Figure 2. As expected, the higher the distance, the lower is the probability that the safety of the nuclear power plant is not guaranteed.

Figure 7 shows the comparison between the probabilities that the nuclear power plant turns into an unsafe condition after the occurrence of an earthquake at epicenter A(40, 40) considering it both as an isolated component provided with its emergency devices (case of independence) and as a part of the system of systems, with the supporting infrastructures (case of dependence). It can be seen that with the given assumptions and data, the probabilities to reach an unsafe state computed in case of dependence are slightly lower than those computed in case of independence, particularly at low earthquake magnitudes. This result shows that in principle the infrastructures in the surrounding of the critical plant can contribute to its resilience, offering additional possibilities for maintaining (or restoring) a safe condition, particularly when the earthquake magnitude is small.

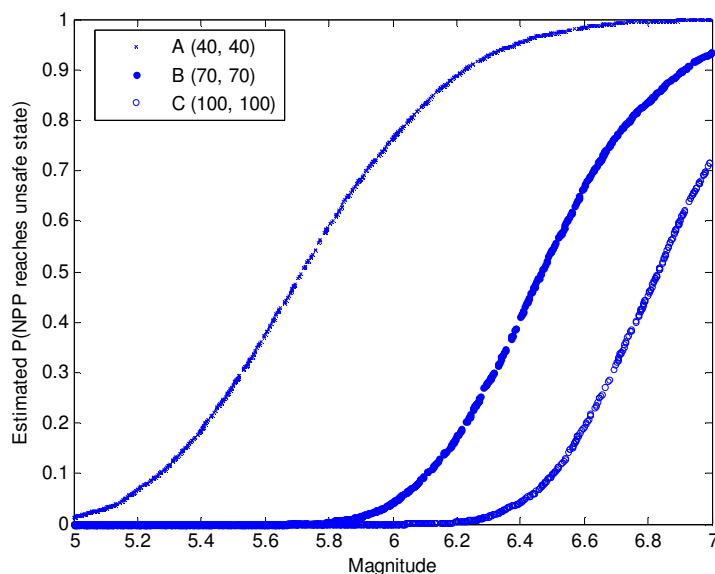


Figure 6. Estimate of the probability that the nuclear power plant reaches an unsafe state upon occurrence of an earthquake of a given magnitude, on the basis of different source-to-site distances. With reference to the map of Figure 2, the coordinates of the earthquake's epicenters considered are A(40, 40), B(70, 70), C(100, 100), expressed in kilometers.

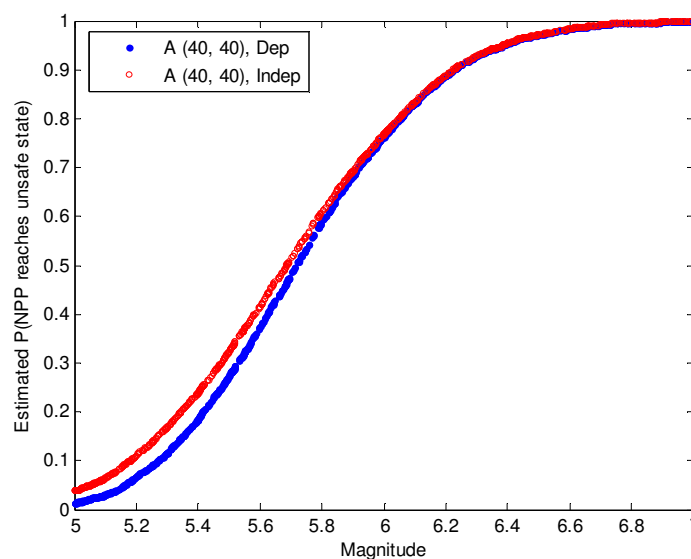


Figure 7. Comparison between the results of the MC simulation in the case of dependence of the nuclear power plant on the connected infrastructure systems, and in the case of independence, for earthquake's epicenter A(40, 40).

5. CONCLUSIONS

We have used Fault Tree analysis and Monte Carlo simulation to perform a quantitative safety analysis of a nuclear power plant under the risk of occurrence of an earthquake, extending the

area of study to its interconnected infrastructure systems (water and power distribution, and transportation networks) within a system-of-systems analysis framework.

The results obtained highlight that the interdependent infrastructure systems may play a role by providing additional support to the safety of a nuclear power plant, and it thus seems advisable to include them in the safety analysis.

More generally, the modeling framework proposed can be used to analyze the contribution to the safety of any critical plant, provided by the interdependent infrastructure systems connected to it.

Future work will concern the inclusion of the time for recovery of a failed component and the duration of emergency service supply.

REFERENCES

- Ambraseys, N.N., Douglas, J., SARMA, S.K. and Smit, P.M. (2005) Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration, *Bulletin of Earthquake Engineering*, 3, 1-53.
- Douglas, J. (2011) Ground-motion prediction equations 1964-2010, *Pacific Earthquake Engineering Research Center*, Final Report BRGM/RP-59356-FR.
- EPRI (2003) *Seismic Probabilistic Risk Assessment Implementation Guide*, EPRI TR-1002989.
- Harary, F. (1995) *Graph Theory*. Perseus, Cambridge, MA
- IAEA (2011) *The great east Japan earthquake expert mission – IAEA international fact finding expert mission of the Fukushima Dai-ichi NPP accident following the great east Japan earthquake and Tsunami*, Mission Report.
- NUREG/CR-6372 (1997) *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, UCRL-ID- 122160 Vol. 1.

Paper II

A framework for the system-of-systems analysis of the risk for a safety-critical plant exposed to external events

E. Zio and E. Ferrario

Reliability Engineering & System Safety 114 (2013) 114-125

A framework for the system-of-systems analysis of the risk for a safety-critical plant exposed to external events

E. Zio^{a,b} and E. Ferrario^a

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supélec, France

enrico.zio@ecp.fr, enrico.zio@supelec.fr

^bDepartment of Energy, Politecnico di Milano, Italy

enrico.zio@polimi.it

Abstract

We consider a critical plant exposed to risk from external events. We propose an original framework of analysis, which extends the boundaries of the study to the interdependent infrastructures which support the plant. For the purpose of clearly illustrating the conceptual framework of system-of-systems analysis, we work out a case study of seismic risk for a nuclear power plant embedded in the connected power and water distribution, and transportation networks which support its operation. The technical details of the systems considered (including the nuclear power plant) are highly simplified, in order to preserve the purpose of illustrating the conceptual, methodological framework of analysis. Yet, as an example of the approaches that can be used to perform the analysis within the proposed framework, we consider the Muir Web as system analysis tool to build the system-of-systems model and Monte Carlo simulation for the quantitative evaluation of the model. The numerical exercise, albeit performed on a simplified case study, serves the purpose of showing the opportunity of accounting for the contribution of the interdependent infrastructure systems to the safety of a critical plant. This is relevant as it can lead to considerations with respect to the decision making related to safety critical-issues.

Keywords: External Events Risk Analysis, System of Systems, Muir Web, Monte Carlo Simulation, Seismic Probabilistic Risk Assessment.

1. INTRODUCTION

The focus of this work is to look at the safety of a critical plant challenged by the occurrence of an external event, like earthquake, flooding, high wind, fire, lightning, volcanic eruption [1]. We assume that properly designed and dimensioned, “internal” emergency devices are available to assure safety of the critical plant upon such disturbances, even in the case of unavailability of the infrastructure services. However, accidental events in the industrial history, e.g., the recent Fukushima disaster [2], show that the post-accident assurance of the full or partial safety of a critical plant in the emergency conditions of an external disastrous event may also need to resort to exceptional recovery means and actions, which need to be supported by the infrastructures connected to the critical plant. In other words, upon the occurrence of the destructive event, the surrounding environment may or may not be left in the conditions to provide “emergency assistance” to the critical plant. Indeed, considering an external event which is spatially distributed, its impact may not affect only the critical plant itself but also the areas around it, with possible damages to the interdependent infrastructures that may or may not be capable of providing the services needed for keeping or restoring the safety of the critical plant.

With these considerations, we propose to extend the boundaries of the analysis for evaluating its safety by adopting a “system-of-systems” framework of analysis [3], [4], [5], [6], [7], [8], [9] which includes the interdependent infrastructures connected to the plant, in addition to its internal emergency devices, and thus examines also the “resilience” properties offered from the overall structure of the system of systems in which the plant is embedded. For the purpose of illustrating the concepts underlying the extended framework, as quantitative indicator we consider the probability that a critical plant remains or not in a “safe state” upon the occurrence of an external event. Safe state is here used to indicate that the plant is in a condition that does not cause health and/or environmental damages.

To provide an example of application of the proposed framework, we consider a case study regarding the occurrence of an earthquake (the external event) impacting on a system of systems which contains a nuclear power plant (the critical plant) that is provided with the needed emergency infrastructure systems. For exemplary purposes, the framework extends the analysis to the power and water distribution, and to the transportation networks (the interdependent infrastructure systems) that can provide services necessary for keeping or restoring the safety of the critical plant. The case study is used only to illustrate the concepts behind the framework of analysis under a system-of-systems viewpoint: for this reason, it is fictitious and admittedly highly simplified in the technical aspects (including those of the nuclear power plant and its safety systems) and strong, possibly at times not too realistic, assumptions are made to keep the focus on the methodological framework. In spite of this, for completeness the modeling and numerical evaluation are carried out by resorting to powerful methods of system analysis and stochastic simulation: Muir Web [10] and Monte Carlo simulation [11], [12], [13].

Muir Web is a system analysis technique to model a complex system and the relationships among its elements. In the context of ecological human community, in which it has been first

introduced [10], traditionally only the major interactions are taken into account in the system modeling: for example, with reference to the food chain, only the connections between predator and prey are usually considered, whereas other relevant and influencing relationships exist between organisms, e.g., one species may take cover for another, and other factors contribute to the food chain, e.g., abiotic elements like water, sun, soil, rainfall, wind [10]. By the representative power of Muir Web, the traditional picture of dependencies is extended through a graph where the nodes represent all the system elements (e.g., species and abiotic factors in the ecological case) and the edges represent their dependency structure.

The concept of Muir Web has been recently applied also to infrastructure systems, exploiting some similarities which exist between the ecological and the infrastructure networks [14]: both are large scale systems with complex interactions and can fail when an external event occurs. In the case of infrastructure systems, the nodes of the web are system components, e.g., a pump, and other factors which influence the infrastructure state, e.g., a stable soil with respect to seismic hazard.

In the case study worked out in this paper, the assessment is performed in two main steps: first, a conceptual map in the form of a Muir Web is built to represent all the dependencies and interdependencies among the components of the infrastructure systems connected to the nuclear power plant; then, Monte Carlo simulation is applied to compute the probability that the nuclear power plant enters in an unsafe state, accounting for the contributions of both the internal emergency devices and the connected infrastructures to support the safety of the critical plant. An analysis is also made to find how much the interdependencies would affect the safety of the nuclear power plant.

The remainder of the paper is organized as follows. In Section 2, the basic concepts of External Event Risk Assessment are introduced, with some specifics of Seismic Probabilistic Risk Assessment (SPRA) for positioning the illustrative case study used to exemplify the methodology; in Section 3, the Monte Carlo simulation framework for SPRA is described for providing the basic ground of the quantification technique used in the case study; in Section 4, the complete assessment of the case study by Muir Web and Monte Carlo simulation is presented, and the results discussed; in Section 5, conclusions and reflections are shared and future developments are provided.

2. NATURAL EXTERNAL EVENT RISK ASSESSMENT

The framework of the analysis considers natural external events as hazard inputs. They can include earthquake, flooding, high wind, fire, lightning, volcanic eruption [1]. The common characteristics of these hazards are the large-scale impacts on the environment and the considerable amount of uncertainty related with their occurrence and their intensity.

To include them in the safety analysis of a critical plant, the following steps should be performed [1]:

- a. Assessment of the frequency of the hazards (i.e., estimation of the frequency of exceedance of particular intensities) and analysis of the loads associated;
- b. Analysis of the plant response to the hazards (i.e., fragilities);

- c. Analysis of the impacts of the hazards on the plant.

To proceed in the analyses, properties and parameters of the hazards should be defined. For example, for seismic hazard, parameters like intensity of the earthquake, ground motion and frequency content (e.g., response spectrum) should be defined; for flooding, relevant parameters include water level of the river/lake, duration of flood and water velocity; for high winds, the dynamic loads from gusts and rotation velocities from tornadoes should be given.

In the present paper, the seismic hazard has been taken into account within a framework of Seismic Probabilistic Risk Assessment (SPRA) based on three parts [15], [16]:

- a. Seismic Hazard Analysis to compute the probabilities of occurrence of different levels of earthquake ground motion at a site of interest.
- b. Seismic Fragility Evaluation to identify the seismic capacity of a component in terms of its conditional probability of failure for any given ground motion level.
- c. System Analysis to integrate the outputs of the hazard and fragility analyses for evaluating the impacts of the earthquake on the infrastructure of interest.

The first part, which is traditionally developed as Probabilistic Seismic Hazard Analysis (PSHA), consists of four procedural steps [15], [16], [17]:

- 1) Identification and characterization of the earthquake source;
- 2) Definition of the earthquake recurrence relationship, i.e., the annual frequency of occurrence of a given magnitude event for each source, typically described by the Gutenberg-Richter law [18] that implies a double-truncated exponential distribution for the magnitude¹ [21], [22]:

$$F_M(m) = \frac{1 - e^{-\beta(m - m_{min})}}{1 - e^{-\beta(m_{max} - m_{min})}} \quad (1)$$

where β represents the relative frequency of smaller to larger events and m_{max} and m_{min} are the upper and lower bounds of the magnitude, respectively, that avoid the high values which are unrealistic and the low values that are negligible.

- 3) Formulation of the ground motion attenuation relationship that identifies the ground motion value at the site of interest, e.g., the peak ground acceleration, given the source-to-site distance and the magnitude. The higher the distance from the source, the lower is the ground motion value. The following relationship described by Ambraseys [23] has been embraced in this paper:

$$\log_{10} z' = C_1 + C_2 m + (C_3 + C_4 m) * \log_{10} \sqrt{r^2 + C_5^2} + C_6 S_s + C_7 S_A + C_8 F_N + C_9 F_T + C_{10} F_O \quad (2)$$

where m is the magnitude, r is the source-to-site distance, S_s and S_A represent the types of soil (soft, stiff or rock, when both variables are set to zero) and F_N , F_T and F_O describe the faulting mechanism (normal, thrust or odd).

¹ The magnitude scale typically used is the moment magnitude defined by Kanamori [19]. For medium size earthquakes it is similar to the Richter values [20].

- 4) Computation of the exceedance probability of ground motion in any time interval by an analytical integration for each magnitude, distance and ground motion value.

In the second part of the SPRA, a fragility evaluation is carried out to provide the parameter values (i.e., the median acceleration capacity A_m and the logarithmic standard deviation due to randomness and to uncertainty in the median capacity β_r and β_u , respectively) for the fragility model that assumes this expression [15]:

$$f' = \Phi \left[\frac{\log\left(\frac{z'}{A_m}\right) + \beta_u \Phi^{-1}(Q)}{\beta_r} \right] \quad (3)$$

where f' is the conditional probability of failure for any given ground motion level z' and Q is the subjective probability of not exceeding a fragility f' .

In the third part, an evaluation of the consequences of the seismic event to the infrastructure under analysis is traditionally performed by the development of event trees and logic models for each event tree top event [15]. In this work, we adopt a Muir Web representation and Monte Carlo simulation for this evaluation.

3. MUIR WEB REPRESENTATION AND MONTE CARLO SIMULATION FOR SEISMIC PROBABILISTIC RISK ASSESSMENT WITHIN A SYSTEM-OF-SYSTEMS FRAMEWORK

In this Section, the objective of the Muir Web modeling is first illustrated (Section 3.1) and the Muir Web representation of a system of systems is then given (Section 3.2). Finally, the operative steps of the Monte Carlo (MC) simulation method for Seismic Probabilistic Risk Assessment (SPRA) are illustrated (Section 3.3).

3.1. Muir Web modeling

The Muir Web is a network representation technique, which allows analysis by graph theory. It has been introduced to explicitly represent the structure of dependence of the physical elements on factors which influence their functionalities. It is a tool to visualize, capture and understand the relations among physical elements and factors of a system, and it organizes the knowledge in a comprehensive way through its multi-dimensional structure. It is inspired by the view of John Muir, the famous naturalist [24]: “When we try to pick out anything by itself we find that it is bound fast by a thousand invisible cords that cannot be broken, to everything in the universe”. The original purpose behind the introduction of the Muir Web was to recreate the landscape and the wildlife of the city of Mannahatta four hundred years ago to see how that place was before it became a city and to reimagine the city’s development taking into account the natural cycles and processes [10]. For this aim, the Muir Web can be converted into maps by an iterative computer program that works through all the relationships and find the right layers in a Geographic Information System [10].

In the Muir web representation there is no difference among the types of relations: they are depicted by arrows that are directed from an element to another dependent on it. Applying it to an engineered system means to consider all the other elements (physical, operational, organizational, etc.) which each single element depends on, including, for example, the type of soil, the maintenance task, the presence of operators, etc. One main objective of the Muir Web is to visualize all the connections among elements. This gives the basis for performing further analysis to characterize the types of relations, the way a failure of an element can affect the state of another connected element, the elements with significant influence on the system functionality and those with little influence.

Table 1 states the advantages and limitations of the Muir Web with respect to other techniques, i.e., Fault Tree Analysis and Hierarchical modelling.

In synthesis, the Muir Web seems to offer a flexible and easy way of representation, with the possibility of managing a large number of nodes and relationships. In addition, extending the analysis to the level of factors (operational, organizational, etc.) that influence the physical elements, it shows the capability of crossing disciplinary boundaries in an integrated representation; then, it can interface straightforwardly with other modelling tools to generate maps representing the spatial localization of the infrastructures, including their interdependences and all related characteristics. This flexibility and ease of representation is paid by the large amount of information needed to further characterize the model and the need of further analyses to associate the logic structure of the system and evaluate it in terms of the quantities of interest, which may be costly simulations for large systems.

Table 1: Comparison of the Muir Web with Fault Tree Analysis and Hierarchical modelling.

	Muir Web		Fault Tree Analysis		Hierarchical modelling	
	Advantages	Limitations	Advantages	Limitations	Advantages	Limitations
Qualitative analysis	Representing the invisible: in addition to the physical elements, the representation includes the factors (operational, organizational, etc.) which the physical elements depend on. The associated knowledge is organized in a comprehensive way through a multi-dimensional structure.	A large amount of information and competences of different disciplinary fields are needed to build the representation.	The physical elements are represented in a well-defined structure, according to the logic of the system, that leads to the identification of the Minimal Cut Sets.	Additional factors (operational, organizational, etc.) are not included. The exhaustive identification and manipulation of the Minimal Cut Sets can be difficult for large systems.	The system is broken up according to its parts and it is analyzed at different levels of detail.	Additional factors (operational, organizational, etc.) are not included.
	Easy to build the network answering the question "why" to identify the depending elements.			Difficult to build the fault tree, in particular in the case of large number of components and complicated logic, dependencies, etc.	Easy to build the hierarchy, identifying the parts of the system with increasing level of detail.	
	Extendable/flexible: the addition of a new component is possible without changing all the structure.	A further analysis is needed to identify the logic structure of the system.	The structured representation allows a rigorous and transparent analysis.	The addition of a new component can change the structure.	Analyzing the system at different levels of detail allows a good understanding of the system structure.	The addition of a new component can change the hierarchy.
	The representation clearly illustrates the dependencies among the components: arrows are directed from one element to another dependent on it. In addition, there is the possibility of including the strength of the relationship ² .		The representation is clear and allows understanding which combinations of components cause the failure of the top event. The modeling is straightforward via few, simple, logic operators.	The Boolean-logic based approach does not allow considering the strength of the relationships.	The representation is clear and allows understanding the composition of the system at different levels of detail.	The representation does not show the relationships among the components at each level of the hierarchy and a further analysis is needed for that.
	Possibility to be converted into maps, resorting to the support of Geographic Information Systems.					
Quantitative analysis	Simulation: propagation of failures in the network.	High computational cost of simulation for large systems.	Numerical calculation of the probability of occurrence of the top event by transforming the logical structure into an equivalent probability form.	Difficulty in treating the dynamics of failures.	Simulation: propagation of failures bottom-up through the hierarchy.	High computational cost of simulation for large systems.

² The strength of the relationship has not been included in the present paper but it has been considered in Sanderson [10], characterizing the relationships by modifiers like "especially" or "often".

3.2. Muir Web representation of a system of systems

For a general representation of the system of systems based on the Muir Web framework (Figure 1), let us consider a plant H that is critical from the safety viewpoint, i.e., if it is not provided with the necessary service inputs it can reach a condition which causes health and environmental damages. The state of the critical plant H is the state of its critical element, E . Connections exist to N_S interdependent systems S_i , $i = 1, \dots, N_S$, numbered in order in such a way that the first q are those inside the plant and the last $N_S - q$ belong to systems outside the plant. The systems internal to the plant S_i , $i = 1, \dots, q$, are designed to provide inherent safety, i.e., the input services required to keep E in a safe state. Each system is composed by N_{c_i} , $i = 1, \dots, N_S$, components and the overall system of systems is therefore formed by $N = N_{in} + N_{out}$ components, where $N_{in} = \sum_{i=1}^q N_{c_i}$ and $N_{out} = \sum_{i=q+1}^{N_S} N_{c_i}$. For the sake of clarity of the representation, we distinguish the intra-system and inter-systems links, i.e., the links among components of the same system and of different systems, respectively, into two types here called “direct dependency” and “support dependency” on the basis of their physical meaning: for the first type, when a component fails, its direct neighbors also fail; for the second one, when a component fails, it does not cause the failure of its neighbors because it assumes the role of “support”, i.e., it is useful to the neighbors when these fail for other reasons. In addition, the links between the interconnected infrastructure systems and the critical plant have been considered. They represent unidirectional dependency, but if a connected system fails, it does not mean that the critical plant fails too; identification, specification and joint analysis of all these dependencies have to be performed to determine their effect on the critical plant, as explained in Section 3.3. The Muir Web of Figure 1 shows an example in which the element E (star) of the critical plant H (dotted-rectangular shape) is connected to four interdependent systems S_i , $i = 1, \dots, 4$ with $N_{c_1} = 5$, $N_{c_2} = 6$, $N_{c_3} = 7$ and $N_{c_4} = 3$ components. The systems S_1 and S_2 are inside the plant and the systems S_3 and S_4 are outside. The direct dependencies are represented by solid lines, the support ones by dashed lines and the connections to the critical plant by bold lines.

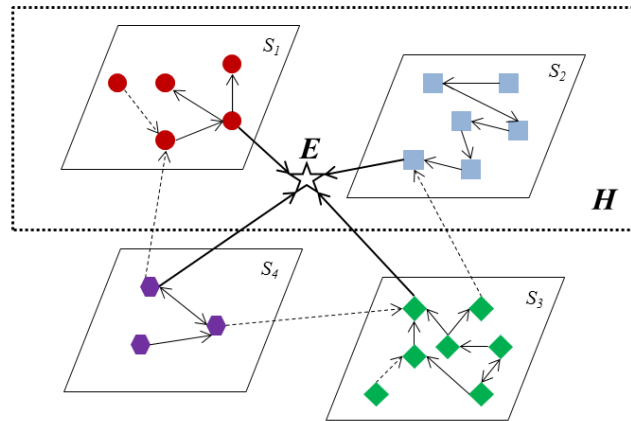


Figure 1: Muir Web representation of a system of systems made of a critical plant, H (dotted-rectangular shape) whose safety is identified in the state of its critical element E , and four interdependent systems S_i , $i = 1, \dots, 4$, whose elements (represented by circles, squares, rhombs and hexagons, respectively) are connected by direct

dependencies (solid lines) and support dependencies (dashed lines). The systems S_1 and S_2 are inside the critical plant, whereas the systems S_3 and S_4 are outside. The links to the critical element E (star) of the critical plant are the bold lines.

3.3. Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a system-of-systems framework

Within the system-of-systems analysis framework here purported, we wish to evaluate the safety of the critical plant H exposed to the risk from earthquakes occurrence, accounting not only for the direct effects of the earthquake on H , in particular on the internal interconnected systems S_i , $i = 1, \dots, q$, inside H , but also for the structural and functional responses of the N_{c_i} , $i = q + 1, \dots, N_S$, components and their impacts on the systems S_i , $i = q + 1, \dots, N_S$, and eventually on the critical plant H through the interconnected web of the underlying dependency structure. To do this, we adopt the Muir Web representation of the system of systems and Monte Carlo (MC) simulation for the quantitative SPRA evaluation [25]. The simulation procedure consists of the following operative steps:

1. sample a magnitude value from the double truncated exponential distribution by equation 1;
2. compute the ground motion value at each of the N_{c_i} , $i = 1, \dots, N_S$, components of the systems S_i , $i = 1, \dots, N_S$, by equation 2;
3. compute the fragility, f , for all the components N_{c_i} , $i = 1, \dots, N_S$, of the systems S_i , $i = 1, \dots, N_S$, by equation 3; f is a vector of N values corresponding to the N components of the system;
4. sample a matrix $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N$, where N_T is the number of simulations, of uniform random numbers in $[0,1)$;
5. determine the fault state matrix $\{g_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N$, by comparing the fragility, f , with the matrix $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N$: if $u_{j,k} < f_k$, $g_{j,k} = 1$; otherwise $g_{j,k} = 0$ for $j = 1, \dots, N_T$ and $k = 1, \dots, N$. When $\{g_{j,k}\}$ assumes value 1, the k -th component is affected by the earthquake, i.e., it enters a faulty state; otherwise, it survives. Each row of the matrix g represents the states of the N system components;
6. determine the state of the critical plant H , considering:
 - a. the impact of the earthquake on H , i.e., taking into account the interconnected systems S_i , $i = 1, \dots, q$, inside the plant.
 - b. the impact of the earthquake both on H , i.e., taking into account the interconnected systems S_i , $i = 1, \dots, q$, inside the plant, and on the interconnected systems S_i , $i = q + 1, \dots, N_S$, outside the plant.

The state of H is identified by the analysis of the states of the N_{in} components of the systems S_i , $i = 1, \dots, q$, for the case a., and of the N components of the systems S_i , $i = 1, \dots, N_S$, for the case b., together with the analysis of the dependence of H from the services provided by the systems, as represented in the Muir Web model. The structure of dependence represented in the Muir Web drives the identification of the

functional logic relations among the components within each system (intra-system links) and among different systems (inter-system links). Knowledge of these relations allows identifying the state of the critical plant H on the basis of the states of the components of its connected systems and their logic links: trivially, if two components of a system are connected in series (*Figure 2, left*), they should be both in an operational state to guarantee its functioning; on the contrary, if they are connected in parallel (*Figure 2, right*), at least one of them should work.



Figure 2: Example of series (left) and parallel (right) configurations between two components.

The state of H is evaluated through the analysis of the logic connections between the components, as explained above, for each row of the matrix $\{g_{j,k}\}$, i.e., for all the k states determined at step 5, where $k = 1, \dots, N_{in}$ and $k = 1, \dots, N$ for the case a. and b. above, respectively, and for all the simulations j , $j = 1, \dots, N_T$. A vector $\{h_j\}$, $j = 1, \dots, N_T$, is then recorded, whose element assumes value 1 when the critical plant H is in an unsafe state and 0 otherwise;

7. estimate the probability of the critical plant H of being unsafe by computing the sample average of the values of the vector $\{h_j\}$, $j = 1, \dots, N_T$.

The procedure above is repeated a large number of times for different values of earthquake magnitude.

Note that the components are considered with binary states: fully operative or completely damaged and also the critical plant can assume only two states: fully operative or totally failed. This approximation is not realistic and leads to pessimistic results: multi-state modeling may be considered for a more realistic description, where different degrees of damage are contemplated.

This framework of analysis should also allow considering the duration of the recovery actions to restore the safe state of the critical plant. This aspect is not here examined, but it is intended to be the objective of future work.

4. CASE STUDY

We consider the analysis of the safe state of a nuclear power plant (the critical plant), provided with proper internal emergency devices, in response to an earthquake (the external event). The nuclear power plant is considered in a safe condition if it does not cause health and environmental damages, i.e., if it does not release radioactive material to the

environment; to maintain this state it must be provided with electrical and water inputs to absorb the heat that it generates. The boundaries of the analysis extend to the responses of the external interconnected systems that provide inputs necessary to keep or restore the plant in the safe state. In Section 4.1, the description of the specific system studied is given under a number of assumptions aimed at simplifying the problem to the level needed to convey the key aspects of the conceptual system-of-systems framework, while maintaining generality.

When an earthquake occurs, the critical plant may not receive the input necessary to be kept in, or restored to, a safe state due to the direct impact on its internal emergency devices (safety systems) and to the damages to the external interconnected infrastructures. In general, two quantities can be of interest with regard to the loss of functionality of the various components of the system of systems embedding the critical plant, upon the occurrence of a damaging external event:

- the probability that the critical plant remains in a safe condition given the possible failure configuration of the components;
- the recovery time of the safe state of the critical plant, i.e., the duration of the recovery actions needed to bring the components back to the level of functionality required to restore the safe condition of the plant.

We limit the analysis to the first quantity, leaving the computation of the second one for future work, and in Section 4.2 we provide the results of the evaluation, accompanied by some critical considerations.

4.1. Description of the physical system and its view as a system of systems

The system under analysis is composed by a critical plant, i.e., a nuclear power plant, H , and five interconnected infrastructure systems, S_i , $i = 1, \dots, 5$, that provide services that can serve keeping the safe state of the nuclear power plant. The systems S_1 and S_2 are inside the nuclear power plant, whereas the systems S_3 , S_4 and S_5 are outside. The external systems are the power system, S_3 , that provides electrical energy, the water system, S_4 , that provides coolant useful for absorbing the heat generated in the nuclear power plant, and the road network, S_5 , that is important for the transport of material and plant operators. The internal systems, S_1 and S_2 , are the power and water systems, respectively, that represent the emergency systems of the plant which needs to obviate at the absence of input from the main systems.

In *Figure 3*, the physical representation of the system is reported referring to a Cartesian plan (x, y) with origin in a river. The nuclear reactor is the element of the nuclear power plant that must be provided with the necessary inputs to assure the safe state of the entire plant. *Table 2* reports the fragility parameters A_m , β_r and β_u , adopted in this analysis, for illustration purposes. The values for the pump and the pipe components have been taken from [26] and [27], respectively, whereas the others fragility parameters have been assumed arbitrarily by the authors to perform the study with different values. Given the large scale system under analysis, two types of soil are considered, rock and soft soil. *Figure 4* represents the spatial localization of the system shown in *Figure 3* with reference to the reciprocal position of all the components (*Figure 4*, left) and to the position of the system, with respect to three

earthquake's epicenters, A , B , C (Figure 4, right). The distances on the axes are expressed in kilometers.

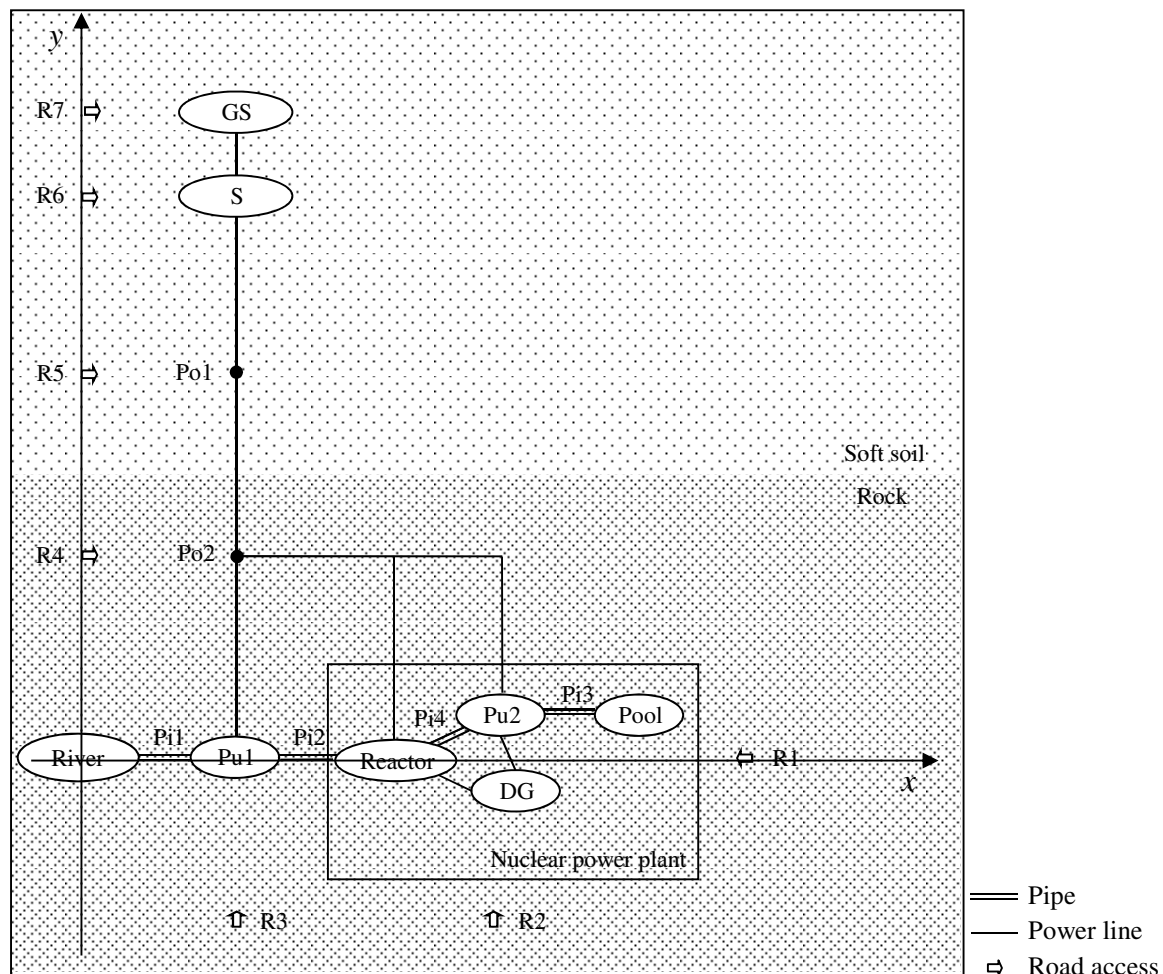


Figure 3: Physical representation of the system. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access.

Table 2: Fragility parameters used in the present work.

	A_m	β_r	β_u
Generation station	0.7	0.3	0.1
Substation	0.9	0.4	0.3
Power Pole	0.8	0.2	0.2
Diesel Generator	0.7	0.4	0.2
Pipe	1.88	0.43	0.48
Pump	0.2	0.2	0.3
Pool	0.2	0.1	0.1
Road	0.3	0.3	0.2

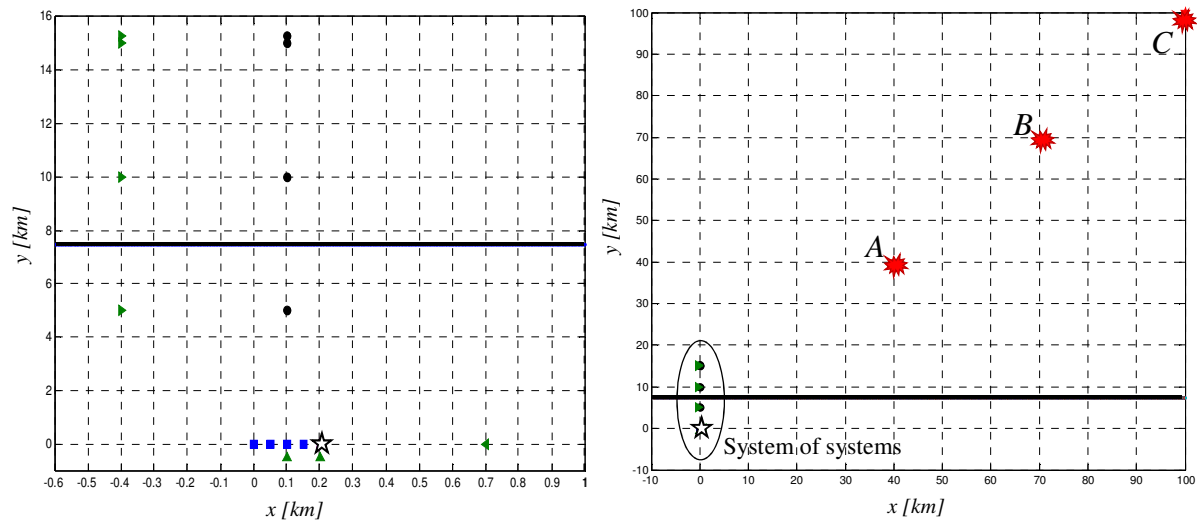


Figure 4: Left: spatial localization of the nuclear reactor (star), which identifies the nuclear power plant, with respect to the components of the electric power system (circle, from top to bottom: Generation Station, Substation, Pole 1, Pole 2), water system (square, from left to right: River, Pipe 1, Pump 1, Pipe 2) and road transportation (triangle, from top to bottom and from left to right: R7, R6, R5, R4, R3, R2, R1). Right: spatial localization of the system of systems with respect to three earthquake's epicenters A(40, 40), B(70, 70), C(100, 100). The horizontal bold line in both Figures represents the division between soft soil (above the line) and rock (below the line).

In Figure 5, the system-of-systems representation is given by the Muir Web showing the physical components of the infrastructure systems and factors which they depend on. In this representation the connections among the elements are depicted without expliciting the types of dependencies introduced in Section 3.2 which are illustrated in Figure 6.

The external water distribution system (Figure 5, left) is formed by a source of water (e.g., a river), a pump and pipes that carry the water. The failure probability of these elements when subjected to earthquake shocks depends on the type of soil, the design and materials of construction, and the maintenance. Operators are in charge of the maintenance of the structural elements and mechanical components.

The external power distribution system (Figure 5, center) is composed by the following elements: a generation station that produces the electrical energy, a substation that transforms the voltage from high to low, power lines and poles to support them, the type of soil on which the infrastructures rest, and the operators that run the generation station and provide the maintenance for all its elements and components.

The components of the emergency water and power distribution systems inside the plant are shown in Figure 5 on the right. The first system is composed by the same elements of the correspondent external system except for the source of water that is an artificial reservoir (e.g. a tank or pool), whereas the power system includes only the emergency diesel generators.

The elements considered for the transportation system are the roads (Figure 5, top). The state of this system is important for access of the materials and operators needed to keep or restore the functionality of the components required for the safe state of the critical plant.

Actually, in view of the methodological character of this work, for the sake of simplicity, the influence of the design construction and materials, the supply of fuel and materials for plant

operation, and the maintenance tasks are not included in the analysis. Failure of the power lines, being aerial elements and therefore being not directly affected by an earthquake, is also not considered. Finally, the assumption is made that the river is not perturbed by the earthquake, so that it is a source of water always available.

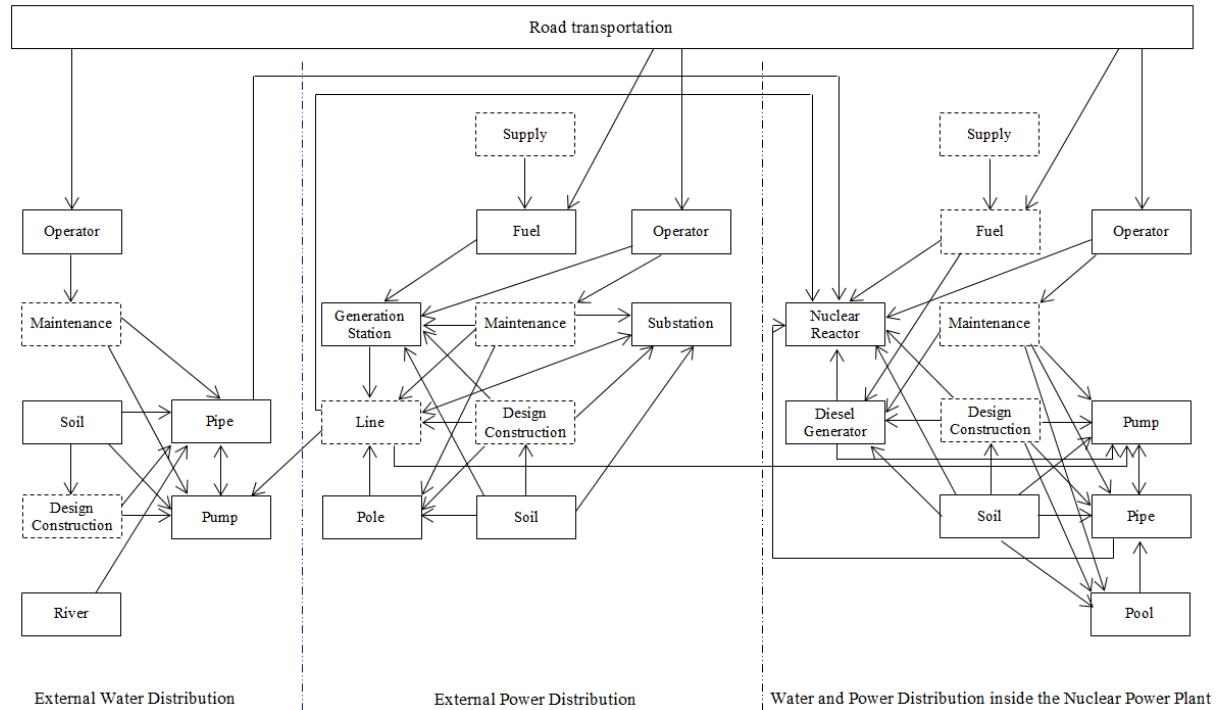


Figure 5: Muir Web of the system of systems: the elements in the dashed box are not considered in the present study.

In Figure 5, the inter-system dependencies are modeled as links connecting components of the five systems, S_i , $i = 1, \dots, 5$; these connections are of the same nature as those linking components of the individual systems (intra-systems dependencies). An example of these connections concerns the water system that depends on the power system as the pump needs electrical energy to work. This component receives the electrical energy from the external power distribution network; on the contrary, it is assumed that the pump inside the nuclear power plant can obtain energy from both the external and internal power systems.

The road transport network allows access to the components of the power and water systems for transporting material (e.g., fuel) and/or operators for operation and/or maintenance.

A representation of the Muir Web illustrating the different types of dependencies introduced in Section 3.2 is shown in Figure 6. The nodes are the components (e.g., generation station, pole, pump...) and the links are the dependencies among them. Note that, differently from Figure 5, the pole Po2 of the external power system is directly connected to the pumps of the external and internal water systems and to the nuclear reactor because the power lines are not considered in this work and the closest element to carry the power to the pumps and to the nuclear reactor is that pole. The transport system, S_5 , is composed by seven road access points to the components of the power and water systems. One access is provided for the

components outside the nuclear power plant, whereas two accesses are provided for the elements inside. In particular, the components of the external power system are considered to have a different road access because they are far from each other (the minimum distance is 300 m between the generation station and the substation, *Figure 4* left), the components of the external water system have the same road access, R3, because they are located close to each other (the total distance from the river to the nuclear power plant is 200 m, *Figure 4* left) and the components of the power and water systems inside the nuclear power plant have the same two road accesses, R1 and R2, since they are contained in the same building. Note that, in the present study, the road plays the role of “reserve component”, since we assume that elements that fail can be immediately repaired/replaced if the access to it through the road system does not fail (recovery times are not considered).

Analyzing the dependency links, it can be noticed that:

- “Direct dependencies” (solid lines) exist for the components of systems S_2 , S_3 , S_4 and between the components Po2 – Pu1, Po2 – Pu2, DG – Pu2 and R7 – GS. These links describe the fact that if a component fails, the connected component fails too because it cannot fulfill anymore its function. For example, in the system S_4 , if the pump Pu1 fails, it cannot pump the water and the pipe Pi2 cannot carry it.
- “Support dependencies” (dashed lines) exist for the road accesses R1, R2, R3, R4, R5, R6 and their corresponding components, since they are useful for transporting operators when maintenance or repair of a component is needed. Therefore, they are a support for those components.
- “Connections to the nuclear reactor” (bold lines) link the components of the systems S_1 , S_2 , S_3 , S_4 to the nuclear reactor.

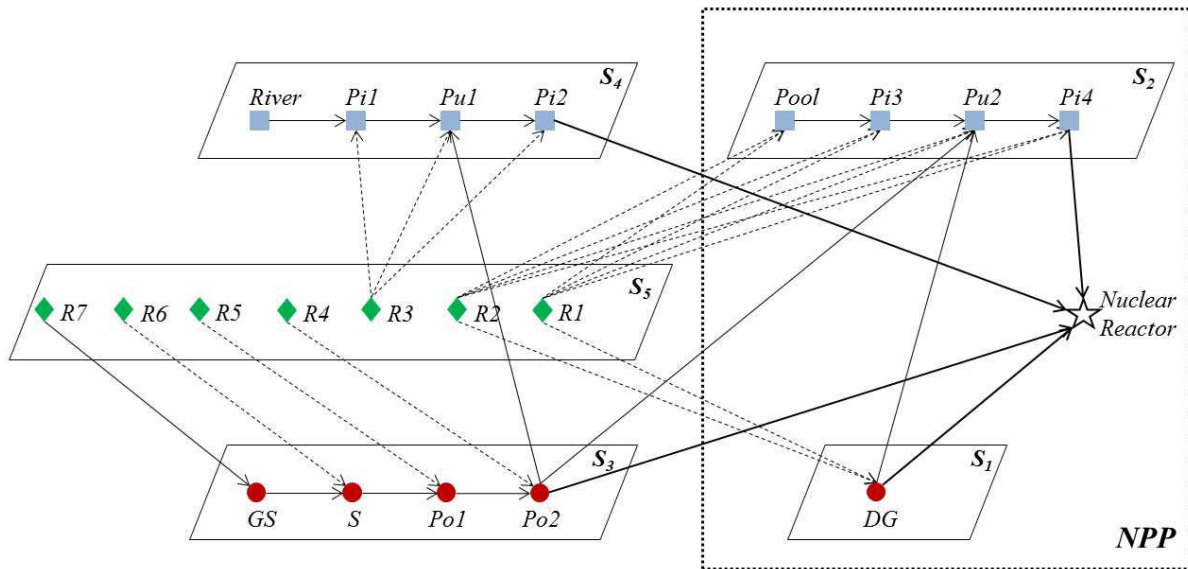


Figure 6: Representation of the physical components of the Muir Web of Figure 5, highlighting the different types of dependencies. The interconnected systems S_i , $i=1, \dots, 5$, can provide services relevant to the safe state of the nuclear power plant (NPP). The links represent the direct dependencies (solid lines), the support dependencies (dashed lines) and the dependencies of the nuclear reactor (star) on its interconnected systems (bold lines). GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R:

Road access, S_1 : internal power system, S_2 : internal water system, S_3 : external power system, S_4 : external water system, S_5 : Road transportation.

From the Muir Web and the knowledge of the functionality of the components of the system of systems, it is possible to identify the logic relations among them. For example, in the system S_3 , there is a flow of energy that starts from the generation station where it is produced, passes into the substation where it is converted into a low voltage and reaches the final destinations, i.e., the nuclear reactor and the water systems S_2 and S_4 , through the poles 1 and 2. All these components are connected in series (*Figure 7*) because if one of them fails, the entire system S_3 fails, i.e., it cannot fulfill anymore its function of providing energy. With the same reasoning, it can be evidenced that the components within the systems S_2 and S_4 , are connected in series too. Instead, the components of the system S_5 are independent and the system S_1 contains only one element.

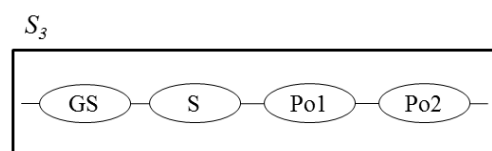


Figure 7: Logic connections between the components of the external power system, S_3 . GS: Generation Station, S: Substation, Po: Pole

Given the assumption of instantaneous recovery of a component if the road access to it is available, we can consider the road access like a “reserve component” in parallel with the corresponding element to which it provide access. In *Figure 8*, the logic connections between the systems $S_3 - S_5$ are provided. Note that the road $R7$ plays a double role in the external energy subsystem: it provides the generation station with access for 1) maintenance and repair (as the other road accesses $R1, R2, R3, R4, R5, R6$) and 2) operators and materials necessary to its operation. Therefore, the damage to this road access can cause the stop of the generation station and, as a consequence, the failure of the external power system S_3 . For this reason, the road $R7$ is in series with the system S_3 and in parallel with the generation station GS.

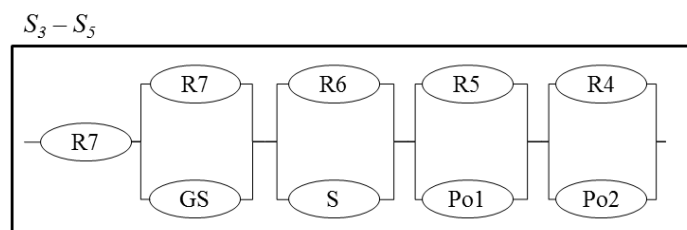


Figure 8: Logic connections between the components of the external power system, S_3 , and those of the road transportation, S_5 . GS: Generation Station, S: Substation, Po: Pole, R: Road access.

Figure 9 reports the logic relations between the systems $S_4 - S_5$ (on the left), $S_2 - S_5$ (in the middle) and $S_1 - S_5$ (on the right).

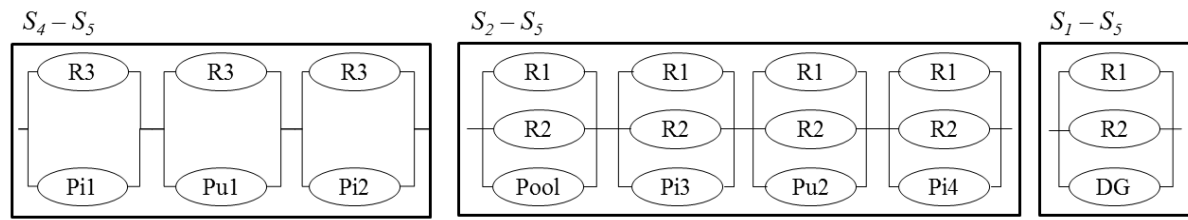


Figure 9: Logic connections between the components of the external water system, S_4 , internal water system, S_2 , internal power system, S_1 , and those of the road transportation, S_5 . GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access.

The power and water systems are connected in series as the pump in the water system needs energy to work. In particular, the external power system, S_3 , is in series with the internal and external water systems, S_2 and S_4 , respectively, and the internal power system, S_1 , is in series with the internal water system, S_2 , as shown in Figure 10.

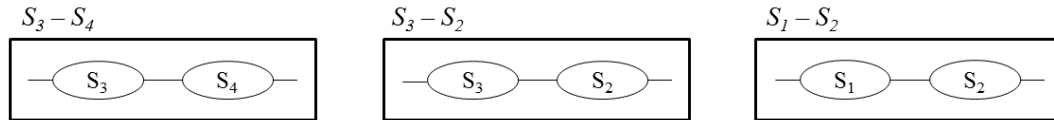


Figure 10: Logic connections between the power and water systems. S_1 : internal power system, S_2 : internal water system, S_3 : external power system, S_4 : external water system.

Figure 11 integrates the logic relations among the systems S_i , $i = 1, \dots, 5$, to maintain the safe state of the nuclear power plant.

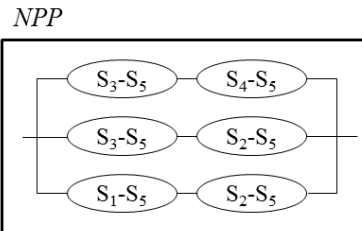


Figure 11: Logic connections between the interconnected systems S_i , $i=1, \dots, 5$, to the nuclear power plant (NPP). S_1 : internal power system, S_2 : internal water system, S_3 : external power system, S_4 : external water system, S_5 : road transportation.

4.2. Results and limitations

Figure 12 shows the results of the Monte Carlo simulation for Seismic Probabilistic Risk Assessment carried out with the operative procedure illustrated in Section 3.3 applied to the case study described above, regarding the system of systems represented by the Muir Web of Section 3.2. For each magnitude level sampled from a truncated exponential probability distribution (1) with lower threshold $m_{min} = 5$ and upper bound $m_{max} = 7$, the estimate of the probability of the nuclear power plant (NPP) to reach an unsafe condition, is computed. The number of magnitude values sampled is 1000 and the number of simulations (N_T) of the components configuration for each value of magnitude is 5000.

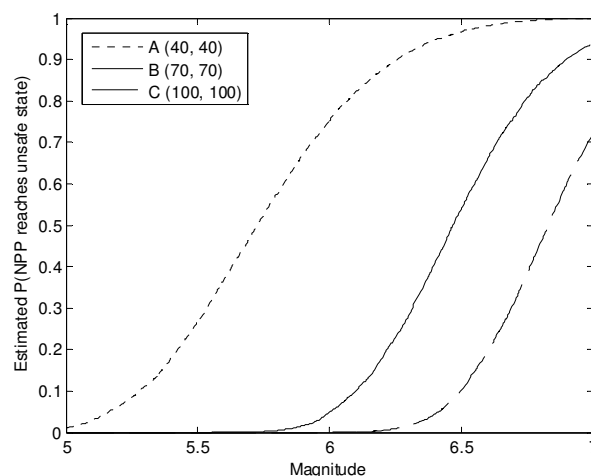


Figure 12: Estimate of the probability that the nuclear power plant reaches an unsafe state after an earthquake of a given magnitude on the basis of different source-to-site distances. With reference to the map of Figure 4, the coordinates of the earthquake's epicenters considered are A(40, 40), B(70, 70), C(100, 100), expressed in kilometers.

The analysis is carried out for the three earthquake's epicenters, A, B, C, shown in Figure 4. As expected, the higher the distance, the lower is the probability that the safety of the nuclear power plant cannot be assured.

Figure 13 shows the comparison between the probabilities that the nuclear power plant turns into an unsafe condition after the occurrence of an earthquake, considering it both as an isolated plant provided with its emergency devices (case of independence) and as embedded in the system of systems of the supporting infrastructures (case of dependence).

It can be seen that, with the given assumptions and data:

- the probabilities to reach an unsafe state computed in case of dependence are slightly lower than those computed in case of independence. This result shows that in principle the infrastructures in the surrounding of the critical plant can contribute to its resilience, offering additional possibilities for maintaining (or restoring) a safe condition;
- the larger difference between the probabilities computed in the case of dependence and independence results for low magnitude values when the source-to-site distance is small, e.g., for magnitudes lower than 5.8 in the case A(40, 40) (Figure 13, top), and for high magnitudes when the source-to-site distance increases, e.g., for magnitudes between 5.8 and 6.5 in the case B(70, 70), and for magnitudes higher than 6.2 in the case C(100, 100) (Figure 13 left and right, respectively). This is expected and can be explained as follows. In case of small source-to-site distance, the higher the magnitude the higher are the impacts on all the systems, so it is probable that also the interdependent systems are damaged and they cannot be used as an additional support to maintain (or restore) the nuclear power plant in a safe state; when, instead, the

magnitude is small not all the components are affected by the earthquake and the interdependent systems could be useful as an additional support for the safety of the critical plant. In the case of high source-to-site distance, the larger difference between the probabilities computed in the case of dependence and independence is for high magnitudes because the lower values do not have impacts on the system components.

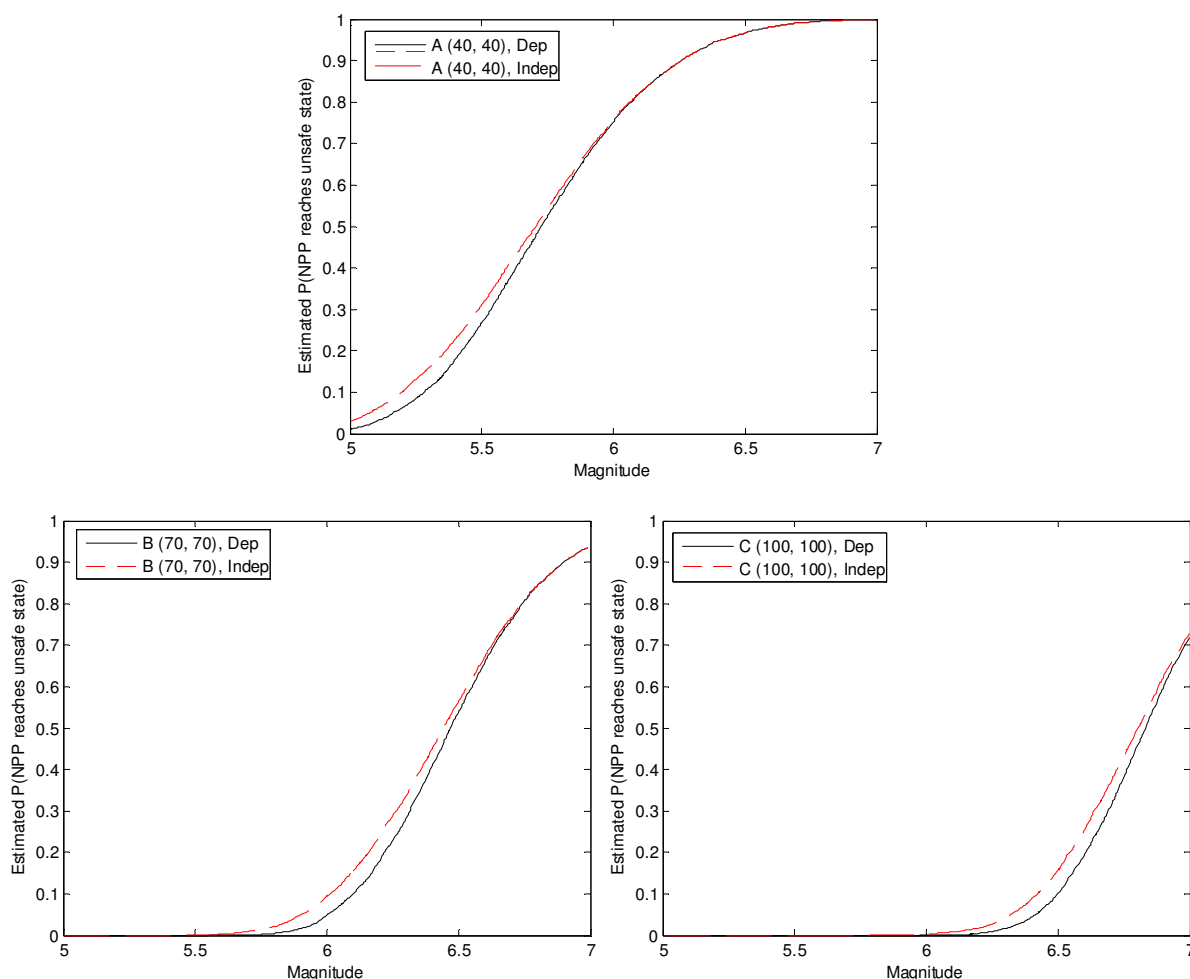


Figure 13: Comparison between the results of the MC simulation in the case of dependence (solid line) of the nuclear power plant on the connected infrastructure systems, and in the case of independence (dashed line). The analysis is carried out for three earthquake's epicenters: A(40, 40) on the top, B(70, 70) on the left and C(100, 100) on the right.

Figure 14, on the right, shows the comparison between the probabilities that the nuclear power plant turns into an unsafe state after the occurrence of an earthquake whose epicenter is in B(70, 70) considering the case of dependence (solid line) presented in Figure 13 (left) and considering each individual infrastructure system as isolated (dashed line), as depicted in the Muir Web of Figure 14, on the left, where all the inter-system links have been removed. This analysis allows highlighting to what extent interdependencies among the infrastructure systems affect the safety of the plant inserting “extra” vulnerabilities to the system of systems.

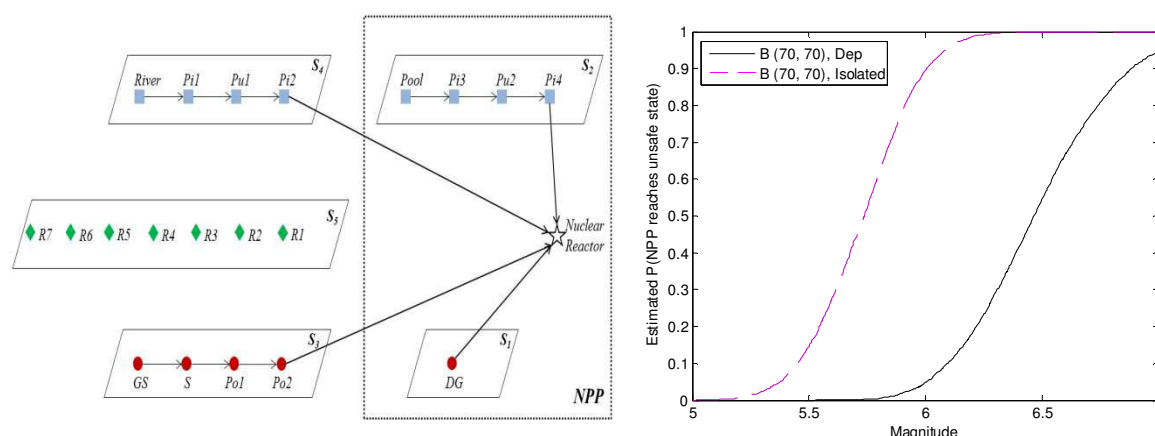


Figure 14: Left: Muir Web of the system of systems of Figure 6 without the inter-system links. Solid lines: direct dependencies, bold lines: connection to the nuclear reactor. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, S₁: internal power system, S₂: internal water system, S₃: external power system, S₄: external water system, S₅: Road transportation. Right: Comparison between the results of the MC simulation in the case of dependence (solid line) of the nuclear power plant on the connected infrastructure systems and in the case of isolated power, water and road transportation infrastructure systems (dashed line) as shown in the scheme on the left. The analysis is carried out for the earthquake's epicenter B(70, 70).

The result shows an increase in the probability of unsafe state of the nuclear power plant: this is due to the assumption of instantaneous recovery of the components (Section 4.1) that implies that if a road access is available, the corresponding component is immediately considered operational. Therefore, under this limiting assumption, the connections between the road system and the other systems increase the safety of the nuclear power plant. On the contrary in the realistic situation, this would not be the case. To show this, we have performed another simulation, considering the road accesses as a part of the systems to which they provide access as shown in Figure 15, left. The results obtained are reported in Figure 15, right. The solid line represents the case of dependence as in Figure 13 (left) and 14 (right) and the dashed line the case of isolated water and power systems, considering the road accesses as part of these systems, as explained above. It can be seen that in the first case the probability values are higher than in the second one, the reduction of safety of the nuclear power plant being due to the interdependences between the water and power systems.

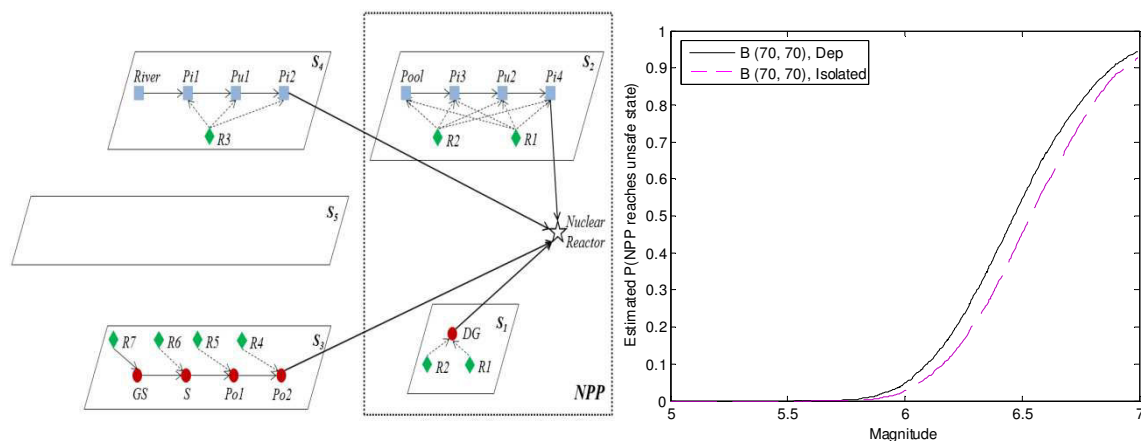


Figure 15: Left: Muir Web of the system of systems of Figure 6 including the road accesses in the corresponding systems to which they provide the access and removing the dependencies between the power and water systems. Solid lines: direct dependencies, dashed lines: support dependencies, bold lines: connection to the nuclear reactor. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, S_1 : internal power system, S_2 : internal water system, S_3 : external power system, S_4 : external water system, S_5 : Road transportation system. Right: Comparison between the results of the MC simulation in the case of dependence (solid line) of the nuclear power plant on the connected infrastructure systems and in the case of isolated power and water systems (dashed line) as shown in the scheme on the left. The analysis is carried out for the earthquake's epicenter B(70, 70).

Some limitations of the case study are pointed out in the following. For example, one concerns the assumption of immediate recovery of an element when it fails. Obviously, in practice, it takes time to bring back a component. In this sense, a time recovery distribution should be considered to perform a more realistic analysis. Similarly, in specific cases the duration of emergency service supply should be included in the analysis to provide a proper time-dependent picture of the conditions of the critical plant. Also, some potentially influential factors identified in the Muir Web representation have been neglected in order to simplify the quantitative analysis, like the design construction and materials, the maintenance task, etc. Furthermore, as mentioned earlier, a multi-state model should be considered to describe partial failures at the components levels and partial safety at the critical plant level. Finally, as in all risk analyses, uncertainties are present and need to be taken into account. In the specific case study, uncertainty is present in the inputs to the Probabilistic Seismic Hazard Analysis that are based on geological, seismological and geophysical data subjected to expert interpretations, but also in the parameters of the component fragility models. These and other uncertainties may have a considerable role in the result, and the decisions associated to it.

In order to develop the case study into a more realistic one, it is necessary to relax some of the assumptions introduced but this will lead to increased analysis and computational costs due to 1) the collection of data needed to evaluate additional factors neglected in the present work and 2) the calculation for evaluating the quantities of interest from a multi-state model with associated uncertainty.

5. CONCLUSIONS

We have presented a system-of-systems framework of analysis of the risk of a critical plant from external events, to account for the influence of the interdependent infrastructures in which the plant is embedded.

For illustrating the conceptual framework of the analysis, we have made reference to an earthquake as the external event, a nuclear power plant as the critical plant and the power and water distribution, and transportation networks as the interdependent infrastructure systems. We admittedly simplified many technical details of the systems considered and made opportunistic assumptions for the purpose of preserving the focus on the conceptual, methodological framework of analysis.

We provided a numerical example by resorting to the Muir Web as system analysis tool to build the system-of-systems model and Monte Carlo simulation for the quantitative evaluation of the model.

In particular, the following analyses have been carried out:

- a. a comparison between the probabilities that the nuclear power plant reaches an unsafe state after an earthquake of a given magnitude, depending on different site-to-source distances: as expected, the higher the distance, the lower is the probability to get to an unsafe state;
- b. a comparison of the previous probabilities (a.), obtained in the case of dependence of the nuclear power plant on the interconnected infrastructure systems, with those obtained in the case of independence, i.e., considering the nuclear power plant as an isolated system provided only by its internal emergency devices: the results show that the probability to reach an unsafe state is higher in this latter case and, in particular, the “resilience” contribution of the interdependent systems to the safety of the nuclear power plant is significant for low magnitudes when the source-to-site distance is small, and for high magnitudes when the source-to-site distance is big;
- c. a comparison of the previous probability (a.) for one earthquake epicenter, obtained in the case of dependence of the nuclear power plant on the interconnected infrastructure systems, with that obtained in the case of isolated infrastructure systems, i.e., removing all the inter-system links and considering all the infrastructure systems as isolated: the results show that the probability to reach an unsafe state is higher in this latter case, due to the particular “redundancy” role of the road accesses under the assumption of immediate recovery of the components;
- d. the same comparison as in c., but considering, for the isolated case, the dependence between the road accesses and the corresponding components and maintaining the independence among the other systems: the results show that in this case the probability to reach an unsafe state is lower; this means that the inter-system links among the power and water systems increase the probability of failure of the system of systems and, thus, of the nuclear power plant being in an unsafe state.

The results of the analyses, albeit performed on a simplified case study and under limiting assumptions, highlight that the interdependent infrastructure systems may play a role for the safety of a critical plant, and it thus seems advisable to include them in the analysis framework. In fact, they can provide additional support to the safety of the critical plant providing inputs necessary for its safe operation (results of case b. above), but their contribution can be reduced by their interconnections as shown in the case d. above. This is relevant as it can lead to considerations with respect to the decision making related to safety-critical issues. One may even imagine considering the optimization of some controllable characteristics of the system of systems with the objective of increasing the safety of the critical plant. This could be done by a thorough analysis to identify the most important elements in the system of systems and a cost/benefit analysis to rationally direct the investments of efforts and resources for improving their structural/functional responses, within a comprehensive system-of-systems approach.

Note that although the driving case study for the illustration of the framework has considered a nuclear power plant as the critical plant, others can be analyzed with their specificities, e.g., chemical process and oil & gas plants or refineries which can release toxic material, develop fires and explosions. For example, loss of offsite power occurred during operation of a vinyl chloride monomer plant at Sodegaura, Chiba (Japan) after a strong earthquake in 1987. In that occasion, the emergency power generator started, as expected, but then it was stopped. As a consequence of the total power failure, the alkali circulation pump of the absorber stopped and the hydrochloric acid gas was released leading to environmental pollution [28].

Future research work will be devoted to apply the framework of analysis presented to diverse systems of systems, with different specificities, and to improve it, for example, by introducing the time needed to recover the safe state of the critical plant and considering a multi-state model for the components of the system of systems. The new case studies will also allow evaluating further the Muir Web representation model and what it is capable to do that other techniques cannot do.

REFERENCES

- [1] International Atomic Energy Agency. Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants. Vienna (AT): IAEA Safety Standards; 2010. Specific Safety Guide No. SSG-3. Chapter 8, Specifics of level 1 PSA for external hazards; p. 91-114.
- [2] International Atomic Energy Agency. The great east Japan earthquake expert mission – IAEA international fact finding expert mission of the Fukushima Dai-ichi NPP accident following the great east Japan earthquake and Tsunami. Mission Report 24 May – 2 Jun 2011. 162 p.
- [3] Adachi T, Ellingwood BR. Serviceability of earthquake-damaged water systems: Effects of electrical power availability and power backup systems on system vulnerability. Reliability Engineering & System Safety. 2008; 93(1):78-88.
- [4] Aven T. On Some Recent Definitions and Analysis Frameworks for Risk, Vulnerability, and Resilience. Risk Analysis. 2011; 31(4):515-522.
- [5] Eusgeld I, Nan C, Dietz S. “System-of-systems” approach for interdependent critical infrastructures. Reliability Engineering & System Safety. 2011; 96(6):679-686.

- [6] Haimes YY. On the Complex Definition of Risk: A Systems-Based Approach. *Risk Analysis*. 2009; 29(12):1647-1654.
- [7] Johansson J, Hassel H. An approach for modelling interdependent infrastructures in the context of vulnerability analysis. *Reliability Engineering & System Safety*. 2010; 95(12):1335-1344.
- [8] Kröger W, Zio E. Vulnerable systems. London: Springer; 2011. 63 p.
- [9] Wang S, Hong L, Chen X. Vulnerability analysis of interdependent infrastructure systems: a methodological framework. *Physica A: Statistical Mechanics and its Applications*. Forthcoming 2012.
- [10] Sanderson EW. *Mannahatta: A natural history of New York City*. New York: Abrams; 2009. 352 p.
- [11] Kalos MH, Whitlock PA. Monte Carlo methods. Vol. 1, Basics. New York: Wiley; 1986. 186 p.
- [12] Marseguerra M, Zio E. Basics of the Monte Carlo Method with Application to System Reliability. Hagen(DE): LiLoLe – Verlag GmbH; 2002. 141 p.
- [13] Zio E. Computational methods for reliability and risk analysis. Series on Quality, Reliability and Engineering Statistics, Vol 14. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2009. Chapter 2, Monte Carlo simulations for reliability and availability analysis; p. 59-69.
- [14] LaRocca S, Guikema SD, Cole J, Sanderson E. Broadening the discourse on infrastructure interdependence by modeling the “Ecology” of infrastructure systems. In: *Application of Statistics and Probability in Civil Engineering*. Faber, Kohler & Nishijima (eds). London: Taylor & Francis Group; 2011. p. 1905-1912.
- [15] Seismic Probabilistic Risk Assessment Implementation Guide, EPRI, Palo Alto, CA: 2003. TR-1002989.
- [16] Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Expert. Main Report, Vol. 1. 1997, NUREG/CR-6372 UCRL-ID- 122160. Supported by U.S. Nuclear Regulatory Commission (NRC), the U.S. Department of Energy (DOE); and the Electric Power Research Institute (EPRI).
- [17] Sen TK. Fundamentals of seismic loading and structures. Singapore: John Wiley & Sons, Ltd; 2009. Chapter 7, Probabilistic Seismic Hazard Analysis; p. 181-218.
- [18] Gutenberg B, Richter CF. Frequency of earthquakes magnitude in California, *Bulletin of the Seismological Society of America*. 1944; 34:185-188.
- [19] Kanamori H. The energy release in great earthquakes. *Journal of Geophysical Research*. 1977; 82(20): 2981–2987.
- [20] Kanamori H, 1983. Magnitude scale and quantification of earthquakes. In: SJ. Duda and K. Aki Editors. *Quantification of Earthquakes*. *Tectonophysics*, 93: 185-199.
- [21] Kramer SL, *Geotechnical Earthquake Engineering*, Prentice Hall, New Jersey. 1996.
- [22] Weatherill GA, Burton PW. The application of multiple random earthquake simulations to probabilistic seismic hazard assessment in the Aegean region. *Firs European Conference on Earthquake Engineering and Seismology*. Geneva, Switzerland. 2006.
- [23] Ambraseys NN, Douglas J, SARMA SK, Smit PM. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering*. 2005; 3:1-53.
- [24] Muir J. Notebook, July 27, 1869. In: *The John Muir Papers, 1858-1957*. Limbaugh and Lewis Editors. Chadwyck-Healey, Alexandria, VA, 1985.
- [25] Huang YN, Whittaker AS, Luco N. A probabilistic seismic risk assessment procedure for nuclear power plants: (I) Methodology, *Nuclear Engineering and Design*. 2011; 241: 3996– 4003.
- [26] Varpasuo P. Seismic fragility analysis of selected heavy components in LNNP unit1 reactor building. *Transactions of the 17th International Conference on Structural Mechanics in Reactor Technology (SMiRT)*. Prague, Czech Republic, August 17-22, 2003.
- [27] Basu PC. Seismic fragility of nuclear installations. Atomic Energy Regulatory Board. Mumbai, India. 2008. Presentation: <http://civil.iisc.ernet.in/basu.pdf>
- [28] Itagaki H, Tamura M. Case details: Partial leakage of hydrochloric acid gas from an absorber due to an earthquake. Hatamura Institute for the Advancement of Technology. Failure Knowledge Database: <http://www.sozogaku.com/fkd/en/cfen/CC1000197.html>

Paper III

Assessing nuclear power plant safety and recovery from earthquakes using a system-of-systems approach

E. Ferrario and E. Zio

Reliability Engineering & System Safety 125 (2014) 103-116

Assessing nuclear power plant safety and recovery from earthquakes using a system-of-systems approach

E. Ferrario^a and E. Zio^{a,b}

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supélec, France

enrico.zio@ecp.fr, enrico.zio@supelec.fr

^bDepartment of Energy, Politecnico di Milano, Italy

enrico.zio@polimi.it

Abstract

We adopt a ‘system-of-systems’ framework of analysis, previously presented by the authors, to include the interdependent infrastructures which support a critical plant in the study of its safety with respect to the occurrence of an earthquake. We extend the framework to consider the recovery of the system of systems in which the plant is embedded. As a test system, we consider the impacts produced on a nuclear power plant (the critical plant) embedded in the connected power and water distribution, and transportation networks which support its operation. The Seismic Probabilistic Risk Assessment of such system of systems is carried out by Hierarchical modeling and Monte Carlo simulation. First, we perform a top-down analysis through a hierarchical model to identify the elements that at each level have most influence in restoring safety, adopting the criticality importance measure as a quantitative indicator. Then, we evaluate by Monte Carlo simulation the probability that the nuclear power plant enters in an unsafe state and the time needed to recover its safety. The results obtained allow the identification of those elements most critical for the safety and recovery of the nuclear power plant; this is relevant for determining improvements of their structural/functional responses and supporting the decision-making process on safety critical-issues. On the test system considered, under the given assumptions, the components of the external and internal water systems (i.e., pumps and pool) turn out to be the most critical for the safety and recovery of the plant.

Keywords: System of systems, Recovery, Seismic Probabilistic Risk Assessment, Hierarchical representation, Monte Carlo simulation.

1. INTRODUCTION

We consider a safety-critical plant, e.g., a nuclear power plant (NPP), exposed to an external hazard, e.g., an earthquake. Internal emergency devices have been designed to provide safety for the plant upon occurrence of the hazardous event, i.e., even if the infrastructure services are not available. However, the history of industrial accidents, including the recent Fukushima nuclear disaster [1], has shown us that the safety of a plant depends also on the infrastructures in which it is embedded, which may or may not provide “resilience” properties. Then, the analysis for the evaluation of the probability that a critical plant remains or not in a safe state, i.e., in a condition that does not cause health and/or environmental damages, upon occurrence of an external accident event, must extend to the interdependent infrastructures connected to it, adopting a “system-of-systems” point of view [2], [3], [4], [5], [6], [7], [8]. For this, we adopt the framework of analysis proposed by the authors in [9] and extend it to include the capacity of the system of recovering from an external aggression or shock, using as representative quantity the recovery time, i.e., the period necessary to restore a desired level of functionality of a system after the shock [10].

As a test system for the developments of our considerations and analyses, we consider the impacts of an earthquake on a nuclear power plant, extending the system boundaries to the power and water distribution, and the transportation networks (the interdependent infrastructure systems) that can provide services necessary for keeping or restoring its safety. The test system is fictitious and highly simplified, intended only to illustrate the way of analyzing the problem under a “system-of-systems” viewpoint, accounting for the effects of the interdependencies.

The systemic analysis is performed in two main steps. In the first step, a conceptual map previously built by the authors [9] to understand all the dependencies and interdependencies between the components of the infrastructure systems connected to the nuclear power plant is exploited to construct a hierarchical representation of the system of systems. Hierarchical modeling is here used for a top-down analysis of the elements that at each level have most influence in restoring safety. Indeed, the hierarchical representation facilitates the identification of the structure of the system of systems, allowing the determination of the critical elements [11]. As a quantitative indicator of the contribution of the components to the recovery of safety, the criticality importance measure is used [12], [13].

In the second step, Monte Carlo simulation [14], [15], [16] is applied to compute 1) the probability that the nuclear power plant enters in an unsafe state and 2) the time of recovery of the safety of the nuclear power plant, accounting for the contributions of both the internal emergency devices and the connected infrastructures.

The reminder of the paper is organized as follows. In Section 2, the basic concepts of a Seismic Probabilistic Risk Assessment are introduced; in Section 3, the hierarchical modeling of a system of systems and Monte Carlo simulation framework for Seismic Probabilistic Risk Assessment are described; in Section 4, the test system and the results of the analysis are presented; in Section 5, conclusions are provided.

2. METHOD FOR SEISMIC PROBABILISTIC RISK ASSESSMENT

To estimate the probabilities of occurrence of different levels of earthquake ground motion that may affect an infrastructure and its response to such event, a Seismic Probabilistic Risk Assessment (SPRA) is typically applied. In a very short and schematic synthesis, it is based on three parts [17], [18]:

- Seismic Hazard Analysis: computes the probabilities of occurrence of different levels of earthquake ground motion at a site of interest.
- Seismic Fragility Evaluation: identifies the seismic capacity of a component in terms of its conditional probability of failure for any given ground motion level.
- System Analysis: integrates the outputs of the hazard and fragility analyses to evaluate the impact of an external event to the infrastructure of interest.

The first part is traditionally developed as a Probabilistic Seismic Hazard Analysis (PSHA) consisting of four procedural steps [17], [18], [19]:

- 1) Earthquake source zones identification and characterization
- 2) Earthquake recurrence relationship definition
- 3) Ground motion attenuation relationship formulation
- 4) Exceedance probability calculation

The first step concerns the identification and characterization of the seismic sources in the proximity of the site of interest. It involves geological, seismological, geophysical data and scientific interpretations; as a consequence it is a critical part of the analysis and it is associated with considerable uncertainty [17], [18]. The major outputs of this step are the seismic map that defines the seismic zones (areas where the earthquake sources have common characteristics like geometry, earthquake activity, earthquake annual recurrence rate), the probability distribution of the source-to-site distance and the identification of the maximum earthquake magnitude, i.e., the largest magnitude that a source can generate [17], [18].

In the second step, the seismic earthquake recurrence relationship, i.e., the annual frequency of occurrence of a given magnitude event for each source, is defined. Typically, it is described by the Gutenberg-Richter law, $\log(n) = a - bm$ where n is the number of earthquakes with magnitude¹ greater than m and a and b are parameters obtained by regression data analysis [17], [18]. This relation implies that the magnitude is exponentially distributed [22], [23]:

$$F_M(m) = 1 - e^{-\beta m} \quad (1)$$

where $\beta = \log_{10} b \cong 2,303b$ represents the relative frequency of smaller to larger events. Equation 1, however, is an unbounded probability distribution so that the magnitude can assume very high values, which are unrealistic and very low values, which are negligible. Therefore, the distribution is double-truncated by upper and lower bounds, m_{max} and m_{min} , respectively, and it is reformulated as follows [17]:

¹ The magnitude scale typically used is the moment magnitude defined by [20]. For medium size earthquakes it is similar to the Richter values [21].

$$F_M(m) = \frac{1 - e^{-\beta(m - m_{min})}}{1 - e^{-\beta(m_{max} - m_{min})}} \quad (2)$$

The third step identifies the ground motion value at the site of interest, given the source-to-site distance and the magnitude. The higher the distance from the source, the lower is the ground motion value. Typical ground motion parameters are the peak ground acceleration and the spectral acceleration. Many ground motion equations have been defined on the basis of the earthquake and site characteristics [24]. They usually assume this expression [17]:

$$\log z' = C_1 + C_2 m + C_3 m C_4 + C_5 \log[r + C_6 \exp(C_7 m)] + C_8 r + g(source) + g(site) \quad (3)$$

where z' is the mean ground motion parameter, $C_i, i=1, \dots, 8$, are the regression coefficients, r is the source-to-site distance, m is the magnitude and $g(source)$ and $g(site)$ are terms that reflect the characteristics of the source and site, respectively.

For example, the peak ground acceleration is well described by [25]:

$$\log_{10} z' = C_1 + C_2 m + (C_3 + C_4 m) * \log_{10} \sqrt{r^2 + C_5^2} + C_6 S_s + C_7 S_A + C_8 F_N + C_9 F_T + C_{10} F_O \quad (4)$$

where S_s and S_A represent the types of soil (soft, stiff or rock, when both variables are set to zero) and F_N, F_T and F_O describe the faulting mechanism (normal, thrust or odd).

In the fourth step, the probability of exceedance of ground motion in any time interval is computed by an analytical integration for each magnitude, distance and ground motion value by the following equation [17]:

$$\nu(z) = \sum_{i=1}^S \lambda_i(m_{min}) \int_{m_{min}}^{m_{max}} \int_{r_{min}}^{r_{max}} f_{R_i}(r|m) f_{M_i}(m) P(Z > z|m, r) dm dr \quad (5)$$

where $i = 1, \dots, S$ represents the source zone, $f_{R_i}(r|m)$ and $f_{M_i}(m)$ are the probability density functions of the source to site distance and of the magnitude, respectively, $P(Z > z|m, r)$ is the probability of exceedance of the ground motion for each source zone, $m_{min}, m_{max}, r_{min}, r_{max}$ are the lower and upper bounds of the magnitude and distance considered and $\lambda_i(m_{min})$ is a rate that removes the contribution of earthquakes with magnitude lower than m_{min} that is not significant.

In the second part of the SPRA, a fragility evaluation is carried out to provide the parameter values (i.e., the median acceleration capacity A_m and the logarithmic standard deviation due to randomness and to uncertainty in the median capacity β_r and β_u , respectively) of the component fragility model of the kind [17]:

$$f' = \Phi \left[\frac{\log\left(\frac{z'}{A_m}\right) + \beta_u \Phi^{-1}(Q)}{\beta_r} \right] \quad (6)$$

where f' is the conditional probability of failure for any given ground motion level z' and Q is the subjective probability of not exceeding a fragility f' .

In the third part, an evaluation of the consequences of the seismic event to the infrastructure under analysis is traditionally performed by the development of event trees and logic models for each event tree top event [17]. In this work we adopt a hierarchical representation and a Monte Carlo simulation for this evaluation.

3. METHOD FOR SAFETY ASSESSMENT AND RECOVERY ANALYSIS

In this Section, the hierarchical representation of a system of systems (Section 3.1) and the operative steps of the Monte Carlo (MC) simulation method for its Seismic Probabilistic Risk Assessment (SPRA) (Section 3.2) are summarized.

3.1. Hierarchical representation of a system of systems

Let us denote a system i at the level L of the hierarchy as $S_i^{(L)}$ and by $N_S^{(L)}$ the number of systems at the level L . In the hierarchical representation of a system-of-systems view of a critical plant, H , at the top of the hierarchy there is only $N_S^{(1)} = 1$ system, the critical plant itself, and it is denoted as $S_1^{(1)}$. At the second level, $L = 2$, this is connected to $N_S^{(2)}$ systems, $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, inside and outside the plant, that provide it with the necessary inputs for its operation. The systems $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, at level $L = 2$, can, in turn, be broken down into subsystems $S_i^{(3)}$, $i = 1, \dots, N_S^{(3)}$ at the third level of the hierarchy, $L = 3$. The hierarchical modeling is built by identifying the elements (or groups) that are “part of” the parent objects, and continuing up to the desired level $L = N_L$, where N_L is the number of levels of the hierarchy. For the analysis of interest here, the hierarchy is continued down to the level of details of the individual components of the system of systems. However, following this procedure for building the hierarchical model, some components may not be considered. Actually, some elements of the system of systems i) may not provide the critical plant H with the inputs necessary for its operation, thus, they cannot be represented in the level-2 of the hierarchy, and ii) may not be part of any system $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, thus, they cannot be identified by the decomposition criteria. These components (hereafter called “recovery supporting elements”) provide the components (or groups) of the systems $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, with the inputs necessary for their functioning or recovery and are here represented as a part of the systems (groups) they support.

By way of example, refer to Figure 1 in which the graph of the system (top), the grouping of its components (middle) and its hierarchical representation (bottom) are depicted. The intra-system dependencies (solid lines), the inter-system ones (dashed lines) and the connections to the critical plant H (bold lines) are identified (Figure 1, top). The increasing resolution in the four levels considered is illustrated (Figure 1, middle): in the first level (square shape), the critical plant H is represented; in the second level (dashed oval shape), the three interdependent systems, $S_i^{(2)}$, $i = 1, \dots, 3$ are reported; in the third and fourth levels (dotted and solid oval shapes, respectively), the grouping of the elements within the systems of level 2 are specified. In Figure 1, top, the recovery supporting elements are those not connected to the critical plant H but linked to other components by dashed lines (i.e., $S_1^{(4)}$ and $S_2^{(4)}$); in Figure 1, middle, they are grouped in the systems to which they provide support, e.g., $S_1^{(4)}$ is both in the systems $S_1^{(2)}$ and $S_2^{(2)}$ and $S_2^{(4)}$ is in the system $S_3^{(2)}$; in Figure 1, bottom, they are

represented in the last levels of the hierarchy according to the grouping of the Figure 1 in the middle. Notice that the recovery supporting elements can belong to more systems (or groups) since they can be a support to different components (or groups), whereas all the others components (or groups) are within just one system since they are built following the criteria “to be a part of”. A final remark is in order with respect to the top-down approach adopted to build the hierarchical model. It is possible that, before reaching the bottom of the hierarchy, some components cannot be subdivided further (e.g., $S_2^{(3)}$ coincides with $S_1^{(4)}$) leading to an incomplete hierarchical representation. Therefore, in this circumstance, a copy of those elements is reported in the levels they are absent [26].

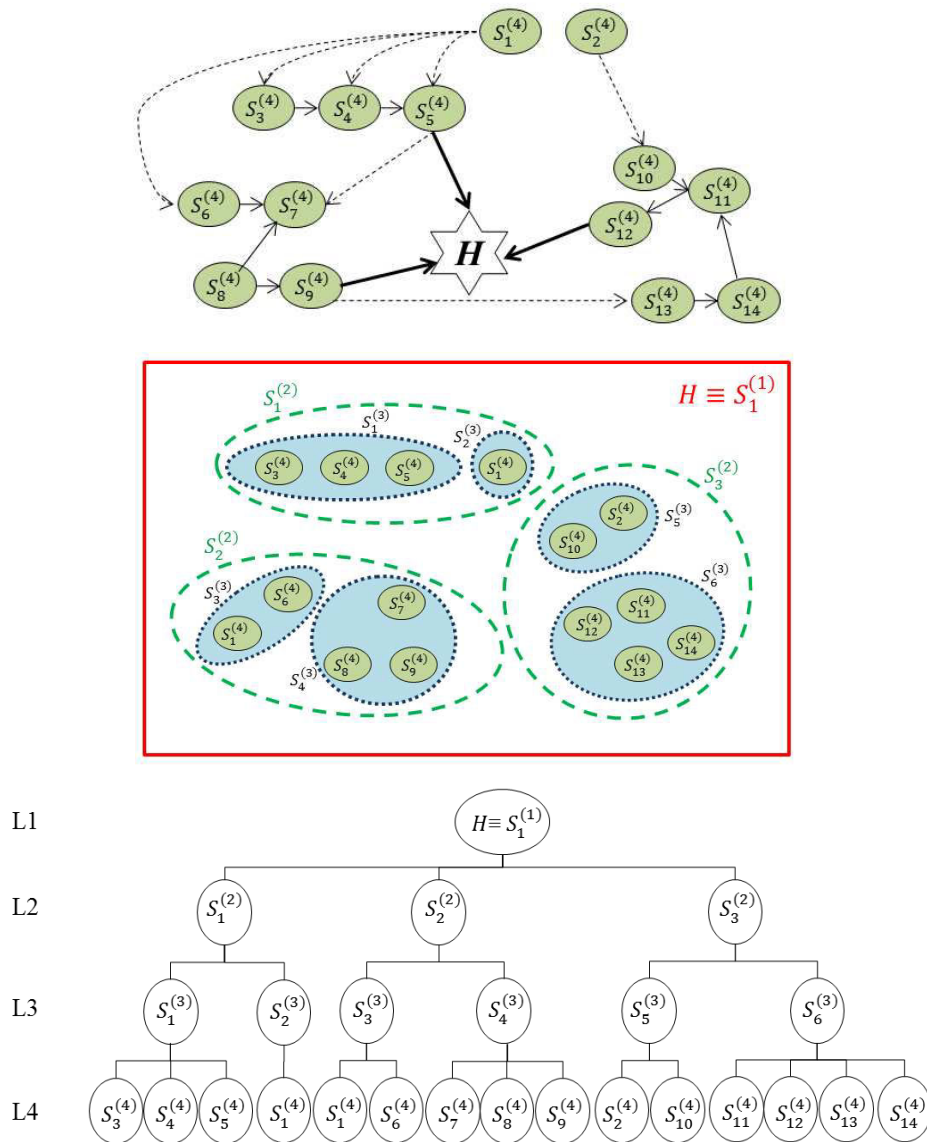


Figure 1: Top: dependencies among the components of the system of systems; the links represent the intra-systems dependencies (solid lines), the inter-systems dependencies (dashed lines) and the dependencies of the critical plant H on its interconnected systems (bold lines). Middle: graphical representation of their grouping; the rectangular, dashed, dotted and solid oval shapes represent the increasing resolution in the hierarchical level. Bottom: corresponding hierarchical representation; L : Level.

A system $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, at level $L - 1$, $L = 2, \dots, N_L$, can be in an operational or in a failure state depending on the states of the systems at the level L , on their functionality and on their logic connections. A state (truth) matrix is associated to each system $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, $L = 2, \dots, N_L$, where the first columns represent the states of the systems $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, at level L and the last column represent the state of the system $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, at level $L - 1$. The entries $\{a_{ij}\}$ are equal to 1 or 0 according to whether the states are in a failure state or not.

By way of example, refer to Table 1 and Figure 2 where three state matrices and the corresponding fault trees are reported, with reference to the system $S_5^{(3)}$ at level $L = 3$ of Figure 1 (middle) composed by the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ at level $L = 4$. The first two state matrices represent, respectively, the series and parallel configurations between the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ (illustrated by the OR and the AND gate in the fault trees): in the first case, the state of $S_5^{(3)}$ can assume only one operational state, since the failure of $S_{10}^{(4)}$ or $S_2^{(4)}$ causes its failure; whereas, in the second case, $S_5^{(3)}$ is in a failure state when both $S_{10}^{(4)}$ and $S_2^{(4)}$ fail. The third matrix shows a case in which the state of $S_5^{(3)}$ depends only on the state of $S_{10}^{(4)}$. The fault tree of this last case is represented by an inhibit gate without condition on the system $S_2^{(4)}$.

Table 1: Three possible state matrices for the system $S_5^{(3)}$ of Figure 1 (middle) on the basis of the states of the systems $S_{10}^{(4)}$ and $S_2^{(4)}$. On the left: $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in series; in the middle: $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in parallel; on the right: $S_5^{(3)}$ depends only on $S_{10}^{(4)}$; 1 represents the failure state.

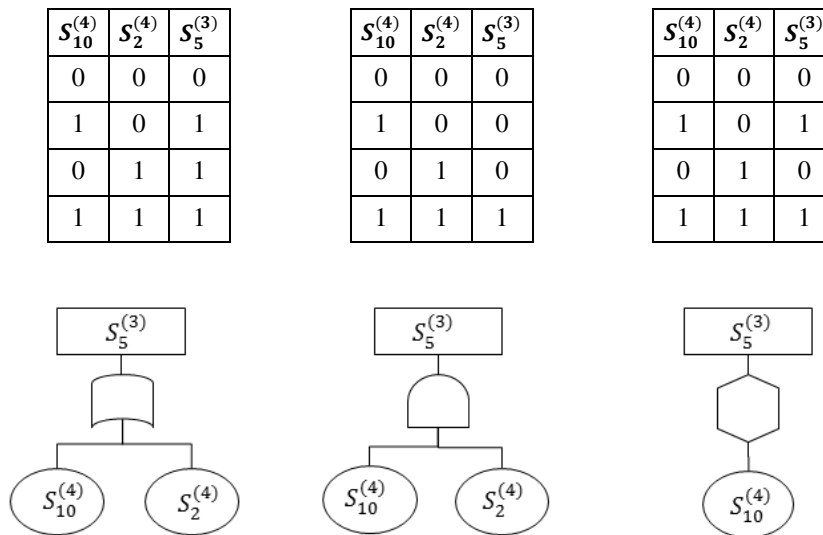


Figure 2: Corresponding fault tree representation of the state matrices reported in Table 1. On the left, $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in series (OR gate); in the middle, $S_{10}^{(4)}$ and $S_2^{(4)}$ are connected in parallel (AND gate); on the right, $S_5^{(3)}$ depends only on $S_{10}^{(4)}$ (INHIBIT gate without condition).

To define the appropriate state matrix for the systems $S_i^{(L-1)}$, $i = 1, \dots, N_S^{(L-1)}$, $L = 2, \dots, N_L$, a deep understanding of their functionality is necessary. The dependencies identified in Figure 1 (top) are a support for this analysis.

3.2. Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a system-of-systems framework

Within the system-of-systems analysis framework here purported, we wish to evaluate the safety of the critical plant H exposed to the risk from earthquakes occurrence, accounting not only for the direct effects of the earthquake on H but also for the structural and functional responses of the connected systems $S_i^{(2)}$, $i = 1, \dots, N_S^{(2)}$, inside and outside the plant, through the analysis of the underlying dependency structure. In addition, we wish to determine the capacity of recovering of the system of systems, evaluating the period necessary to restore the safety of the critical plant. To do this, we adopt the hierarchical representation of the system of systems and Monte Carlo (MC) simulation for the quantitative SPRA evaluation [27]. The simulation procedure consists of the following operative steps:

1. choose a value of magnitude with respect to which the analysis is performed;
2. compute the ground acceleration value at each of the $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, $L = N_L$, elements of the system of systems, by equation 4; $N_S^{(N_L)}$ is the number of elements at the last level of the hierarchy, i.e., in our case, the number of individual components;
3. compute the fragility, f , for all the components $S_i^{(N_L)}$, $i = 1, \dots, N_S^{(N_L)}$, of the system of systems by equation 6; f is a vector of $N_S^{(N_L)}$ values, one for each individual component in the system;
4. sample a matrix of uniform random numbers in $[0,1)$ $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N_S^{(N_L)}$, where N_T is the number of simulations;
5. determine the fault state matrix $\{g_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N_S^{(N_L)}$, by comparing the fragility, f , with the matrix $\{u_{j,k}\}$, $j = 1, \dots, N_T$, $k = 1, \dots, N_S^{(N_L)}$: if $u_{j,k} < f_k$, set $g_{j,k} = 1$; otherwise set $g_{j,k} = 0$ for $j = 1, \dots, N_T$ and $k = 1, \dots, N_S^{(N_L)}$. When $g_{j,k}$ assumes value 1, it means that in the j -th simulation the k -th component is hit by the earthquake, i.e., it enters a faulty state; otherwise, it survives. Each row of the matrix g represents the states of the $N_S^{(N_L)}$ system components in the j -th simulation;
6. determine the state of the critical plant H . This is done by propagating bottom-up through the hierarchy the faulty states of the components: the states of the $S_i^{(N_L)}$ components and the state matrix at the level $N_L - 1$ of the hierarchy are used to determine the states of the $S_i^{(N_L-1)}$ systems at the upper hierarchical level, $L = N_L - 1$, and the evaluation is repeated for the states of the systems of the level $N_L - 2$ and so on until the top level of the hierarchy, $L = 1$.

In doing so, the state of H is evaluated for each row of the matrix $\{g_{j,k}\}$, i.e., for each configuration of the system sampled. A vector $\{h_j\}$ is then recorded, whose element h_j , $j = 1, \dots, N_T$, assumes value 1 when the critical plant H is in an unsafe state and 0 otherwise;

7. estimate the probability of the critical plant H of being unsafe by computing the sample average of the values of the elements of the N_T –dimensional vector $\{h_j\}$, $j = 1, \dots, N_T$.
8. for each configuration of the system sampled that turns the critical plant H in an unsafe state, evaluate the recovery time (RT) by the following steps:
 - a. sample a matrix $\{R_{T,r,k}\}$, $r = 1, \dots, N_{R,T}$, $k = 1, \dots, N_S^{(N_L)}$, where $N_{R,T}$ is the number of recovery time simulations of the $S_i^{(N_L)}$, $i = 1, \dots, N_S^{(N_L)}$, elements of the system of systems that are in a faulty state; for each element the sampling is done from the respective recovery time distribution;
 - b. determine the recovery time of the critical plant H , computing the recovery times at each hierarchical level accounting for the configurations of the systems $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, $L = N_L, \dots, 1$, from bottom to top of the hierarchy. For example, if the systems at level L , are connected in series to the system at level $L - 1$, the recovery time of the latter is the maximum recovery time of the systems or components at the lower level L (Figure 3, left); if they are connected in parallel, the recovery time is the minimum (Figure 3, middle). In the other cases, specific evaluation should be performed. For example, if the failure of a given system $S_i^{(L)}$, $i = 1, \dots, N_S^{(L)}$, does not affect the state of another system $S_j^{(L)}$, $j = 1, \dots, N_S^{(L)}$, $j \neq i$, but plays a role in the operations of its recovery from failure it should be considered in the analysis like an increasing time for operations of recovery of the system at level $L - 1$ (Figure 3, right).

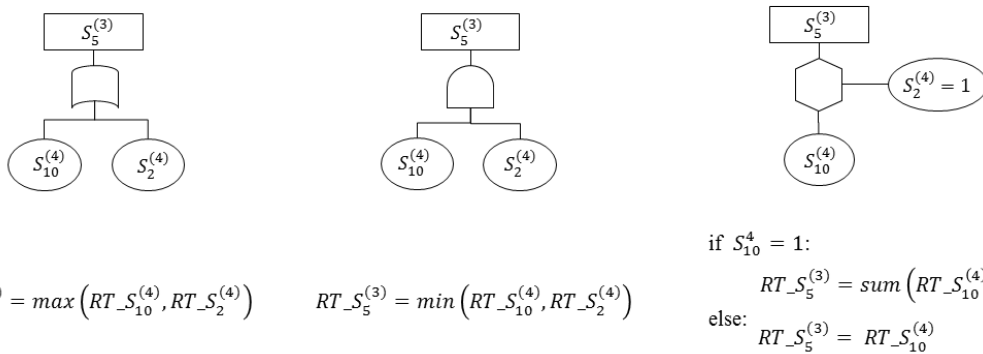


Figure 3: Computation of recovery time (RT) of the system $S_5^{(3)}$ with reference to three different configurations of the systems $S_{10}^{(4)}$ and $S_2^{(4)}$ represented in the fault tree. On the left: OR gate, the recovery time of $S_5^{(3)}$ is the maximum recovery time of $S_{10}^{(4)}$ and $S_2^{(4)}$. In the middle: AND gate, the recovery time of $S_5^{(3)}$ is the minimum recovery time of $S_{10}^{(4)}$ and $S_2^{(4)}$. On the right, INHIBIT gate: the recovery time of $S_5^{(3)}$ is the recovery time of $S_{10}^{(4)}$ but if the condition $S_2^{(4)} = 1$ is verified, the recovery time is the sum between the recovery times of $S_{10}^{(4)}$ and $S_2^{(4)}$. 1 represents the failure state.

Notice that it is assumed that infinite resources (e.g., repair teams and material) are available for the restoration process so that the recovery can be performed at the same time on all components in need. This assumption is made considering that in emergency situations all the possible means, resources and actions are deployed to keep or restore the critical plant safety. In any case, extension to the situation of limited resources does not pose significant difficulties in both the modelling and its quantification. Finally, the components are considered with binary states: fully operative or completely damaged and also the critical plant can assume only two states: fully operative or totally failed. This approximation is not realistic and leads to pessimistic results: multi-state modeling may be considered for a more realistic description, where different degrees of damage are contemplated.

4. EXEMPLIFICATION OF THE PROPOSED METHOD ON A TEST SYSTEM

We consider the mock-up problem of [9] concerning the safety of a nuclear power plant (the critical plant), provided with proper internal emergency devices, in response to an earthquake (the external hazardous event) in a system-of-systems framework, i.e., extending the boundaries of the analysis to the responses of the interconnected systems that could help keeping or restoring the plant safe state. The nuclear power plant is considered in a safe condition if it does not cause health and environmental damages, i.e., if it does not release radioactive material to the environment; to maintain this state it must be provided with electrical and water inputs to absorb the heat that it generates. We analyze the capacity of recovering of the system of systems, in terms of the period necessary to restore the safe state of the plant.

When an earthquake occurs, the critical plant may not receive the input necessary to be kept in, or restored to, a safe state due to the direct impact on its emergency devices (safety systems) and to the damages to the interconnected infrastructures. Two quantities are used to characterize the loss of functionality of the various components of the system of systems embedding the critical plant, upon the occurrence of a damaging external event:

- from the safety viewpoint, the probability that the critical plant remains in safe state;
- from the recovery viewpoint, the time needed to restore the safe state of the critical plant.

Both quantities are here computed for two values of earthquake magnitude, 5.5 and 6, on the Richter scale.

In Section 4.1, the description of the system studied is given under a number of assumptions which simplify the problem to the level needed to convey the key aspects of the conceptual system-of-systems framework, while maintaining generality. In Section 4.2, the hierarchical representation of the system and some considerations about its capacity of recovering are given. In Section 4.3, we provide the results of the evaluation of the two quantities of interest above mentioned.

4.1. Description of the system

The system under analysis is composed by a critical plant, i.e., a nuclear power plant, a water system that provides coolant useful to absorb the heat generated in the nuclear power plant, a power system that provides electrical energy for the running of the nuclear power plant and the water system, and a road network relevant to the power and water systems for the transport of material and/or plant operators.

The water and power systems are subdivided into two independent parts, external and internal to the plant; the latter one represents the emergency system of the plant which needs to obviate at the absence of input from the main external system.

In Figure 4, the physical representation of the system is reported referring to a spatial plane (x , y) with origin in the river. Table 2 reports the fragility parameters A_m , β_r and β_u , adopted in this analysis, for illustration purposes. The values for the pump and the pipe components have been taken from [28] and [29], respectively, whereas the others fragility parameters have been assumed arbitrarily by the authors to perform the study with different values. Given the large-scale system under analysis, two types of soil are considered, rock and soft. Figure 5 represents the spatial localization of the system shown in Figure 4 with reference to the reciprocal position of all the components (Figure 5, left) and to the position of the system with respect to the considered earthquake epicenter $A(70, 70)$ (Figure 5, right). The distances on the axes are expressed in kilometers.

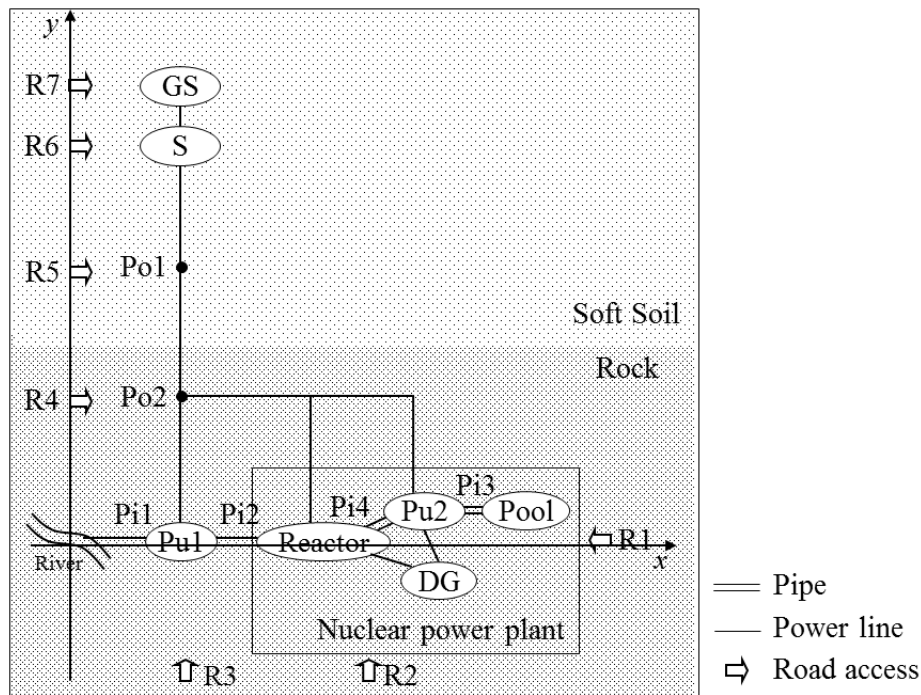


Figure 4: Physical representation of the system of systems. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access.

Table 2: Fragility parameters used in the present work.

	A_m	β_r	β_u
Generation station	0.7	0.3	0.1
Substation	0.9	0.4	0.3
Power Pole	0.8	0.2	0.2
Diesel Generator	0.7	0.4	0.2
Pipe	1.88	0.43	0.48
Pump	0.2	0.2	0.3
Pool	0.2	0.1	0.1
Road	0.3	0.3	0.2

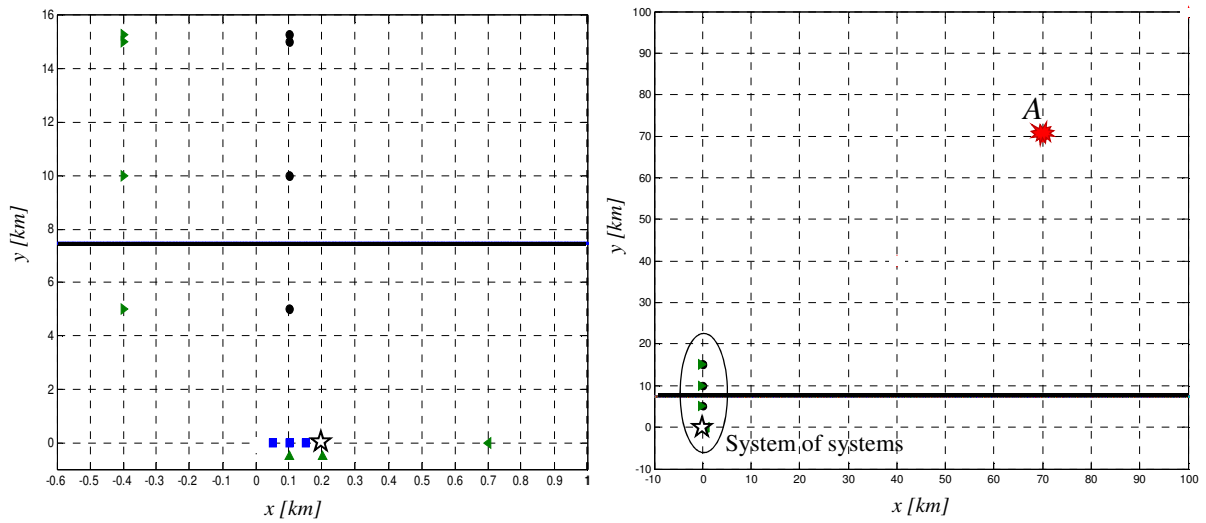


Figure 5: Left: spatial localization of the nuclear power plant (star) with respect to the components of the electric power system (circle, from top to bottom: Generation Station, Substation, Pole 1, Pole 2), water system (square, from left to right: Pipe 1, Pump 1, Pipe 2) and road transportation (triangle, from top to bottom and from left to right: R7, R6, R5, R4, R3, R2, R1). Right: spatial localization of the system of systems with respect to the earthquake's epicenter A(70, 70). The horizontal bold line in both Figures represents the division between soft soil (above the line) and rock (below the line).

In Figure 6, the system-of-systems representation is given by a conceptual map showing the components of the systems and their relationships, intra- and inter-systems. The intra-system dependencies are represented by the solid lines, the inter-system ones by dashed lines and those with the critical system by the bold lines. In addition, in the Figure the dependence of the system of systems on the type of soil on which the infrastructures rest is illustrated.

The external water distribution system (Figure 6, left) is formed by a pump and pipes that carry the water. The external power distribution system (Figure 6, center) is composed by the following elements: a generation station that produces the electrical energy, a substation that transforms the voltage from high to low, and poles that support power lines.

The components of the emergency water and power distribution systems inside the plant are shown in Figure 6 on the right. The first system is composed by the same elements of the

corresponding external system considering in addition an artificial reservoir (i.e., the source of water), whereas the power system includes only the emergency diesel generators.

The elements considered for the transportation system are the roads (Figure 6, top). The state of this system is important for access of the materials and operators that are needed to restore the components required for the safe state of the critical plant. Given their role, they are considered as recovery supporting elements (see Section 3.1).

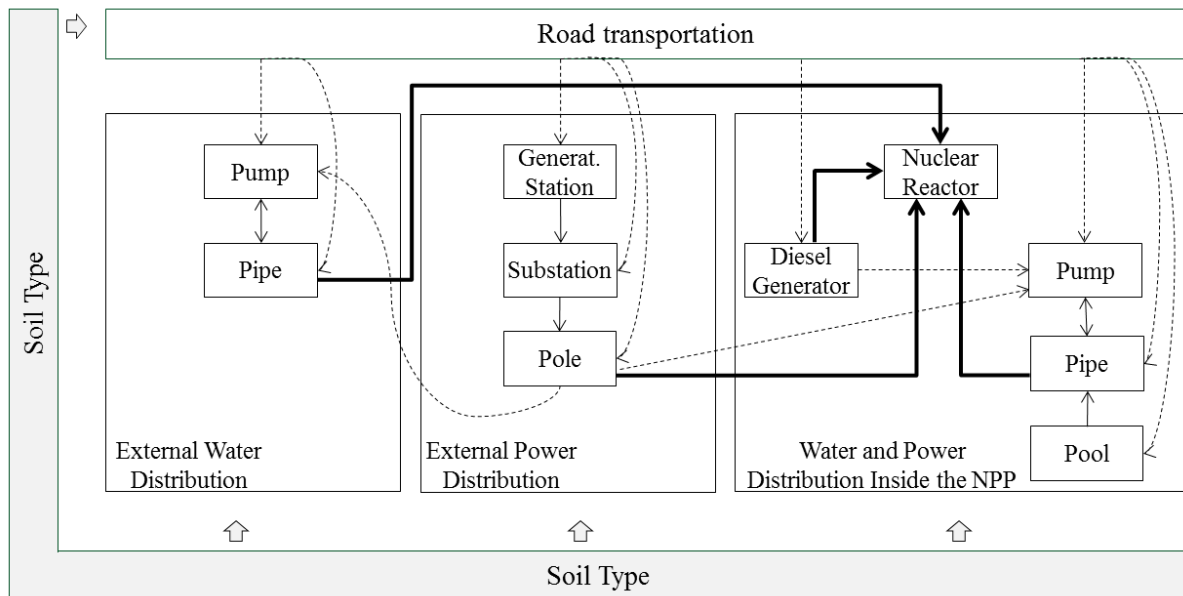


Figure 6: System of systems: conceptual map; the links represent the intra-systems dependencies (solid lines), the inter-systems dependencies (dashed lines) and the dependencies of the nuclear power plant on its interconnected systems (bold lines).

The inter-system dependencies are modeled as links connecting components of the power, water and road transportation systems (Figure 6, dashed lines); these links are conceptually similar to those linking components of the individual systems (intra-systems dependencies), and are considered bidirectional with respect to the “flow” of dependence between the connected systems. For example, the external water system depends on the external power system as the pump needs electrical energy to work. Notice that this relation is expressed by a link from the pole to the pump because the first one, supporting the power lines, is the closest element to the pump that carries the power (the same reason explains the connection of the pole to the nuclear reactor and to the pump inside the nuclear power plant). While the pump of the external water system can receive electrical energy only from the external power distribution network, it is assumed that the pump inside the nuclear power plant can obtain it from both the external and internal power systems.

The road transport network allows access to the components of the power and water systems for transporting material (e.g., fuel) and/or operators for operation and/or recovery.

The transport system is composed by seven interdependent road access points to the components of the power and water systems. They are distributed as follows: one road access is available for the components outside the nuclear power plant and two road accesses for

those inside, i.e., the components outside the nuclear power plant can only be reached by one road access, whereas the ones inside by two road accesses (the same two accesses are provided for all the components inside) (Figure 4). In particular, the components of the external power system are considered to have a different road access because they are far from each other (the minimum distance is 300 m between the generation station and the substation, Figure 5 left), the components of the external water system have the same road access, R3, because they are located close to each other (the total distance from the river to the nuclear power plant is 200 m, Figure 5 left) and the components of the power and water systems inside the nuclear power plant have the same two road accesses, R1 and R2, since they are contained in the same building.

Among these road access points, only the one connected to the generation station, R7 in Figure 4, has an impact on the state of the system of systems because it contributes to the running of the generation station, carrying materials and operators. On the contrary, the other road accesses have no direct impact on the state of the system of systems since they are used only to repair the elements that enter in a faulty state. Therefore, their contribution is not of interest for the evaluation of the safety of the critical plant, but they are relevant for the analysis of the capacity of recovering of the system of systems.

In this work we have not considered i) the power lines that, being aerial elements, are not directly affected by an earthquake and ii) the river, i.e., the source of water of the external water system, that it is assumed to be always available. Other aspects could be introduced in the analysis as i) the influence of the design, construction and materials of the infrastructures considered, ii) the supply of fuel and materials for plant operation, and iii) the maintenance tasks. However, in view of the methodological character of this work, for the sake of simplicity, we have not included them in the modelling.

4.2. Hierarchical representation of the system of systems and its capacity of recovering

From the conceptual map shown in Figure 6, the connections between the physical elements of the system of systems are presented in Figure 7. The solid, dashed and bold lines represent the intra-system dependencies, the inter-systems dependencies and the links to the nuclear power plant (NPP), respectively. The clusters taken into account in the analysis are identified in Figure 8, and they are structured hierarchically in Figure 9.

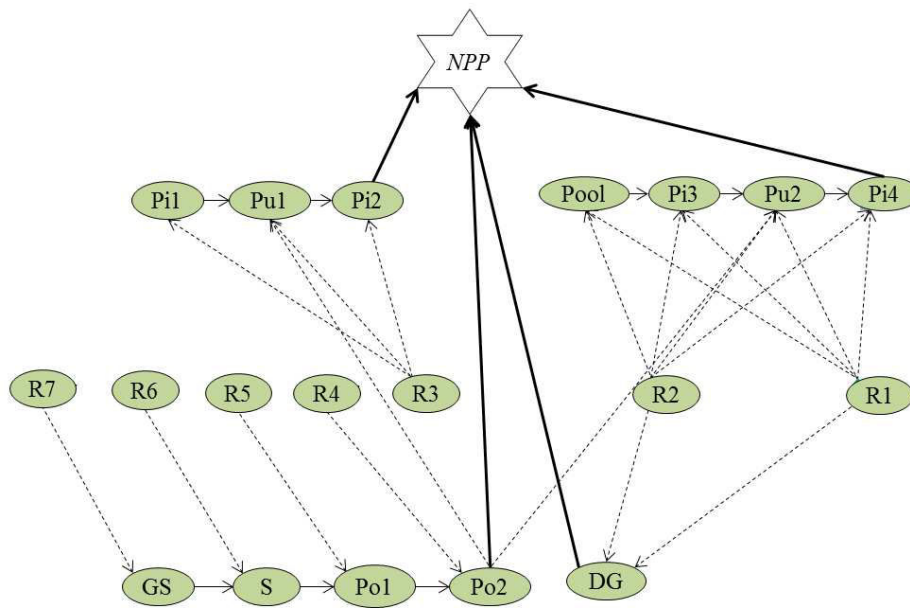


Figure 7: Dependencies among the components of the system of systems; the links represent the intra-systems dependencies (solid lines), the inter-systems dependencies (dashed lines) and the dependencies of the nuclear power plant (NPP) on its interconnected systems (bold lines). GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access.

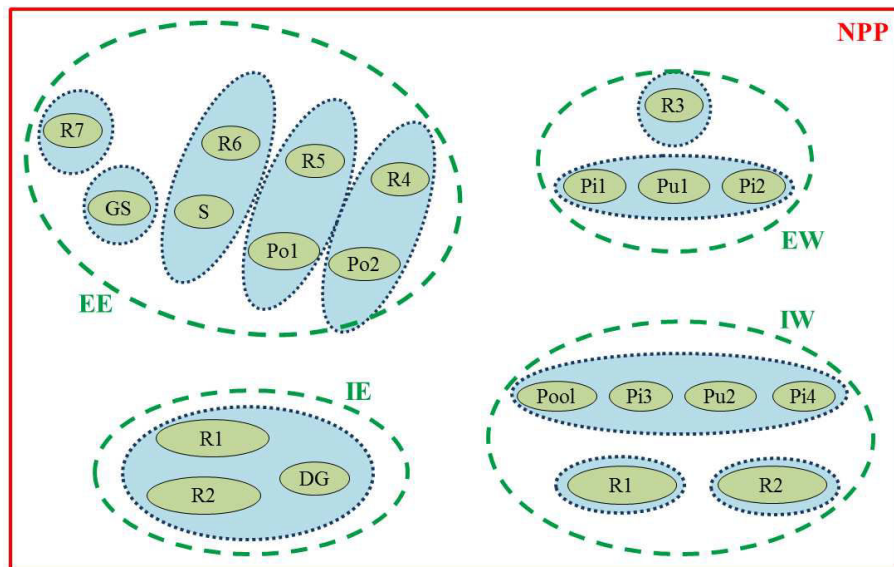


Figure 8: Representation of the system of systems highlighting its underlying structure of four hierarchical levels represented by the rectangular (level 1), the dashed (level 2), the dotted (level 3) and the solid (level 4) oval shapes. NPP: Nuclear Power Plant, EE: External Energy, EW: External Water, IE: Internal Energy, IW: Internal Water, GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access.

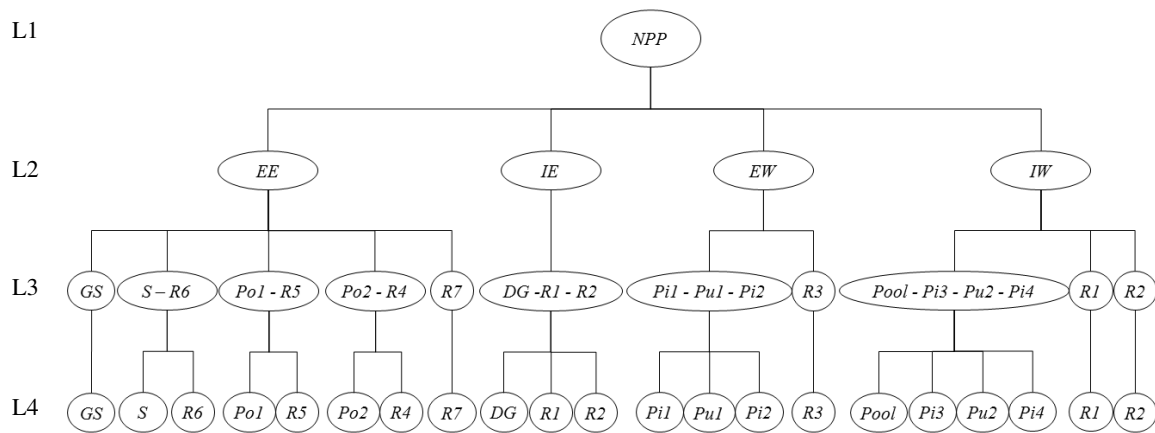


Figure 9: Hierarchical representation of the system of systems. NPP: Nuclear Power Plant, EE: External Energy system, EW: External Water system, IE: Internal Energy system, IW: Internal Water system, GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, Pu: Pump, DG: Diesel Generator, R: Road access, L: Level.

The nuclear power plant is at the top (level 1) of the hierarchy. Its safety is supported by the power and water systems that are partitioned, at the level 2, into external and internal parts: external energy (EE), internal energy (IE), external water (EW) and internal water (IW). The road accesses are the recovery supporting elements and, as explained in Section 3.1, they belong to the systems to which they provide support, i.e., in this test system they belong to the corresponding EE, IE, EW and IW systems. The level 3 is, then, composed by single individual components or road accesses or a combination of them, and the level 4, the most specified level, is formed by the individual elements (components and road accesses) of the system of systems. Notice that only the recovery supporting elements can belong to different systems (or groups), e.g., R1 and R2 are within both the IE and IW systems, whereas the other components appear in just one system, e.g., the pole Po2 belongs to the EE system.

The roads (elements R1, R2, R3, R4, R5, R6) are used only for the recovery task and, thus, do not influence the state of other parts of the system of systems, i.e., their failures do not cause the stop of the running of other components. On the contrary, they play a role for system recovery because if they are damaged they have to be recovered to allow reaching the system components that are failed for repairing them, and eventually restoring the safety of the critical plant. In other words, if a component fails, the road access to it has to be available for its recovery. For this reason, the components of the level 3 of the hierarchy are grouped together with the corresponding road, e.g., the substation (S) is grouped with the road R6, the diesel generator (DG) is grouped with the two roads R1 and R2, etc. Instead, when a road is connected with more than one component, the first grouping is among the components and, then, at the next higher level, the components are grouped with the road, e.g., the components of the external water systems (pipes and pump) are grouped together at level 3 and then they are grouped with the road R3 at level 2. This grouping at level 3 allows highlighting the contribution of a road with respect to all the components (one or more) to which it provides access.

The road R7, plays a role in the external energy subsystem which goes beyond the access for recovery, as it provides the generation station with the access for the operators and materials necessary to its functioning. Therefore, the damage to this access road can cause the stop of the generation station and, as a consequence, the failure of the external energy subsystem. For this reason, it is not grouped with the generation station at the third hierarchical level.

The capacity of recovering of the system of systems is quantified in terms of the time needed to recover the safe state of the critical plant. To compute this, the evolution in time of the system of systems is included in the SPRA framework. For the sake of simplicity, damages from aftershocks are not considered in the time-dependent analysis.

As illustrated in the procedure of Section 3.2, the recovery time of the nuclear power plant is computed starting from the recovery time of the individual components at the bottom level of the hierarchy which is climbed from bottom to top through the configurations of the components or systems at each level.

To account for the uncertainty in the duration of the recovery, lognormal distributions have been associated to the recovery time of the individual components. Table 3 shows the means and the variances used in this study; these values have been taken on the basis of the following consideration. The time to recover a component depends on its size, its location, and the type of damage and the easiness to find the failure. It is assumed that, the components inside the nuclear power plant need more time for the recovery than the components outside. In particular, this happens when it is necessary to replace part of the component or the entire component given its huge dimensions and the difficulty to operate inside the plant.

For this reason, we have assumed that the mean of the time needed to recover the pump inside the nuclear power plant is larger than that needed for the pump outside. The large mean value of the time to recover the pool is due to its size, location inside the plant and difficulty in restoration. The time to repair a pipe could be very short (even few hours), but we have assumed a mean value equal to 4 days to account for the difficulty in locating the break. The diesel generator has a time of repair with a high uncertainty (variance equal to 5), because it may vary significantly depending on the type of damage. The components with lowest mean value of the recovery time are the power pole, the road, the generation station and the substation that are outside the plant; the latter are affected by large uncertainty (variances of 5 and 10, respectively), because their recovery depends on the intensity of the damage, e.g., a generation station can be slightly perturbed by the earthquake and its repairing can last few hours but it can also be destroyed and in this case the time to build it again is obviously much higher.

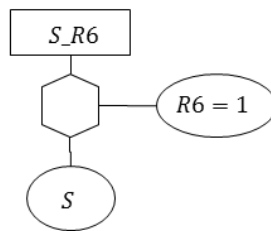
Table 3: Parameters of the lognormal distributions that describe the recovery time of the single components of the system of systems.

Components	Mean [days]	Variance
Pump (inside the plant)	75	3
Pump (outside the plant)	5	3
Pipe	4	3
Pool	75	3
Diesel Generator	30	5
Power pole	1.5	3
Generation Station	1	10
Substation	1	5
Road	2	3

By way of example, the explanation of the procedure for the evaluation of the time to recover power at the hierarchical level 3 and 2 for the test system under analysis is illustrated in the following, with reference to the Figures 10 – 11.

At level 3 of the hierarchy, there are five groups for the external energy (EE) system and one for the internal energy (IE) system. For the individual components of the EE system, i.e., generation station and road R7, the recovery times are described by lognormal distributions whose parameters are reported in Table 3, whereas for the groups made by the pairs of components and road access, e.g., substation and road R6 (S_R6), the recovery time is computed on the basis of the relations among them represented by the fault tree in Figure 10. For the group of the IE system, the fault tree of the recovery time of the triplet “DG_R1_R2” is reported in Figure 11.

As reported in the procedure of Section 3.2, given the assumption of unlimited resources for restoration, the recovery starts at the same time (i.e., immediately after the earthquake) on all the components in need. Actually, one exception is made for those components whose access is disrupted; in this case, the recovery is sequential: first, the access to them is restored and, then, components recovery starts.

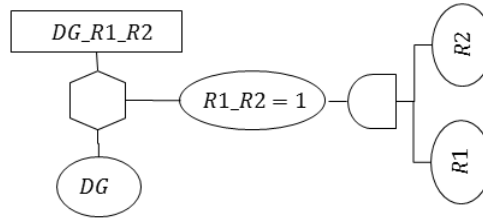


```

if R6 = 1:
    RT_S_R6 = sum(RT_S, RT_R6)
else:
    RT_S_R6 = RT_S

```

Figure 10: Fault tree representation for the computation of the recovery time (RT) of the pair “S_R6” at level 3 of the hierarchy; S: Substation, R: Road access. 1 represents the failure state.

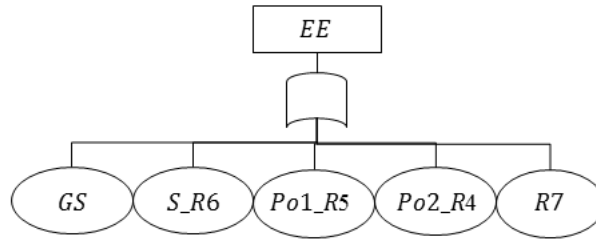


```

if  $R1\_R2 = 1$ :
     $RT\_DG\_R1\_R2 = \text{sum}(RT\_DG, \min(RT\_R1, RT\_R2))$ 
else:
     $RT\_DG\_R1\_R2 = RT\_DG$ 
    
```

Figure 11: Fault tree representation for the computation of the recovery time (RT) of the triplet “DG_R1_R2” at level 3 of the hierarchy; DG: Diesel Generator, R: Road access. 1 represents the failure state.

At level 2, the recovery time of the EE system is the maximum recovery time of the elements of level 3, since they are connected in series (Figure 12). The recovery time of the IE system is that of the triplet “DG – R1 – R2” computed at level 3.



$$RT_{EE} = \max(RT_{GS}, RT_{S_R6}, RT_{Po1_R5}, RT_{Po2_R4}, RT_{R7})$$

Figure 12: Fault tree representation for the computation of the recovery time (RT) of the external energy system (EE) at level 2 of the hierarchy; GS: Generation Station, S: Substation, Po: Pole, R: Road access.

Analogous reasoning is used to define the recovery time for the water system at level 3 and 2.

To compute the recovery time at level 1, the logic relations (LR) between the external and internal energy and water systems at level 2 are given in Figure 13 and the corresponding state matrix of the nuclear power plant is reported in Table 4.

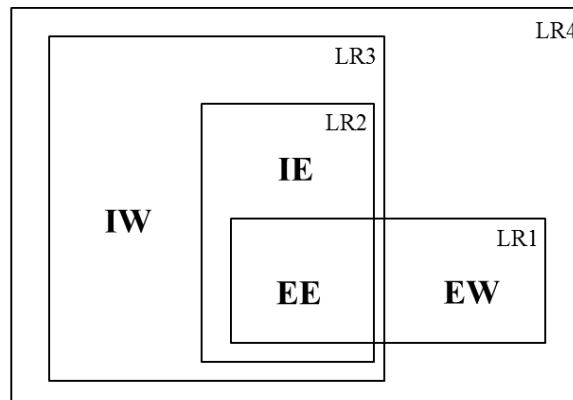


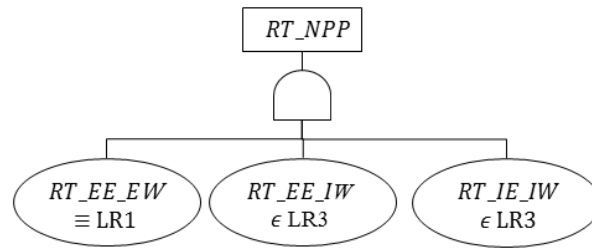
Figure 13: Schematic representation of the relations (LR) that exist between the external energy (EE), internal energy (IE), external water (EW) and internal water (IW) systems at the level 2 of the hierarchy.

Table 4: State matrix of the nuclear power plant (NPP) (level 1) on the basis of the states of the external energy (EE), internal energy (IE), external water (EW) and internal water (IW) systems (level 2); 1 represents the failure state.

EE	IE	EW	IW	NPP
1	1	1	1	1
1	1	1	0	1
1	1	0	1	1
1	1	0	0	1
1	0	1	1	1
1	0	1	0	0
1	0	0	1	1
1	0	0	0	0
0	1	1	1	1
0	1	1	0	0
0	1	0	1	0
0	1	0	0	0
0	0	1	1	1
0	0	1	0	0
0	0	0	1	0
0	0	0	0	0

The EE and EW systems are grouped together in the relation LR1 because the EW system needs the EE system to work. The relation LR2 considers the IE and EE systems with respect to the relation LR3, since the IW system can receive electrical inputs both from the IE and EE systems and at least one of these two systems must work. The relation LR4 includes all the relations LR1, LR2 and LR3 and represents the nuclear power plant.

The recovery time of the nuclear power plant (Figure 14) is obtained by the minimum of the recovery time of the systems involved in the relations LR1 and LR3, since its safety is guaranteed when it is provided with both energy and water inputs. Therefore it is computed by the minimum recovery time of the pairs “EE – EW”, “EE – IW” and “IE – IW”.



$$RT_{NPP} = \min(RT_{EE_EW}, RT_{EE_IW}, RT_{IE_IW})$$

Figure 14: Sketch of the computation of the recovery time (RT) of the nuclear power plant (NPP) at level 1 of the hierarchy on the basis of the recovery time of the external energy (EE) internal energy (IE), external water (EW) and internal water (IW) systems, grouped according the relations LR1 and LR3 identified in Figure 13.

For the sake of simplicity, the assumption has been made that the internal emergency devices will not stop functioning once successfully started. In fact, the diesel generator can be refueled in operation without causing an interruption of the production of the electrical energy and the pool of the internal water system has been assumed of infinite capacity.

4.3. Results

The Monte Carlo simulation for Seismic Probabilistic Risk Assessment illustrated in Section 3.2 has been applied to the test system of Section 4.1 for two values of earthquake magnitudes, $M= 5.5$ and $M = 6$ on the Richter scale at the epicenter of coordinates $(x, y) = (70, 70)$ (Figure 4). The number of simulations (N_T) of the components configurations for each magnitude value is 2000 and the number of recovery time simulations (N_{R_T}) for each configuration that turns the nuclear power plant (NPP) in an unsafe state is 5000. These numbers have been arbitrarily chosen by the authors in such a way to reach a good trade-off between precision of the results and computational cost.

Figure 15 shows the estimated probabilities (under all assumptions made) that the nuclear power plant reaches an unsafe state upon the occurrence of an earthquake of magnitude equal to 5.5 (left) and 6 (right) on the Richter scale. The estimated conditional probabilities of failure of the external energy (EE), external water (EW), internal energy (IE) and internal water (IW) systems, given that the NPP has entered into an unsafe state, are also indicated.

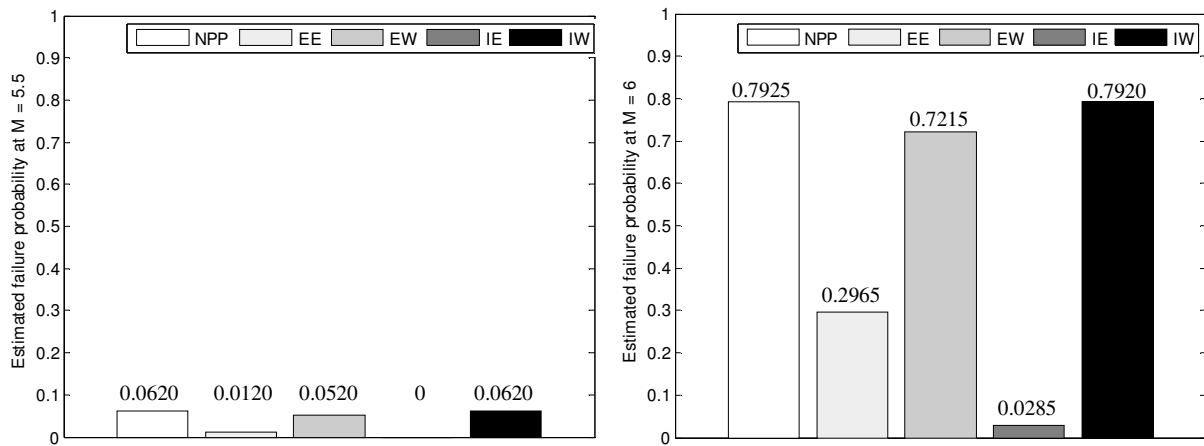


Figure 15: Estimate of the probability that the nuclear power plant (NPP) reaches an unsafe state upon occurrence of an earthquake of magnitude equal to 5.5 (left) and 6 (right) on the Richter scale, and the estimates of the conditional probability of failure of the external energy (EE), external water (EW), internal energy (IE) and internal water (IW) systems, given that the NPP has reached an unsafe state.

As expected, the higher the magnitude of the earthquake, the higher is the probability that the safety of the nuclear power plant cannot be assured.

The estimated probabilities of failure of the IW and EW systems are similar to that of the NPP at both magnitudes. This is because the two systems mostly contribute to the reaching of the NPP unsafe state. A qualitative analysis of the fragility values of the elements of the power and water systems, given in Table 5 in decreasing order for $M = 5.5$, on the left, and $M = 6$, on the right, shows that the first two components with higher fragility values are the pumps of the IW and EW systems. At magnitude 5.5 on the Richter scale, the third element in Table 5 is the road R7 that belongs to the EE system followed by the DG of the IE system that never fails in the simulation performed, due to its low fragility value ($2.52 \cdot 10^{-3}$). At magnitude 6 on the Richter scale, the third element with higher fragility is represented by the pool that in the ranking at magnitude 5.5 is in the 10th position; this represents a further weak element of the internal water system. The other components remain in the same ranking order both at magnitude 5.5 and 6 on the Richter scale, with increased fragility values for the higher magnitude.

Table 5: Conditional probability of failure of the components of the system of systems given an earthquake of magnitudes 5.5 (left) and 6 (right) on the Richter scale. The values are reported in decreasing order. GS: Generation Station; S: Substation; R: Road access; Po: Pole; Pi: Pipe; DG: Diesel Generator; Pu: Pump; M: Magnitude.

	M = 5.5		M = 6
Pu2	3.78E-01	Pu2	9.32E-01
Pu1	1.27E-01	Pu1	7.46E-01
R7	3.66E-02	Pool	3.80E-01
DG	2.52E-03	R7	3.08E-01
S	1.94E-03	DG	2.86E-02
Pi4	7.40E-04	S	2.74E-02
Pi3	7.40E-04	Pi4	9.64E-03
Pi2	7.35E-04	Pi3	9.64E-03
Pi1	7.27E-04	Pi2	9.61E-03
Pool	4.57E-05	Pi1	9.53E-03
GS	7.05E-06	GS	1.13E-03
Po2	6.54E-10	Po2	1.00E-05
Po1	1.01E-10	Po1	5.28E-06

We now proceed with the evaluation of the capacity of recovering of the system of systems, starting from the top level of the hierarchy (recovery of the critical plant safety) and proceeding downward with the analysis of the lower levels to identify the causes and major contributors to the higher levels. The criticality importance measure [13], $I_i^{Cr,L}(t)$, of the component (or group) i at level L , $L = 2, \dots, N_L$, of the hierarchy at time t is used to guide the analysis through the hierarchical model. It is defined as the probability that the component (or group) i at level L , $L = 2, \dots, N_L$, of the hierarchy is critical for the system and failed at time t , given that the system is failed at time t :

$$I_i^{Cr,L}(t) = \frac{I_i^{B,L+1}(t) \cdot (1 - r_i^{L+1}(t))}{1 - R(\mathbf{r}^{L+1}(t))} \quad (7)$$

where $r_i^{L+1}(t)$ is the reliability of the component (or group) i at level $L+1$ of the hierarchy, $\mathbf{r}^{L+1}(t)$ is the vector of reliabilities of the components (or groups) at level $L+1$ of the hierarchy, $R(\mathbf{r}^{L+1}(t))$ is the system reliability, dependent on the reliabilities of the individual components (or groups) at level $L+1$ of the hierarchy and on the system configuration, $I_i^{B,L+1}(t)$ is the Birnbaum's measure of importance of the i -th component (or group) at level $L+1$ of the hierarchy and it is defined as $I_i^{B,L+1}(t) = \frac{\partial R(\mathbf{r}^{L+1}(t))}{\partial r_i^{L+1}(t)}$ [13].

With respect to the test system under analysis, the system reliability (level 1) depending on the reliabilities of the groups of level 2 and on their logic relations reported in Table 4, has been computed as follows:

$$\begin{aligned} R(\mathbf{r}^2(t)) = & (1 - r_{EE}^2(t))r_{IE}^2(t)(1 - r_{EW}^2(t))r_{IW}^2(t) + (1 - r_{EE}^2(t))r_{IE}^2(t)r_{EW}^2(t)r_{IW}^2(t) + \\ & r_{EE}^2(t)(1 - r_{IE}^2(t))(1 - r_{EW}^2(t))r_{IW}^2(t) + r_{EE}^2(t)(1 - r_{IE}^2(t))r_{EW}^2(t)(1 - r_{IW}^2(t)) + \\ & r_{EE}^2(t)(1 - r_{IE}^2(t))r_{EW}^2(t)r_{IW}^2(t) + r_{EE}^2(t)r_{IE}^2(t)(1 - r_{EW}^2(t))r_{IW}^2(t) + \end{aligned}$$

$$r_{EE}^2(t)r_{IE}^2(t)r_{EW}^2(t)(1 - r_{IW}^2(t)) + r_{EE}^2(t)r_{IE}^2(t)r_{EW}^2(t)r_{IW}^2(t) = r_{EE}^2(t)r_{EW}^2(t) + r_{EE}^2(t)r_{IW}^2(t) + r_{IE}^2(t)r_{IW}^2(t) - r_{EE}^2(t)r_{EW}^2(t)r_{IW}^2(t) - r_{EE}^2(t)r_{IE}^2(t)r_{IW}^2(t)$$

The reliability $r_{EE}^2(t)$, $r_{IE}^2(t)$, $r_{EW}^2(t)$ and $r_{IW}^2(t)$ of the EE, IE, EW and IW systems, respectively, at level 2 of the hierarchy, depend on the reliability of the groups at level 3, that in turns depend on the individual components at level 4. For example, the reliability $r_{EW}^2(t)$ at level 2 depends on the reliability of the groups Pi1-Pu1-Pi2 and R3 at level 3 (Figure 9); the first group is composed by three components, Pi1, Pu1 and Pi2, in series, thus, its reliability is the product of the single reliability of the corresponding elements at level 4 of the hierarchy ($r_{Pi1-Pu1-Pi2}^3(t) = r_{Pi1}^4(t)r_{Pu1}^4(t)r_{Pi2}^4(t)$), whereas the second group, having no impacts on the state of the system EW (as explained in Section 4.2) is not considered in the computation of the reliability $r_{EW}^2(t)$. The reliabilities of the individual components at level 4 are the complement to 1 of the corresponding conditional probabilities of failure, given a magnitude value, reported in Table 5.

Figure 16 shows the probability density functions (PDFs) (on the left) and the respective cumulative distribution functions (CDFs) (on the right) of the time it takes to restore the safety of the nuclear power plant when an earthquake of magnitude 5.5 (solid line) and 6 (dashed line) on the Richter scale occurs. The 95th percentile of the distributions is used as indicator of the time it takes to recover safety. As expected, at the lower magnitude the time for recovering safety is shorter.

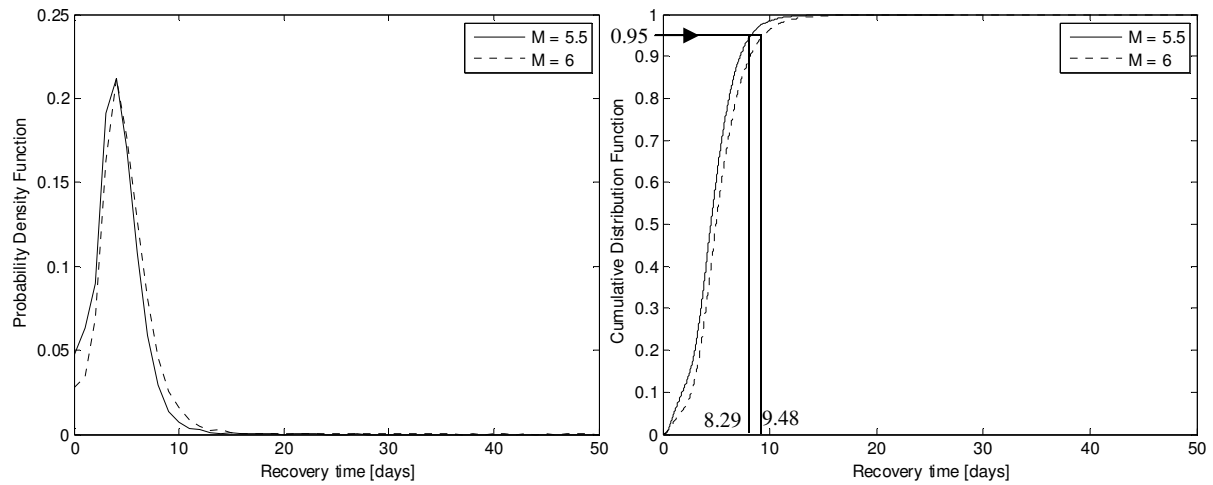


Figure 16: Left: probability density functions of the recovery time of the safety of the nuclear power plant when an earthquake of magnitude 6.5 (solid line) and 7 (dashed line) on the Richter scale occurs. Right: corresponding cumulative distribution functions.

In Table 6, the values of the criticality importance measure of the systems at level 2 (external and internal power and water systems) with respect to the level 1 of the hierarchy (critical plant) are reported. It can be seen that the EW and IW systems have a significantly higher impact than EE and IE systems both at lower and higher magnitudes.

Table 6: Criticality importance measures of the external (E) and internal (I) power (E) and water (W) systems for magnitudes equal to 5.5 and 6 on the Richter scale.

	M = 5.5	M = 6
$I_{EE}^{cr,2}$	0.2081	0.0984
$I_{IE}^{cr,2}$	9.8E-04	4.8E-04
$I_{EW}^{cr,2}$	0.7614	0.6059
$I_{IW}^{cr,2}$	0.9984	0.9883

Figures 17 and 18 show the probability density functions of the time it takes to recover the internal and external parts of the power and water systems (level 2 of the hierarchy) after the occurrence of an earthquake of magnitude equal to 5.5 and 6 on the Richter scale, respectively.

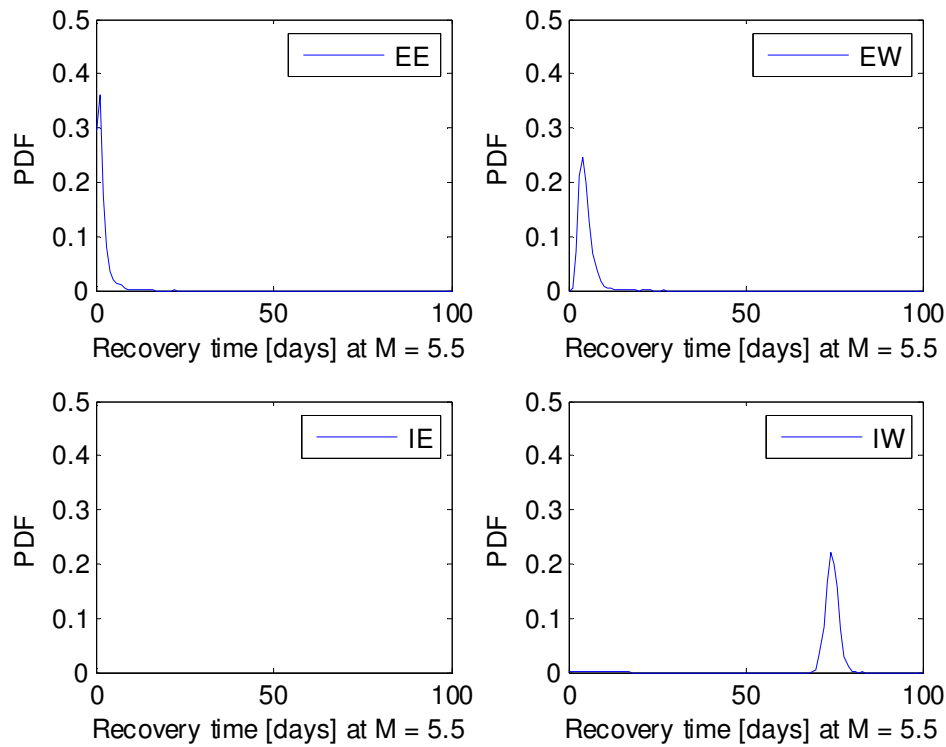


Figure 17: Probability density functions of the recovery time of the internal (I) and external (E) parts of the power (E) (left) and water (W) (right) systems, given the occurrence of an earthquake of magnitude (M) equal to 5.5 on the Richter scale.

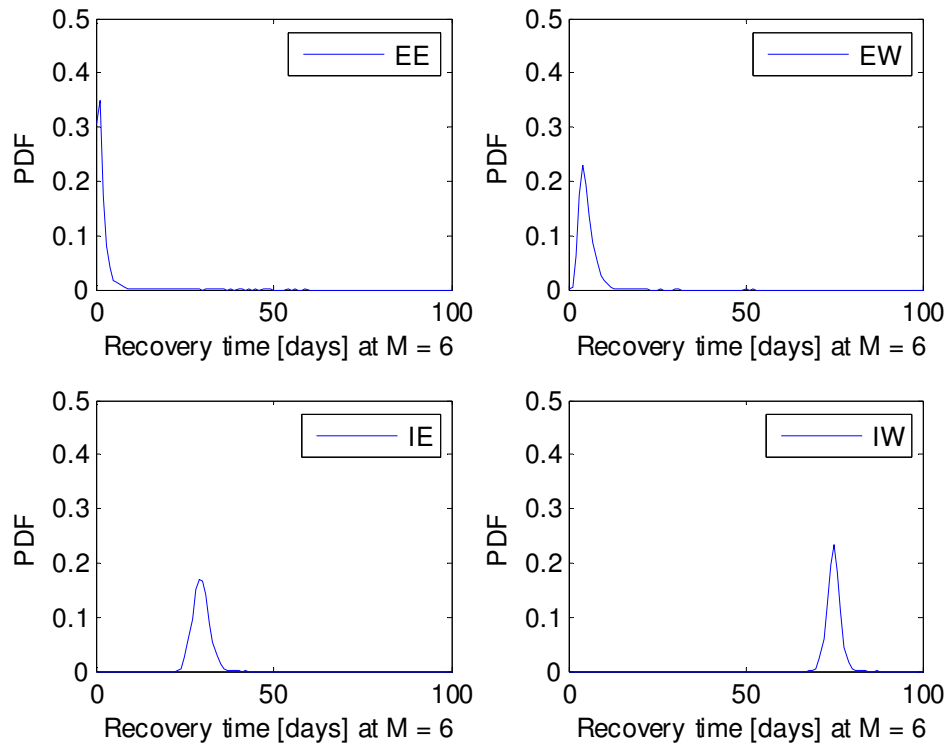


Figure 18: Probability density functions of the recovery time of the internal (I) and external (E) parts of the power (E) (left) and water (W) (right) systems, given the occurrence of an earthquake of magnitude (M) equal to 6 on the Richter scale.

At magnitude 5.5 on the Richter scale, the recovery time of the IE system is not present since this system has never failed in the simulation.

At magnitude 6, the recovery times of the external parts of the energy and water systems are concentrated at values lower than the recovery times of the internal parts, which means that the recovery times of the systems at level 2 depend on the recovery of the external parts.

Figure 19 shows the probability density functions of the time it takes to recover the groups of the external water system at the level 3 of the hierarchy, for an earthquake of magnitude 6 on the Richter scale.

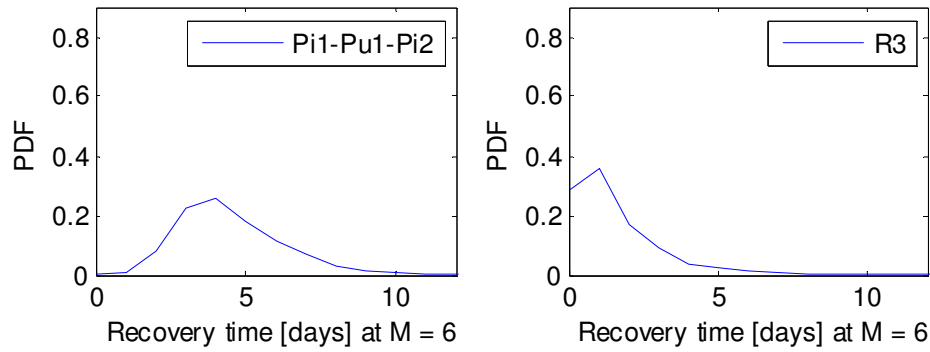


Figure 19: Probability density functions (PDFs) of the recovery time of the groups at level 3 of the hierarchy for the external water system, given the occurrence of an earthquake of magnitude (M) equal to 6 on the Richter scale.

The group of components Pi1-Pu1-Pi2 contributes mostly to the recovery time of the EW system since the state of the road R3 has no impact in the state of the EW system, as explained in Section 4.2; then, the criticality importance measure of Pi1-Pu1-Pi2 is 0.6059, i.e., it is equal to $I_{EW}^{cr,2}$ as shown in Table 6.

Figure 20 illustrates the recovery time distributions of the components Pu1, Pi1 and Pi2 at level 4 of the hierarchy, and Table 7 reports the corresponding criticality importance measure values: at level 4, the major contributor to the recovery time is the component Pu1 that has the highest importance measure value equal to 0.5906.

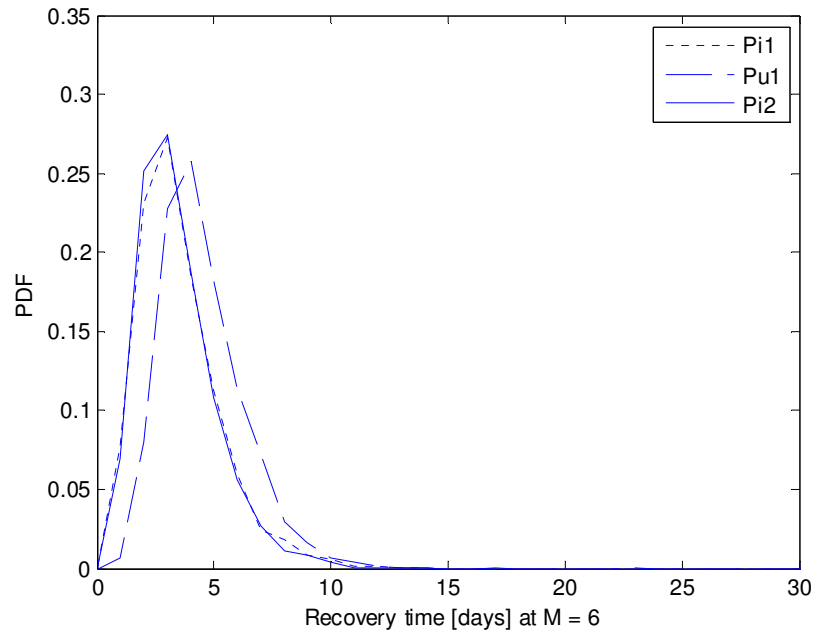


Figure 20: Probability density functions (PDFs) of the recovery time of the components Pi1, Pu1 and Pi2 given the occurrence of an earthquake of magnitude (M) equal to 6 on the Richter scale.

Table 7: Criticality importance measures of the groups at the level 4 of the hierarchy, for an earthquake of magnitude (M) equal to 6 on the Richter scale.

	M = 6
$I_{Pi1}^{cr,4}$	1.93E-03
$I_{Pu1}^{cr,4}$	5.91E-01
$I_{Pi2}^{cr,4}$	1.95E-03

A similar analysis on the internal water system (here not reported, for brevity), leads to the conclusion that the pump and the pool are the most relevant components for the time of recovery of such system.

3. CONCLUSIONS

We have adopted a system-of-systems framework previously proposed by the authors for the analysis of the risk of a critical plant (a nuclear power plant in the example worked out) exposed to hazardous external events (earthquakes in the example worked out), so as to account for the influence of the interdependent infrastructures in which the plant is embedded. We have represented the system of systems with a hierarchical model and used Monte Carlo simulation for its probabilistic evaluation in terms of the safety of the nuclear power plant and its capacity of recovering, measured in terms of the time needed to restore safety.

The plus of this framework is that it allows performing a systematic analysis through the hierarchical levels of the model, and identifying the contribution to the safety recovery time of the system-of-systems individual elements (here measured by the criticality importance measure). The results which are obtained by such type of analysis can be useful to point out which systems are recovered early and which take more time to be recovered. These findings can help identifying margins for improvement of the structural/functional responses of the critical elements, for improving the global recovery of the system of systems so as to increase the safety of the critical plant. In the end, they can inform decision makers in their planning choices of actions for increasing the safety of critical plants.

Future work will be devoted to explore other system modeling and analysis approaches for comparison, like for example Multilevel Flow Modelling (MFM) [30], Stochastic Flowgraphs [31], Goal Tree Success Tree – Master Logic Diagram (GTST – MLD) [32], with the aim of pointing out limitations and benefits with respect to their application.

REFERENCES

- [1] International Atomic Energy Agency. The great east Japan earthquake expert mission – IAEA international fact finding expert mission of the Fukushima Dai-ichi NPP accident following the great east Japan earthquake and Tsunami. Mission Report 24 May – 2 Jun 2011. 162 p.

- [2] Adachi T, Ellingwood BR. Serviceability of earthquake-damaged water systems: Effects of electrical power availability and power backup systems on system vulnerability. *Reliability Engineering & System Safety*. 2008; 93(1):78-88.
- [3] Aven T. On Some Recent Definitions and Analysis Frameworks for Risk, Vulnerability, and Resilience. *Risk Analysis*. 2011; 31(4):515-522.
- [4] Eusgeld I, Nan C, Dietz S. "System-of-systems" approach for interdependent critical infrastructures. *Reliability Engineering & System Safety*. 2011; 96(6):679-686.
- [5] Haimes YY. On the Complex Definition of Risk: A Systems-Based Approach. *Risk Analysis*. 2009; 29(12):1647-1654.
- [6] Johansson J, Hassel H. An approach for modelling interdependent infrastructures in the context of vulnerability analysis. *Reliability Engineering & System Safety*. 2010; 95(12):1335-1344.
- [7] Kröger W, Zio E. *Vulnerable systems*. London: Springer; 2011. 63 p.
- [8] Wang S, Hong L, Chen X. Vulnerability analysis of interdependent infrastructure systems: a methodological framework. *Physica A: Statistical Mechanics and its Applications*. 2012; 391(11): 3323-3335.
- [9] Ferrario E, Zio E. A system-of-systems framework of Nuclear Power Plant Probabilistic Seismic Hazard Analysis by Fault Tree analysis and Monte Carlo simulation. *Proceedings of the joint 2012 International Conference on Probabilistic Safety Assessment and Management (PSAM 11) & European Safety and RELiability Conference (ESREL 2012)*, Helsinki, Finland, June 2012.
- [10] Cimellaro, GP. Reinhorn, AM. Bruneau, M. 2010. Framework for analytical quantification of disaster resilience. *Engineering Structures* 32: 3639-3649.
- [11] Gomez C, Sanchez-Silva M, Duenas-Orsorio L, Rosowsky D. Hierarchical infrastructure network representation methods for risk-based decision-making. *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance*. 2011. <http://dx.doi.org/10.1080/15732479.2010.546415>
- [12] Cheok MC, Parry GW, Sherry RR. Use of importance measures in risk informed applications. *Reliability Engineering and System Safety*. 1998; 60: 213-226.
- [13] Zio E. Computational methods for reliability and risk analysis, Chapter 2, *Series on Quality, Reliability and Engineering Statistics*, Vol 14, World Scientific Publishing Co. Pte. Ltd., 2009.
- [14] Kalos MH, Whitlock PA. *Monte Carlo methods*. Vol. 1, Basics. New York: Wiley; 1986. 186 p.
- [15] Zio E. Computational methods for reliability and risk analysis. *Series on Quality, Reliability and Engineering Statistics*, Vol 14. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2009. Chapter 2, Monte Carlo simulations for reliability and availability analysis; p. 59-69.
- [16] Zio E. *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*. London: Springer Series in Reliability Engineering. 2012.
- [17] *Seismic Probabilistic Risk Assessment Implementation Guide*, EPRI, Palo Alto, CA: 2003. TR-1002989.
- [18] *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Expert*. Main Report, Vol. 1. 1997, NUREG/CR-6372 UCRL-ID- 122160. Supported by U.S. Nuclear Regulatory Commission (NRC), the U.S. Department of Energy (DOE); and the Electric Power Research Institute (EPRI).
- [19] Sen TK. *Fundamentals of seismic loading and structures*. Singapore: John Wiley & Sons, Ltd; 2009. Chapter 7, Probabilistic Seismic Hazard Analysis; p. 181-218.

- [20] Kanamori H. The energy release in great earthquakes. *Journal of Geophysical Research*. 1977; 82(20): 2981–2987.
- [21] Kanamori H. Magnitude scale and quantification of earthquakes. In: SJ. Duda and K. Aki Editors. *Quantification of Earthquakes*. Tectonophysics, 1983; 93: 185-199.
- [22] Kramer SL. *Geotechnical Earthquake Engineering*, Prentice Hall, New Jersey. 1996.
- [23] Weatherill GA, Burton PW. The application of multiple random earthquake simulations to probabilistic seismic hazard assessment in the Aegean region. *First European Conference on Earthquake Engineering and Seismology*. Geneva, Switzerland. 2006.
- [24] Douglas J (Bureau de Recherches Géologiques et Minières). *Ground-motion prediction equations 1964-2010*. Berkeley (California): Pacific Earthquake Engineering Research Center: 2011. 455 p.
- [25] Ambraseys NN, Douglas J, SARMA SK, Smit PM. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering*. 2005; 3:1-53.
- [26] Gomez C, Sanchez-Silva M, Dueñas-Osorio L. Clustering methods for risk assessment of infrastructure network systems. *Applications of statistics and probability in civil engineering – Faber, Koler and Nishijima (eds)*. Taylor and Francis Group, London. 2011: 1389-1397.
- [27] Huang YN, Whittaker AS, Luco N. A probabilistic seismic risk assessment procedure for nuclear power plants: (I) Methodology, *Nuclear Engineering and Design*. 2011; 241: 3996–4003.
- [28] Varpasuo P. Seismic fragility analysis of selected heavy components in LNNP unit1 reactor building. *Transactions of the 17th International Conference on Structural Mechanics in Reactor Technology (SMiRT)*. Prague, Czech Republic, August 17-22, 2003.
- [29] Basu PC. Seismic fragility of nuclear installations. Atomic Energy Regulatory Board. Mumbai, India. 2008. Presentation: <http://civil.iisc.ernet.in/basu.pdf>
- [30] Lind M. An introduction to multilevel flow modeling. *Nuclear safety and simulation*. 2011; 2(1): 22-32.
- [31] Huzurbazar, A. V, Williams B. J. Flowgraph models for complex multistate system reliability. *Modern statistical and mathematical methods in reliability*. 2005; 10: 247-262.
- [32] Modarres M, Cheon SW. Function-centered modeling of engineering systems using the goal tree–success tree technique and functional primitives, *Reliability Engineering & System Safety*. 1999; 64(2): 181-200.

Paper IV

Goal Tree Success Tree – Dynamic Master Logic Diagram and Monte Carlo simulation for the safety and resilience assessment of a multistate system of systems

E. Ferrario and E. Zio

Engineering Structures 59 (2014) 411-433

Goal Tree Success Tree – Dynamic Master Logic Diagram and Monte Carlo simulation for the safety and resilience assessment of a multistate system of systems

E. Ferrario^a and E. Zio^{a,b}

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supélec, France

enrico.zio@ecp.fr, enrico.zio@supelec.fr

^bDepartment of Energy, Politecnico di Milano, Italy

enrico.zio@polimi.it

Abstract

We extend a system-of-systems framework previously proposed by the authors to evaluate the safety and physical resilience of a critical plant exposed to risk of external events. The extension is based on a multistate representation of the different degrees of damage of the individual components and the different degrees of safety of the critical plant. We resort to a hierarchical model representation by Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST – DMLD), adapting it to the framework of analysis proposed. We perform the quantitative evaluation of the model by Monte Carlo simulation. To the best of the author's knowledge this is the first time that a multistate framework of combined safety and resilience analysis relating the structural and functional behaviour of the components to the system function in a GTST – DMLD logic modelling of a system of systems is adopted in Seismic Probabilistic Risk Assessment. To illustrate the approach, we adopt a case study that considers the impacts produced by an earthquake and its aftershocks (the external events) on a nuclear power plant (the critical plant) embedded in the connected power and water distribution, and transportation networks which support its operation.

Keywords: Physical Resilience, Multistate Model, System of Systems, Goal Tree Success Tree – Dynamic Master Logic Diagram, Monte Carlo simulation, Seismic Probabilistic Risk Assessment.

1. INTRODUCTION

Resilience is the capacity of a system to survive to aggressions and shocks by changing its non-essential attributes and rebuilding itself [1]; it includes technical, organizational, social and economic facets [2]. In this work, we consider the “physical” resilience of a critical plant exposed to risk of an external event. We limit the analysis to the capacity of recovering from an external aggression or shock, using as representative quantity the recovery time, i.e., the period necessary to restore a desired level of functionality of a system after the shock [2]. For the resistance to the shock and the recovery from the shock, the critical plant is provided with internal emergency devices (internal barriers) to keep it in, or restore it to, a safe state when the main inputs devoted to this purpose fail. Since the internal emergency devices can fail too, we extend the boundaries of the study to the infrastructure systems (external supports) in which the plant is embedded, which also may or may not be left in the conditions to maintain the safety of the plant after the occurrence of a disruptive event. Supporting elements (e.g., roads for access to the sites struck by the disruptive external event) are also considered for the recovery of the failed components of the main inputs, internal barriers and external supports. We adopt the system-of-systems framework of analysis proposed by the authors in [3] and extend it to a multistate representation where different degrees of damage of the individual components are contemplated [2], [4], [5]. In particular, we consider an original multistate model of structural damage and functional performance at component level, that integrates into a multistate model of safety at system level for well-being analysis [6].

The modelling of the system of systems includes: i) the connections among the main inputs ii) the links among the internal barriers, iii) the dependencies among the external supports, iv) the interdependencies between the systems in i), ii), iii), and the relationships among systems in i), ii), iii) and the recovery supporting elements. We propose a hierarchical model representation by Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) [7]. This provides an efficient and clear description of the system-of-systems complexity through different hierarchical levels of system goals and functions, by the GT, and objects and parts, by the ST. The interrelationships are represented in a DMLD that translates into a dependency matrix and redefined logic gates, e.g., “AND” and “OR”, that assume a different meaning with respect to a binary state model, e.g., Fault Tree [7]. We extend the GTST-DMLD representation adapting it to the framework of analysis proposed. To the best of the author’s knowledge this is the first time that a multistate framework of combined safety and resilience analysis relating the structural and functional behaviour of the components to the system function in a GTST – DMLD logic modelling of a system of systems is adopted in Seismic Probabilistic Risk Assessment (SPRA). We use Monte Carlo simulation [8], [9], [10] for the probabilistic evaluation of such system of systems considering multiple levels of safety of the critical plant and physical resilience, measured in terms of the time needed to restore the different levels of safety.

To illustrate the approach, we adopt a simplified case study that considers a nuclear power plant (the critical plant) exposed to the risk of an earthquake and its subsequent aftershocks (the external events). The plant is provided with proper internal emergency devices (internal

barriers), and embedded in the connected power and water distribution (external supports), and transportation networks (recovery supporting elements) which support its operation and provide resilience to it.

The reminder of the paper is organized as follows. In Section 2, the multistate model for the safety assessment of a critical plant in a system-of-systems framework is presented; in Section 3, the Goal Tree Success Tree – Dynamic Master Logic Diagram and Monte Carlo simulation are described in relation to Seismic Probabilistic Risk Assessment and within the multistate system-of-systems framework; in Section 4, the case study and the results of the analysis are presented; in Section 5, conclusions are provided. Finally, in Appendix A, an exemplification of qualities, parts and GTST-DMLD within a system-of-systems framework is showed with respect to Sections 2 and 3; in Appendix B, the basic concepts of a Seismic Probabilistic Risk Assessment are introduced, to provide the reference elements needed for the case study; in Appendix C, details of the operative steps of the GTST-DMLD and Monte Carlo simulation for Seismic Probabilistic Risk Assessment are given.

2. MULTISTATE MODEL FOR THE SAFETY ASSESSMENT OF A CRITICAL PLANT WITHIN A SYSTEM-OF-SYSTEMS FRAMEWORK

In Section 2.1, the system-of-systems framework is illustrated with reference to three levels of safety and distinguishing its goal and functions, i.e., its qualities, and its objects, i.e., its parts; in Section 2.2, a multistate model for the system of systems is introduced.

2.1. System-of-systems framework: safety, qualities and parts

When due to an accident the main inputs to a critical plant stop, safety is assured by internal barriers which provide the inputs in the amount necessary for the safety conditions. These barriers are designed to withstand postulated accidents (design basis accidents) and include multiple, independent and redundant layers of defense to compensate for potential human and mechanical failures (defense in depth) [11]. As mentioned in the Introduction (Section 1), we adopt a system-of-systems view [3] extending the analysis to the external supports for emergency management actions and additional, redundant infrastructure systems to provide the safety-required inputs in case of failure of both the main inputs and the first (internal) barriers. In all generality, we consider also recovery supporting elements, as physical components (e.g., roads for access to the site) and organizational elements (e.g., technical competence of operators), that provide help in the recovery of the internal and external safety systems. On the basis of this system-of-systems framework, we can identify three levels of safety distinguishing the internal barriers (first level), the external supports (second level) and the recovery supporting elements (third level), as illustrated in Figure 1.

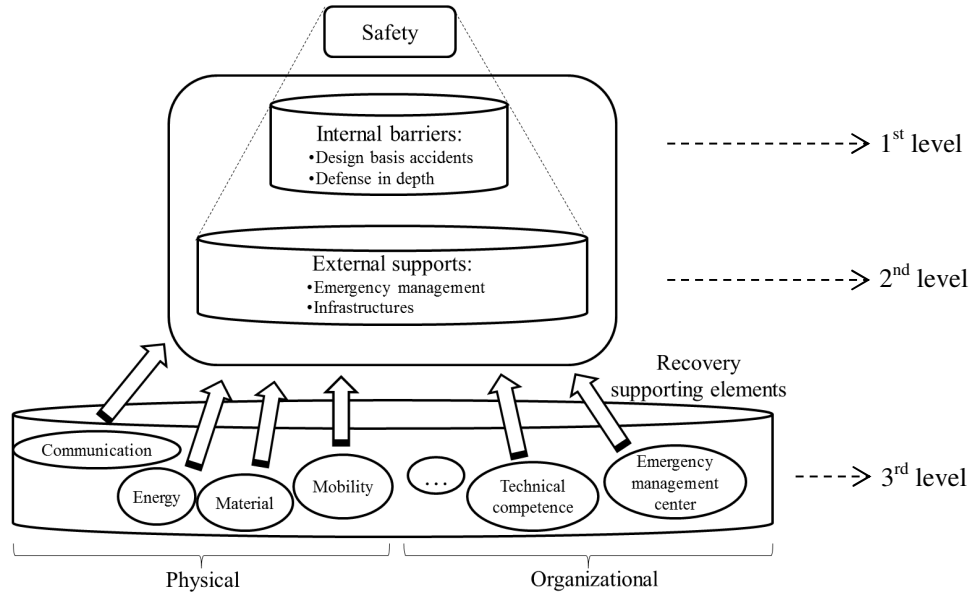


Figure 1: Safety levels of a system-of-systems framework considering a critical plant in emergency conditions. The first level (top) considers internal barriers; the second one (middle) extends to the external supports; the third one (bottom) accounts for the elements supporting the recovery.

In the present work, for the sake of simplicity, emergency management and organizational supporting elements are not considered. The concept of resilience is limited to the physical characteristics of the components and systems: then, we refer to physical resilience as the underlying concept. On the other hand, the Goal Tree Success Tree Dynamic Master Logic Diagram (GTST-DMLD) illustrated in Section 3 can accommodate elements of fuzzy logic theory to describe imprecisely known characteristics and logic relations of non-physical facets by linguistic fuzzy terms [7]. For example, specific inputs like the level of experience of the operators can have an impact on the degree of safety of the critical plant in emergency condition: these inputs could be described in the GTST-DMLD by including threshold values [7]. This kind of considerations will be subject of further development in the future research.

In the framework under analysis, we can distinguish between qualities and parts. The former are referred to the goals and functions, i.e., the objectives, of the system of systems; the latter are related to the objects, i.e., the physical elements, that interact with each other to attain the objectives.

In the following, we introduce a formal description of the qualities and parts, which can be organized in hierarchies, with respect to a critical plant H whose state corresponds to the state of its critical element, E .

The *qualities* are identified by the main goal F^* concerning the safety of H , i.e., E , that is attained by F_α , $\alpha = 1, \dots, N^*$, functions ordered in such a way that the first r directly achieve the goal F^* (i.e., they are principal functions) and the last $N^* - r$ support the first ones (i.e., they are auxiliary functions), as illustrated in Figure 2, on the left. The F_α , $\alpha = 1, \dots, N^*$, functions may be hierarchically divided into other functions that can be further decomposed into other ones until the required level of functional detail is reached. The last $N^* - r$

functions are represented in a parallel branch of the same hierarchy of F^* and they are connected to it by a dashed line to highlight their auxiliary role.

The *parts* are composed by N infrastructure systems $S^{(a)}$, $a = 1, \dots, A$, divided in: n^{MI} infrastructure systems of main inputs, n^{IB} internal barriers, n^{ES} external supports, n^{RS} recovery supporting elements (Figure 2, right). Each system $S^{(a)}$, $a = 1, \dots, A$, can be hierarchically decomposed into other systems that can be in turn divided into other ones until the desired level of detail of system components is reached. Some of the n^{MI} , n^{IB} and n^{ES} systems directly provide necessary supplies to the critical element E (i.e., they are principal systems), whereas some others among them are needed for the operation of the principal systems (i.e., they are auxiliary systems); to point out the different role of the last ones, they are connected to the corresponding principal systems by a dashed line (Figure 2, right), as for the functional hierarchy. The n^{RS} recovery supporting elements are considered apart from the other n^{MI} , n^{IB} and n^{ES} systems since they are involved in the recovery of system safety.

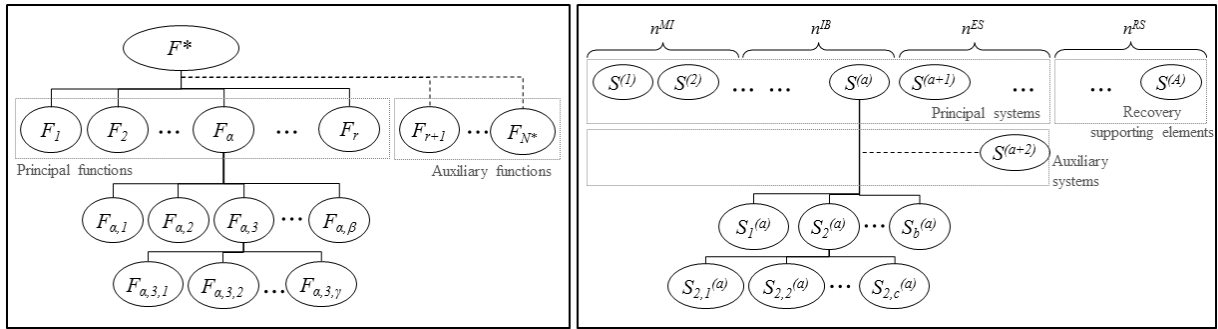


Figure 2: Scheme of the hierarchies of the qualities (left) and parts (right) of a system of systems. The auxiliary functions and parts are connected by a dashed line to the hierarchy branch that they support. The indices $\alpha, \beta, \gamma, a, b, c$ are used to indicate the systems/elements in the hierarchies; n^{MI} , n^{IB} , n^{ES} , n^{RS} refer to the number of main inputs, internal barriers, external supports and recovery supporting elements, respectively.

Notice that in a system-of-systems view only one main function (F^*) is analyzed, whereas more than one physical systems, involved in achieving that function, are considered ($S^{(a)}$, $a = 1, \dots, A$).

For illustration purpose, refer to Appendix A where an exemplification of qualities and parts is given.

2.2. System-of-systems framework: multistate model

The safety assessment of the critical plant is based on multistate modeling. In particular, at component level two aspects are described by the model: structural damage and functionality (Section 2.2.1); at system-of-systems level, only functionality, which is based on the structural and functional states of the components, is considered (Section 2.2.2).

2.2.1. Multistate model at component level: structural damage and functionality

Let us denote as η , $\eta = 1, \dots, L$, the generic component in the last level of the physical hierarchies of the systems, $S^{(a)}$, $a = 1, \dots, A$, where L is the total number of components that

are not further decomposed. A disruptive external event can affect both the physical structure and the functional performance of the generic component η , but not necessarily with a one-to-one correspondence. For example, a road can be affected at different levels of damage by an external event: from no damage to slight (few inches), moderate (several inches) or major (few feet) settlements of the ground. When the road is slightly damaged it can still perform its function (of connection) as in normal condition because the damage is negligible: then, the functional performance associated to the structural states “no damage” and “slightly damage” is the same. On the other hand, the correspondence between structural and functional states strongly depends on their definition and on the scope of the application, e.g., in a transportation planning the function of the road can be related to the traffic flow per hour and in this case the performance may be reduced even for slight settlements of the ground due to a decreasing speed of the vehicles, leading to a one-to-one correspondence between structural and functional states.

We define as g_i^η , $i = 1, 2, \dots, G$, and z_j^η , $j = 1, 2, \dots, Z$, the structural and functional states of the generic component η , respectively, where the indices i and j are ordered such that when $i, j = 1$, the component is fully damaged and cannot perform its function (worst condition); when $i = G$ and $j = Z$, the component shows no damage and can fully perform its function (best condition). Relations exist among the structural and functional states: a structural state corresponds to one functional state but one functional state can be associated to one or more structural states (Figure 3).

The evaluation of the safety of the critical plant is based on the functional state of the components that in turn depends on their structural state. The analysis of the functional state could be enough for evaluating the safety of the critical plant in the case of one-to-one correspondence between structural and functional states. On the contrary, considering more structural states than functional states allows us taking into account hidden (structural) criticalities that can suddenly turn the functionality of a component into a worse state, e.g., upon occurrence of aftershocks. In fact, a same functional state can be reached from different structural states, i.e., from different degrees of damage: even if functional performance is the same, a component with worse structural state is more fragile if exposed to other external events that can further degrade it structurally and at the same time cause a reduction of its functionality. For example, with respect to Figure 3, it can be seen that the functional state z_j^η , $j = 3$, can be reached when the component η is in the structural state g_i^η , $i = 4$, $i = 5$ or $i = 6$, but in the case $i = 4$ the component is weaker to withstand subsequent stresses than in the case $i = 6$, and therefore it is more inclined to pass into a lower structural state, i.e., if the structural state is lower than 4 (g_i^η , $i < 4$), the functionality will be lower than 3 (z_j^η , $j < 3$). With respect to the example of the road above, when the road is slightly damaged it is more exposed to aftershocks than when it is not damaged.

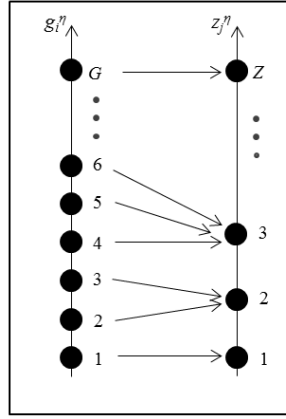


Figure 3: Relations between the structural, g_i^η , $i = 1, 2, \dots, G$, and functional z_j^η , $j = 1, 2, \dots, Z$, states for a component η .

In the case study exemplification of this work, we consider three structural and functional states, i.e., g_i^η and z_j^η with $i, j = 1, 2, 3$. They represent risk, marginal and healthy conditions, adopting the scheme of well-being analysis [6]. Denoting as $y^{\eta, \min}$ the lowest output value that it is requested by a component η to keep a safe state (it represents the risk threshold) and $y^{\eta, \text{opt}}$ the optimal output value that should be provided by the component η to keep a safe state with a safety margin, sm , ($sm = y^{\eta, \text{opt}} - y^{\eta, \min}$), we define:

1. Risk state:

- Structural (g_i^η , $i = 1$): the component η is strongly damaged by the external event.
- Functional (z_j^η , $j = 1$): the component η cannot fulfill its function; its output y^η is lower than the minimal requested $y^{\eta, \min}$, i.e., $y^\eta < y^{\eta, \min}$.

2. Marginal state:

- Structural (g_i^η , $i = 2$): the component η is slightly damaged by the external event.
- Functional (z_j^η , $j = 2$): the component η can fulfill its function, providing an output y^η that is lower than the optimal output $y^{\eta, \text{opt}}$, but higher than the minimal requested, i.e., $y^{\eta, \min} \leq y^\eta < y^{\eta, \text{opt}}$, the safety margin is not satisfied.

3. Healthy state:

- Structural (g_i^η , $i = 3$): the component is not damaged by the external event.
- Functional (z_j^η , $j = 3$): the component can fulfill its function, providing an output y^η that is equal or higher than the optimal output $y^{\eta, \text{opt}}$, i.e., $y^\eta \geq y^{\eta, \text{opt}}$.

The relations between structural and functional states depend on the scope of the application, as exemplified above, but also on the intrinsic characteristics of the components. The combinations considered for the case study of this work are illustrated in Figure 4 for a generic component η . The relations among three structural and functional states (Figure 4.a) are typical of elements of the water system since their functional performance is associated to their flow: a reduction of the water flow due to a structural damage means a reduction of their

functional performance, e.g., a leak in a pipe reduces the flow capacity. In the following, we refer to these elements as components of the first group. The combinations among three structural states and two functional states (Figure 4.b) occur when a component not damaged ($g_i^\eta, i = 3$) or slightly damaged ($g_i^\eta, i = 2$) can perform totally its function ($z_j^\eta, j = 3$), i.e., the structural damage of state 2 has no effects on the functional performance. The components characterized by these relations are referred to the second group and, for example, they are the road accesses, as shown above, and the elements of the power system, e.g., the power pole that can fulfill its function to carry the power line even if its structure presents some damage. Finally, binary components (Figure 4.c), included in the third group, present two structural and functional states: no degrees of damage are considered since also a slight damage lead a component to loose completely its functionality (e.g., in the case of a valve).

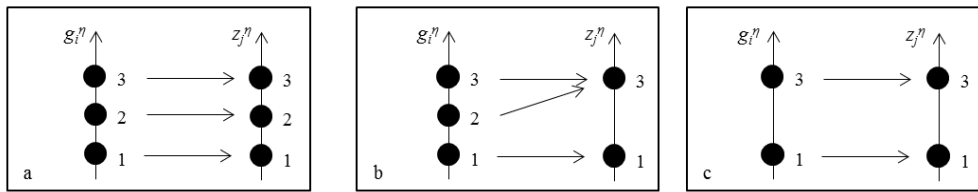


Figure 4: Three types of relations between the structural, $g_i^\eta, i = 1, 2, \dots, G$, and functional $z_j^\eta, j = 1, 2, \dots, Z$, states of a component η .

2.2.2. Multistate model at system-of-systems level: functionality

For the scope of the present application, we are not interested in the definition of an indicator of the structural state of the system of systems but rather in its functional performance, i.e., the degree of fulfillment of the goal function F^* (in this case, the degree of safety of the critical plant H). To obtain a functional state at system-of-system level, we combine the systems $S^{(a)}, a = 1, \dots, A$, into K alternative (or redundant) logic paths, $\xi_k^F, k = 1, \dots, K$, that attain the same function F^* , as illustrated in Figure 5 for four systems, $S^{(a)}, a = 1, \dots, 4$.

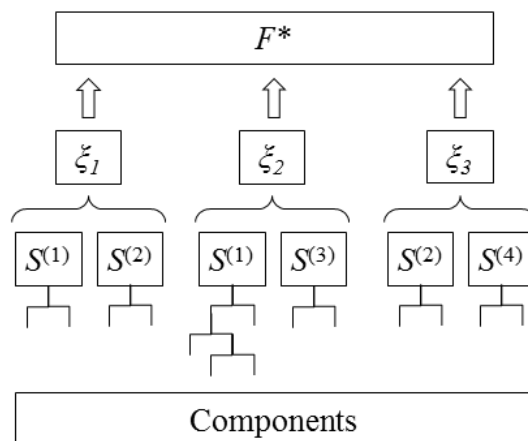


Figure 5: Exemplification of the combination of $S(a), a = 1, \dots, 4$, systems into 3 redundant logic paths $\xi_k^F, k = 1, \dots, 3$, that attain the same function F^* .

The functionality of the $S^{(a)}$, $a = 1, \dots, A$, systems is based on the functional performance and on the structural state of the components η , $\eta = 1, \dots, L$: then, we can identify a healthy, marginal and risk state for these systems on the basis of the states of their components. The functional state of the logic paths, ξ_k^F , $k = 1, \dots, K$, is in turn obtained from the states and the reciprocal relationships of the $S^{(a)}$, $a = 1, \dots, A$, systems. Finally, the functional performance at system-of-systems level is determined on the basis of i) how many and which logic paths, ξ_k^F , $k = 1, \dots, K$, are available and ii) their functional state. The evaluation of the function F^* is different case by case, depending on the characteristics of the system of systems and on the expert judgment. In the present work, we still consider three functional states, z_j^H , $j = 1, 2, 3$, i.e., risk, marginal and healthy, respectively, for the critical plant H . In all generality, we assume that both the healthy and marginal states assure the safety of the critical plant. While the first one can provide inputs to the critical plant by different available ξ_k^F , $k = 1, \dots, K$, alternative logic paths, i.e., safety margin is satisfied, the second one can assure inputs by only one of the redundant logic paths without possibility of replacing it in case of its accidental interruptions, i.e., a safety margin is not satisfied. Further details about the multistate model at system-of-systems level adopted in this work are reported in Section 4.2.

3. GOAL TREE SUCCESS TREE – DYNAMIC MASTER LOGIC DIAGRAM AND MONTE CARLO SIMULATION FOR SEISMIC PROBABILISTIC RISK ASSESSMENT WITHIN A MULTISTATE SYSTEM-OF-SYSTEMS FRAMEWORK

3.1. Goal Tree Success Tree - Dynamic Master Logic Diagram

The Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) is a goal-oriented method based on a hierarchical framework [7]. It gives a comprehensive knowledge of the system describing the complex physical systems in terms of functions (qualities), objects (parts) and their relationships (interactions). The first part is developed by the Goal Tree (GT), the second one by the Success Tree (ST) and the third one by the DMLD [7].

The GT identifies the hierarchy of the qualities of the system decomposing the objective of the analysis, i.e., the goal, into functions that are in turn divided into other functions and so on by answering the question “how” they can attain the parent function (looking from top to bottom of the hierarchy) and “why” the functions are needed (looking from bottom to top of the hierarchy). Two types of qualities, i.e., main and support functions, are considered on the basis of their role: the first ones are directly involved in achieving the goal, whereas, the second ones are needed to support and realize the main functions [12]. For example, the goal function of safely generating electric power in a nuclear power plant is attained by many functions as heat generation, heat transport, emergency heat transport, heat to mechanical energy transformation, mechanical to electrical energy transformation [13]. Each of these functions require the support of other functions, e.g., emergency heat transport may require internal cooling [13] or a pump whose function is to “provide pressure” require the support functions “provide ac power”, “cooling and lubrication”, “activation and control” [13].

The ST represents the hierarchy of the objects of the system from the whole system to the parts necessary to attain the last levels of the GT. This hierarchy is built identifying the elements that are “part of” the parent objects. As for the GT, two types of objects are distinguished: main and support objects. The first ones are directly needed to achieve the main functions, whereas the second ones are needed for the operation of the main objects [12]. For example, generating power plants, electric power transmission and distribution networks are the support objects to provide ac power to a pump.

The DMLD is an extension of the Master Logic Diagram (MLD) [7] to model the dynamic behavior of a physical system. It identifies the interactions between parts, functions and parts and functions, in the form of a dependency matrix and it adds the dynamic aspect by introducing time-dependent fuzzy logic [7].

Further details are not given here for brevity sake: the interested reader is referred to the cited literature [12], [7]. In the next Section, the adaption of the GTST-DMLD for a multistate system-of-systems framework is illustrated.

3.2. Goal Tree Success Tree - Dynamic Master Logic Diagram of a system of systems

We adapt the GTST-DMLD presented in Section 3.1 to a proper representation of a system of systems. Figure 6 shows a conceptual scheme of GTST-DMLD for a system of systems.

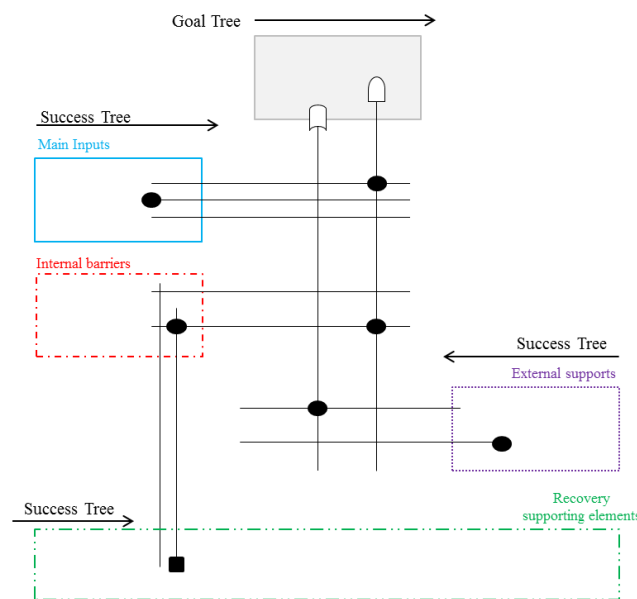


Figure 6: Scheme of GTST-DMLD for a system of systems.

The Goal Tree (GT) is located at the top; the Success Tree (ST), below the GT, is divided into three different parts to put in evidence the different role and importance of the physical elements with respect to the safety levels introduced in Section 2.1. The main inputs and the internal barriers are placed on the top-left, the external supports on the middle-right and the recovery supporting elements on the bottom.

We call the “main” and “supporting” functions/parts of the original GTST-DMLD representations as “principal” and “auxiliary” functions/parts, respectively, in order to avoid

confusion with the main inputs, the external supports and the recovery supporting elements of the system-of-systems framework.

The relationships among elements and functions are illustrated by the MLD. In particular, the connections among components of i) the main inputs, ii) the internal barriers, iii) the external supports are shown; the interdependencies between the systems i), ii), iii) are depicted; the links of the recovery supporting elements with the systems i), ii), iii) are indicated; the connections between the systems i), ii), iii) and the functions of the Goal Tree are given. Two types of dependencies have been taken into account: direct and support dependencies. The first ones, identified by a dot in the representation and called in the following “dot-dependencies”, express the need to have the element on the bottom in operation to achieve (with respect to a function) or to let working (with respect to an object) the element on the top. The support dependencies, depicted by a square and called hereafter “square-dependencies”, mean that the element on the bottom is needed for the recovery of the element on the top: its failure does not cause the failure of the corresponding elements, but it increases the recovery time of the connected element in the case that this fails too. It acts like a delay in the repairing of the connected components. Thus, the square-dependencies are “time dependent”: when a component does not need recovery they can be neglected, whereas, in the opposite case, they become fundamental until the complete restoration of the component; at this point, they can be neglected again. They are key elements of the model for the evolution in time of the recovery process and they can modify (increase) the total recovery time of the component that needs to be restored.

The dynamic aspect, consisting in the functional multistate of the components, is represented by the logic gates “AND” and “OR” that assume the same meaning as in [7] to evaluate the state of the connected components and functions from the bottom to the top of the diagram: the minimum and the maximum values of inputs are the output values in case of “AND” and “OR” gates, respectively. In this state analysis only the dot-dependencies are considered. In the present work the inputs are discrete states (see Section 2.2) but are not described by fuzzy intervals as in [7].

On the contrary, in the evaluation of the physical resilience both the dot- and square-dependencies are included and the logic gates “AND” and “OR” have an opposite meaning with respect to the state evaluation. In fact, the output values of the “OR” and “AND” gates are the minimum and the maximum values of the inputs, respectively. In this case, the inputs are the recovery time values. For example, refer to Figure 7 where two systems $S^{(a)}$, $a = 1, 2$, contribute to the realization of the function F^* (dot-dependencies) and two other systems $S^{(a)}$, $a = 3, 4$, are relevant only to allow the recovery of the system $S^{(a)}$, $a = 2$, (square-dependencies). Assuming that $S^{(1)}$ and $S^{(4)}$ are in functional state 3, $z_j^{S(1)}$ and $z_j^{S(4)}$, $j = 3$, with associated recovery time ($RT_{S(1)}$ and $RT_{S(4)}$) equal to 0, and $S^{(2)}$ and $S^{(3)}$ are in state 1, $z_j^{S(2)}$ and $z_j^{S(3)}$, $j = 1$, with associated recovery times ($RT_{S(2)}$ and $RT_{S(3)}$) equal to 2 and 5, respectively, the function F^* is in state 1, $z_j^{F^*}$, $j = 1$, since the “AND” gate (G1) means “minimum values between $z_j^{S(1)}$ and $z_j^{S(2)}$ ”. The time needed to realize the function F^* is 7 ($RT_{F^*} = 7$) since the “AND” gate (G1) means “maximum values between $RT_{S(1)}$ and $RT_{S(2)}$ ”, where the total time

needed to recover $S^{(2)}$ depends on the time to recover $S^{(2)}$ itself and the maximum value (“AND” gate G2) between $RT_{S(3)}$ and $RT_{S(4)}$. Replacing the “AND” gate G2 with an “OR” gate, the total time needed to recover $S^{(2)}$ is 2, since the minimum value between $RT_{S(3)}$ and $RT_{S(4)}$ is zero. Replacing both the “AND” gates, G1 and G2, with two “OR” gates, the function F^* is in state 3, $z_j^{F^*}$, $j = 3$, thus, it is not necessary to recover it ($RT_{F^*} = 0$).

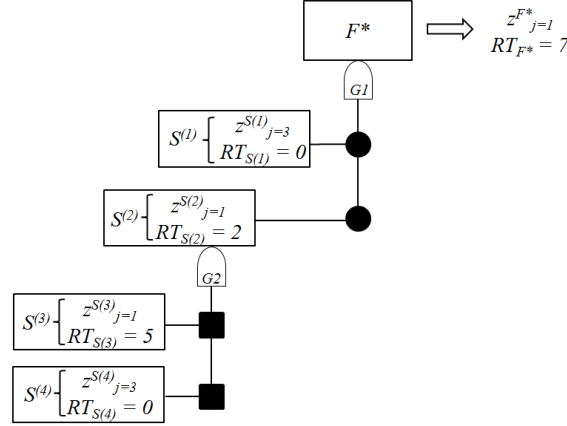


Figure 7: Example of the use of the “AND” logic gate together with the dot- and square- dependencies for computing the state and the recovery time of the function F^* .

In Appendix A, an example of GTST-DMLD is reported.

3.3. Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a system-of-systems framework

Within the system-of-systems analysis framework here purported, in the case study of the next Section 4 we wish to evaluate the safety of the critical plant H (a nuclear power plant) exposed to the risk from earthquakes and aftershocks occurrence (see Appendix B), accounting for the structural and functional responses of the systems inside and outside the plant, i.e., main inputs, internal barriers, external supports and recovery supporting elements, through the analysis of the underlying dependency structure. In addition, we wish to determine the physical resilience of the system of systems, evaluated in terms of the time of recovery of safety states 2 and 3 (marginal and healthy, respectively) of the critical plant. To do this, we adopt the GTST-DMLD representation of the system of systems and Monte Carlo (MC) simulation for the quantitative SPRA evaluation [14]. The simulation procedure is illustrated in Appendix C.

4. CASE STUDY

We recall the case study of [3] concerning the safety of a nuclear power plant (the critical plant), in response to an earthquake (the external hazardous event). The problem is analyzed in a system-of-systems framework, distinguishing main inputs, internal barriers, external supports and recovery supporting elements. We adopt a multistate model to identify different degrees of component damage and, consequently, different degrees of system safety. In

particular, at the system level we consider three states of the nuclear power plant of which two correspond to safe conditions (marginal and healthy, see Section 2.2). Safe condition means that the nuclear power plant does not cause health problems and environmental damages, i.e., it does not release radioactive material to the environment. To maintain these conditions it must be provided with energy and water flow inputs to absorb the heat that it generates.

We analyze also the physical resilience of the system of systems, in terms of the time necessary to recover the safe states (marginal and healthy) of the plant including the occurrence of aftershocks that can further degrade the system of systems.

When an earthquake occurs, the critical plant may not receive the input necessary to be kept in, or restored to, a safe state due to the direct impact on its emergency devices and to the damage to the interconnected infrastructures. Two quantities are used to characterize the loss of functionality of the various components of the system of systems embedding the critical plant, upon the occurrence of a damaging external event:

- from the safety viewpoint, the probability that the critical plant remains in marginal and healthy states;
- from the physical resilience viewpoint, the time needed to recover the marginal and healthy states of the critical plant facing the occurrence of aftershocks.

Both quantities are here computed for an earthquake of magnitude equal to 5.5 on the moment magnitude scale.

In Section 4.1, the description of the system studied is given under a number of assumptions which simplify the problem to the level needed to convey the key aspects of the conceptual system-of-systems framework, while maintaining generality. In Section 4.2, the Goal Tree Success Tree – Dynamic Master Logic Diagram representation of the system-of-systems considered in the case study is given. In Section 4.3, we provide the results of the evaluation of the two quantities of interest above mentioned.

4.1. Description of the system of systems

The critical plant, i.e., the nuclear power plant (NPP), is composed by a Main Feedwater (MFW) system that provides coolant useful to absorb the heat generated and four internal barriers: High Pressure Coolant Injection (HPCI) and Low Pressure Coolant Injection (LPCI) systems that provide water to cool the reactor, an automatic depressurization system (ADS) that reduces the pressure in the reactor vessel and a diesel generator (DG) that can provide the LPCI system with power.

The MFW system is formed by a condenser where the unused steam coming from a turbine is condensed into water that is pumped to the reactor vessel by the feedwater pump (FWP) and pipes (Pi1 and Pi2). In case of accident damaging the MFW system function, the HPCI and LPCI systems need to provide the necessary function. Both systems are composed by a condensate storage tank (CST1 and CST2, respectively), a pump (HPP and LPP, respectively) and pipes (Pi3, Pi4 and Pi5, Pi6, respectively). To operate, the LPCI system needs the automatic depressurization system (ADS) to reduce the pressure inside the vessel. Apart from

the pump of the HPCI system that is a turbine-driven pump, the pumps of the MFW and LPCI systems need electrical power to work. This is usually provided by the offsite power and in case of its loss, the emergency diesel generator can be activated to supply the LPP.

The external supports of the critical plant are the offsite power system (EE) and an external water (EW) system. The first one is composed by a generation station (GS) that produces the electrical energy, a substation (S) that transforms the voltage from high to low, power lines and poles (Po1 and Po2) to support them. The second one is formed by the river, i.e., the source of water, a pump (RP) that receives electrical power from the offsite power system and pipes (Pi7 and Pi8) that carry the water.

The recovery supporting elements are the road accesses to the components of the system of systems. The state of the roads is important for access of materials and operators that are needed to restore the components required for the safe state of the critical plant.

Actually, in view of the methodological character of this work, for the sake of simplicity, power lines are not here considered and the assumption is made that the river is not perturbed by the earthquake so that it is a source of water always available.

In Figure 8, the physical representation of the system of systems is reported referring to a spatial plane (x, y) with origin in the river; one type of soil, i.e., soft soil, has been considered.

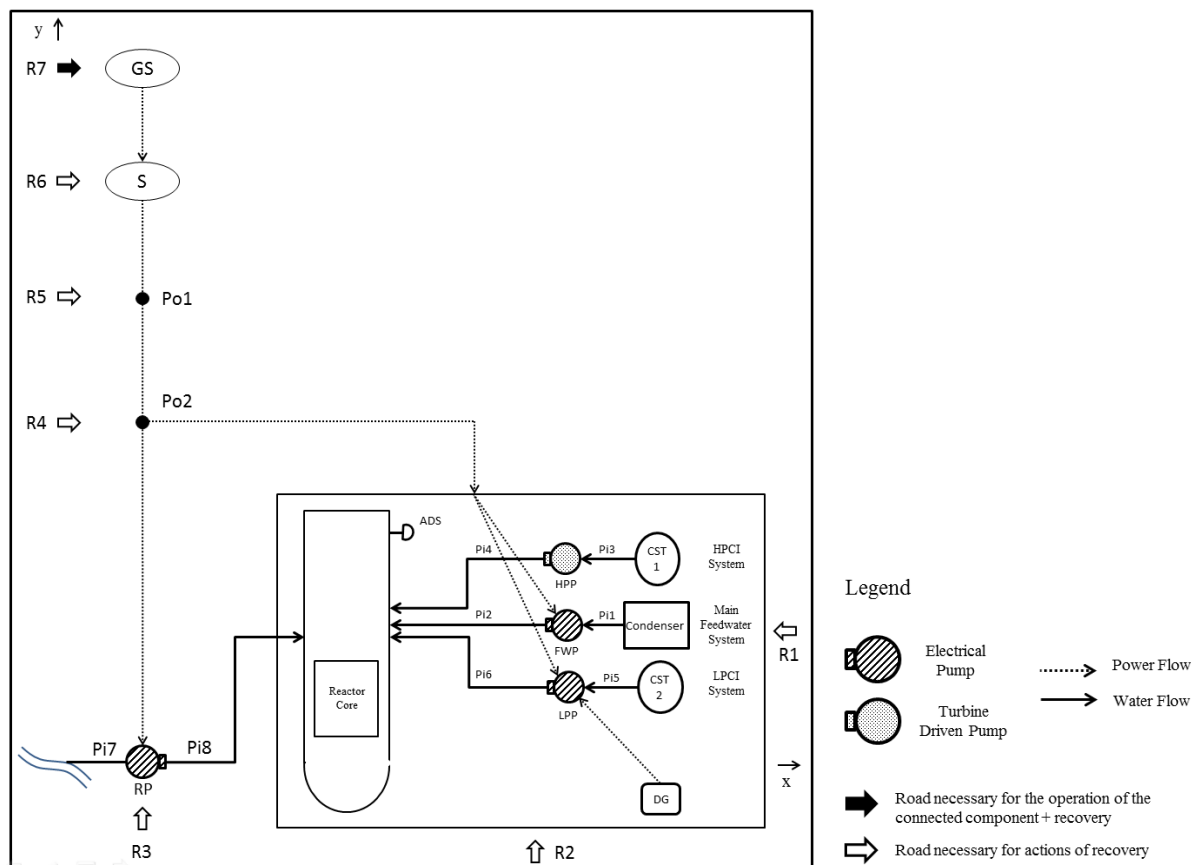


Figure 8: Physical representation of the system. GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, RP: River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator, R: Road access.

Only the road access connected to the generation station, R7 in Figure 8, has an impact on the state of the system of systems because it contributes to the running of the generation station, carrying materials and operators. On the contrary, the other road accesses have no direct impact on the state of the system of systems since they are used only to repair the elements that enter in faulty and marginal states. Therefore, their contribution is not of interest for the evaluation of the safety of the critical plant, but they are relevant for the analysis of the physical resilience of the system of systems. Given the different role of the road access R7 we will consider it, in the following, as an auxiliary element of the offsite power system.

Figure 9 represents the spatial localization of the system shown in Figure 8 with reference to the reciprocal position of all the components (Figure 9, left) and to the position of the system with respect to the considered earthquake epicenter A(70, 70) (Figure 9, right). The distances on the axes are expressed in kilometers.

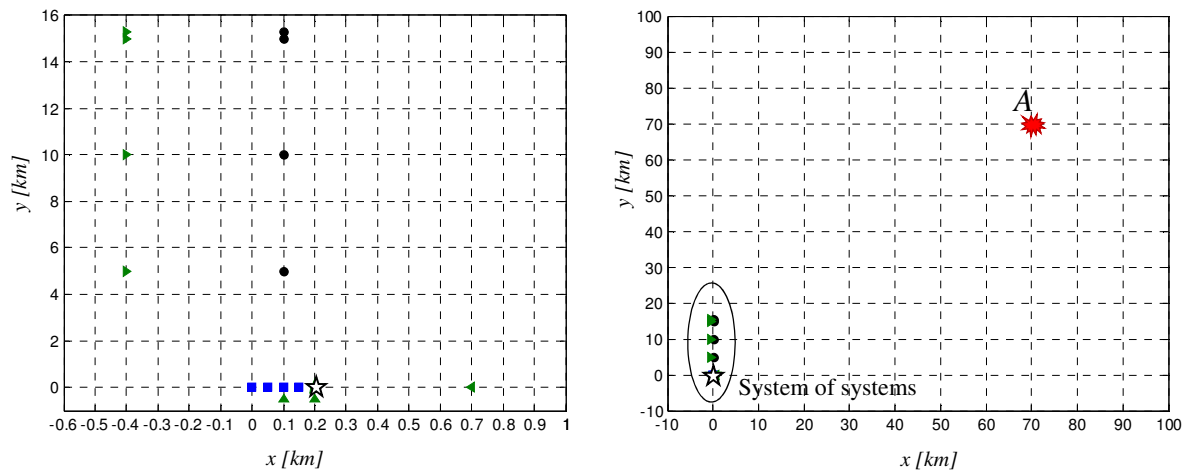


Figure 9: Left: spatial localization of the nuclear power plant (star) with respect to the components of the electric power system (circle, from top to bottom: Generation Station, Substation, Pole 1, Pole 2), water system (square, from left to right: River, Pipe 7, RP, Pipe 8) and road transportation (triangle, from top to bottom and from left to right: R7, R6, R5, R4, R3, R2, R1). Right: spatial localization of the system of systems with respect to the earthquake's epicenter A(70, 70).

Figure 10 shows the graph of the system of systems with respect to the safety levels of Section 2.1. The arrows are directed from one element to another one which depends on it.

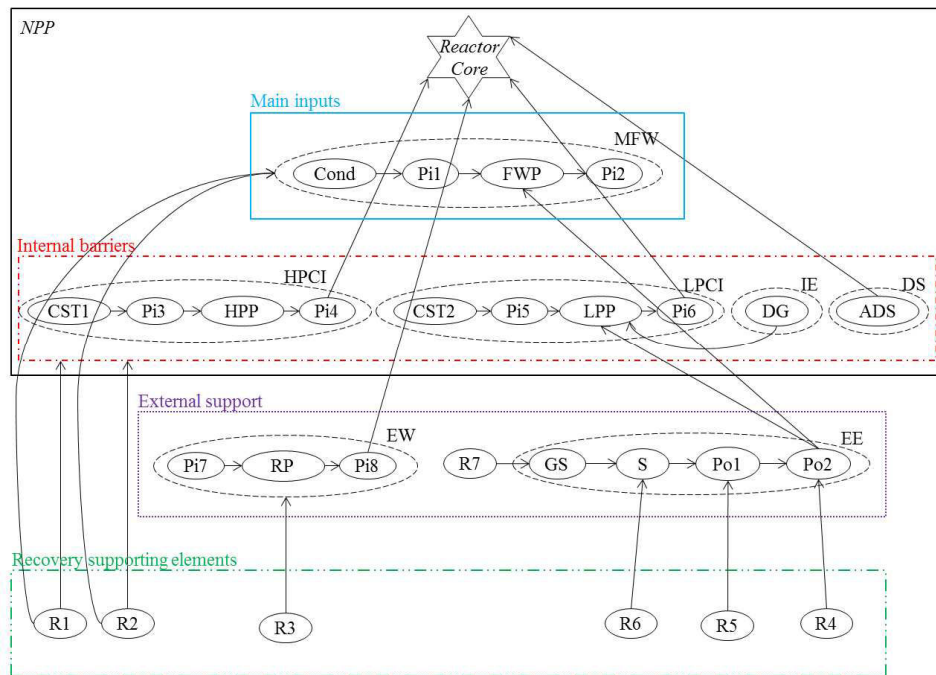


Figure 10: Graph of the system of systems. MFW: Main Feedwater System; HPCI: High Pressure Coolant Injection System; LPCI: Low Pressure Coolant Injection System; IE: Internal Energy System; DS: Depressurization System; EW: External Water System; EE: Offsite power system; R: Road access; GS: Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, Cond: Condenser; RP: River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator.

4.1.1. Resistance of the components in terms of fragility

We assume that all the components are in a structural state 3 (healthy) when the earthquake occurs. After that, they can remain in the state 3, turn into a state 2 (marginal) or directly pass into a state 1 (risk). If they enter in a state 2, they can degrade to a state 1 as a consequence of subsequent aftershocks.

For illustration purposes, Table 1 reports the fragility parameters A_m , β_r and β_u (see Appendix B.1), adopted in this analysis with reference to the two degrees of damage considered (marginal and risk). In the first three columns, the fragility parameters to enter in a risk state given that the component was in a healthy state are reported; these values are the same adopted by the authors in [3], adding the values for the automatic depressurization system that was not considered in the previous work. The fragility parameters to enter in a marginal state given that the component was in a healthy state are reported in the three columns, in the middle. These values are obtained decreasing arbitrarily the median acceleration capacity, A_m , by 40%, assuming that it is easier to enter into a marginal state than in a risk state. In the last three columns, the fragility parameters to enter into a risk state given that the component was in a marginal state are illustrated. These values are identified by decreasing the median acceleration capacity, A_m , of the healthy state by 55%, since a component in a marginal state is more prone to pass into a risk state than a component in a healthy state. In Figure 11, the fragility curves obtained by the parameters of Table 1 are depicted: the fragility curves of

exceeding a risk threshold given that the initial states were healthy and marginal are illustrated in dashed and solid lines, respectively, the fragility curve of exceeding a marginal threshold given that the initial state was healthy is represented in dotted line.

Table 1: Fragility parameters used in the present work with respect to the transitions healthy-risk, healthy-marginal and marginal-risk.

	Healthy \rightarrow Risk			Healthy \rightarrow Marginal			Marginal \rightarrow Risk		
	A_m	β_r	β_u	A_m	β_r	β_u	A_m	β_r	β_u
Generation station	0.70	0.30	0.10	0.42	0.30	0.10	0.32	0.30	0.10
Substation	0.90	0.40	0.30	0.54	0.40	0.30	0.41	0.40	0.30
Power Pole	0.80	0.20	0.20	0.48	0.20	0.20	0.36	0.20	0.20
Diesel Generator	0.70	0.40	0.20	0.42	0.40	0.20	0.32	0.40	0.20
Pipe	1.88	0.43	0.48	1.13	0.43	0.48	0.85	0.43	0.48
Pump	0.20	0.20	0.30	0.12	0.20	0.30	0.09	0.20	0.30
Condensate storage tank / Condenser	0.20	0.10	0.10	0.12	0.10	0.10	0.09	0.10	0.10
Automatic depressurization system	1.5	0.3	0.3	-	-	-	-	-	-
Road	0.30	0.30	0.20	0.18	0.30	0.20	0.14	0.30	0.20

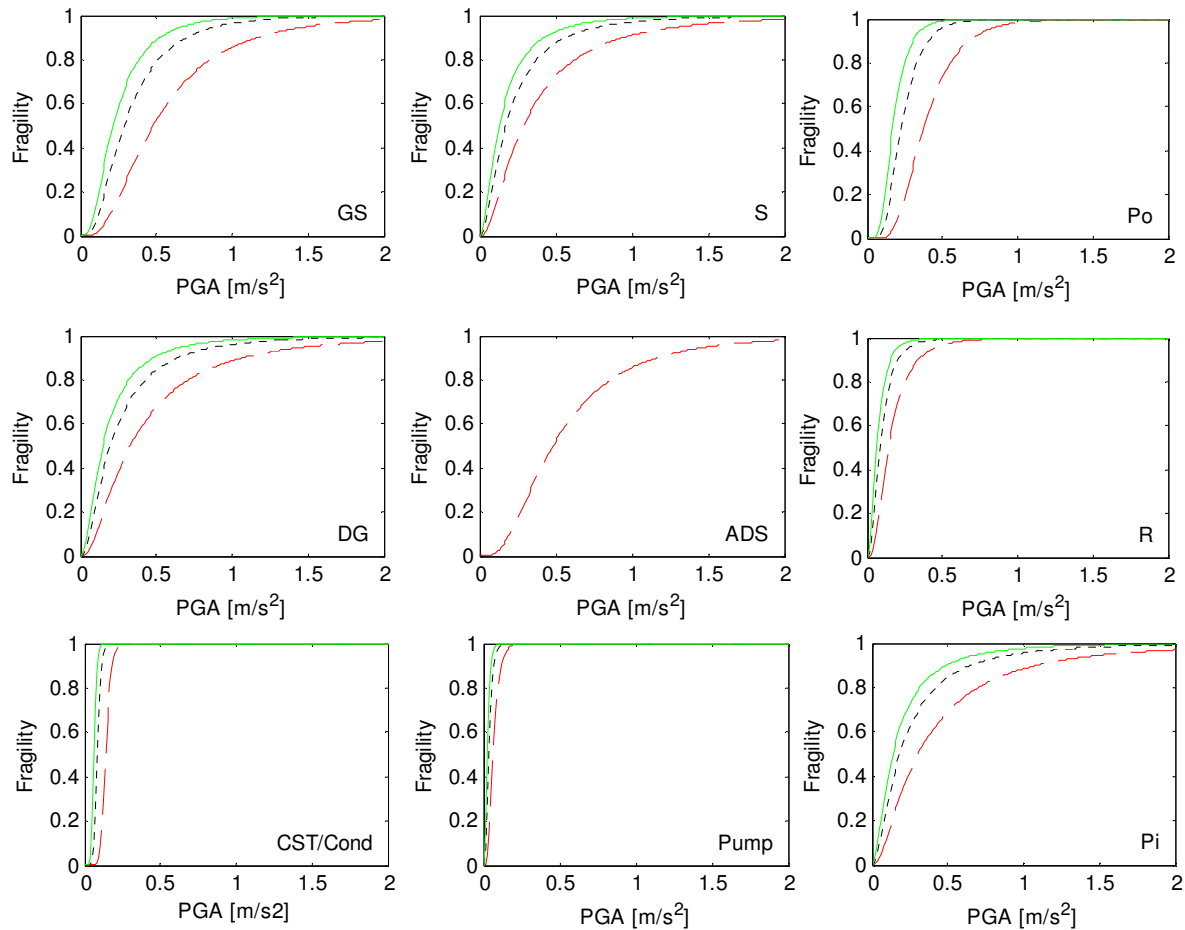


Figure 11: Fragility curves as a function of the peak ground acceleration (PGA) [m/s^2] for the following components: Generation Station (GS), Substation (S), Power Pole (Po), Diesel Generator (DG), Automatic Depressurization System (ADS), Road Access (R), Condensate Storage Tank (CST), Condenser (Cond), Pump, Pipe (Pi). The fragility curves of exceeding a risk threshold given that the initial states were healthy and marginal are illustrated in dashed and solid lines, respectively, the fragility curve of exceeding a marginal threshold given that the initial state was healthy is represented in dotted line.

Notice that the automatic depressurization system presents fragility parameters only to enter into a risk state from a healthy state, since we describe it with a binary state model: with respect to the taxonomy of combinations of structural and functional states introduced in Section 2.2.1, it belongs to the third group of components.

On the contrary, we consider the pumps and pipes in the first group (three structural and three functional states) since their functional performance is associated to the water flow. For the sake of simplicity, the condensate storage tank and the condenser are included in the second group even if they concern the water flow. The elements of the power systems and the road access belong to the second group too, since a slight damage in their parts does not affect their functionality: a power pole can or cannot support the power lines, a generation station can or cannot produce the quantity of energy requested, a road can or cannot provide access to the connected component.

Table 2 reports examples of structural damage to show the meaning of a specific component being in a healthy, marginal or risk states. These values have been extracted from [15] where five levels of structural damage (none, slight/minor, moderate, extensive, complete) are identified for some components of the power, water and transportation systems. For example, for a substation a *slight damage* is defined as the failure of 5% of the disconnected switches, or the failure of 5% of the circuit breakers, or by the building being in minor damage state; a *moderate damage* is defined as the failure of 40% of the disconnected switches, or the failure of 40% of the circuit breakers, or the failure of 40% of the current transformers, or by the building being in moderate damage state; an *extensive damage* is defined as the failure of 70% of the disconnected switches, or the failure of 70% of the circuit breakers, or the failure of 70% of the current transformers, or by the building being in extensive damage state; a *complete damage* is defined as the failure of all disconnected switches, or the failure of all the circuit breakers, or the failure of all the current transformers, or by the building being in complete damage state [15]. In the Table, the values are grouped into the three structural states: healthy (i.e., none damage) marginal (i.e., slight/minor and moderate) and risk (i.e., extensive and complete). The structural state for the pipes is taken from [16] that distinguish between small ($< 2\%$), intermediate ($2\% \div 10\%$) and large breaks ($> 10\%$). Here it is considered that the marginal state includes the small and intermediate breaks.

In Table 2, also the functional performance of a component that is in a specific state is reported. Values of flow are identified for the components of the group 1; whereas percentages of 100% or 0% of functionality are associated with the components of the groups 1 and 2 that have binary functional states. To identify the flow values, we consider that in shutdown conditions the flow rate to cool the reactor is between 4625 gpm [16] and 5010 gpm [17]. Therefore, a component of a water system of the group 3 is in a healthy functional state if it can provide a quantity of water equal or higher than 5010 gpm, it is in a marginal functional state if it can provide a quantity in the interval 4625 gpm - 5010 gpm, otherwise it is in a risk functional state.

Note that, in this work we have not considered interdependence between structural and functional thresholds since we have assumed that the functionality depends on the structural

state. A further study will be performed to identify the correspondence between structural and functional state quantitatively, or to determine fragility curves that are based on multiple limit states parameters and can include both the aspects of structural safety and functionality, as illustrated in [18].

Table 2: Physical meaning of structural damage and functional performance with respect to the healthy, marginal and risk states of the components of the case study.

	State	Structural damage	Functional performance
Pumps (FWP, HPP, LPP, RP)	Healthy	0%	5010 [gpm]
	Marginal	-	4625 ÷ 5010 [gpm]
	Risk	-	< 4625 [gpm]
Pipes (Pi1, ..., Pi8)	Healthy	0%	5010 [gpm]
	Marginal	0 ÷ 10% (break size)	4625 ÷ 5010 [gpm]
	Risk	> 10% (break size)	< 4625 [gpm]
Condensate Storage Tank (CST1 and CST2) / Condenser	Healthy	0%	100%
	Marginal	Damage without loss of its content or with minor loss of content	
	Risk	Major damage with loss of its contents	0%
Automatic Depressuriz. System (ADS)	Healthy	0%	100%
	Risk	> 0%	0%
Generation Station (GS)	Healthy	0%	100%
	Marginal	Turbine tripping, building in minor/moderate damage state...	
	Risk	Considerable damage to motor driven pumps or building in extensive damage state,...	0%
Substation (S)	Healthy	0%	100%
	Marginal	0 ÷ 40% failure of the disconnected switches, or of the circuit breakers, or of the current transformers...	
	Risk	> 40% failure	0%
Pole (Po1 and Po2)	Healthy	0%	100%
	Marginal	0 ÷ 12% failure of distribution circuits	
	Risk	> 12% failure	0%
Diesel Generator (DG)	Healthy	0%	100%
	Marginal	-	
	Risk	-	0%
Roads (R1, ..., R7)	Healthy	0%	100%
	Marginal	Slight/moderate settlement (few/several inches) or offset of the ground	
	Risk	Major settlement of the ground (few feet)	0%

4.1.2. Physical resilience in terms of time of recovery

The physical resilience of the system of systems is quantified in terms of the time needed to recover the healthy state of the critical plant starting from a risk and marginal state, and its marginal state starting from a risk state. To compute this, the evolution in time of the system of systems is included in the SPRA framework.

As illustrated in the procedure of Appendix C, the recovery time of the nuclear power plant is computed starting from the recovery time of the individual components and analyzing the dependency structure identified by the GTST-DMLD.

To account for the uncertainty in the duration of the recovery, lognormal distributions have been associated to the recovery time of the individual components. Table 3 shows the means and the error factors used in this study to recover the safety i) from risk to healthy state (first two columns), ii) from marginal to healthy state (two columns in the middle) and iii) from risk to marginal state (last two columns). The values of recovery from risk to healthy state are the same used by the authors in [19] and they are based on the following consideration. The time to recover a component depends on its size, its location, the type of damage and easiness to locate the failure. It is assumed that the components inside the nuclear power plant need more time for the recovery than the components outside. In particular, this happens when it is necessary to replace part of the component or the entire component given its huge dimensions and the difficulty to operate inside the plant. For this reason, we have assumed that the mean of the time needed to recover the pump inside the nuclear power plant is larger than that needed for the pump outside. The large mean value of the time to recover the condensate storage tanks and condenser is due to their size, location inside the plant and difficulty in restoration. The time to physically repair a pipe could be very short (even few hours), but we have assumed a mean value equal to 4 days to account for the potential difficulty in locating the break. The diesel generator has a time of repair with a high uncertainty (error factor equal to 5), because it may vary significantly depending on the type of damage. The components with lowest mean value of the recovery time are the power pole, the road, the generation station and the substation that are outside the plant; the latter are affected by large uncertainty (error factors of 5 and 10, respectively), because their recovery depends on the intensity of the damage, e.g., a generation station can be slightly perturbed by the earthquake and its repairing can last few hours but it can also be destroyed, and in this case the time to build it again is obviously much higher. Finally, also the automatic depressurization system, even if inside the plant, presents a short recovery time, because we assume that it is easy to replace it with another one.

The mean values of recovery for the cases ii) and iii) above are identified by considering that the time to recover a component from risk to marginal state is longer than that from marginal to healthy state and their sum is equal to the direct recovery from risk to healthy state. Thus, we define the mean values for the cases ii) and iii) as the 30% and 70%, respectively, of the mean value from risk to healthy state.

Table 3: Mean, μ , and Error Factor, EF , of the recovery time lognormal distribution used in the present work with respect to the transitions risk-healthy, marginal-healthy, risk-marginal.

	Risk \rightarrow Healthy		Marginal \rightarrow Healthy		Risk \rightarrow Marginal	
	μ [days]	EF	μ [days]	EF	μ [days]	EF
Generation station	1	10	0.3	10	0.7	10
Substation	1	5	0.3	5	0.7	5
Power Pole	1.5	3	0.45	3	1.05	3
Diesel Generator	30	5	9	5	21	5
Pipe	4	3	1.2	3	2.8	3
Pump (inside the plant)	75	3	22.5	3	52.5	3
Pump (outside the plant)	5	3	1.5	3	3.5	3
Condensate storage tank / Condenser	75	3	22.5	3	52.5	3
Automatic depressurization system	1	3	-	-	-	-
Road	2	3	0.6	3	1.4	3

4.2. GTST-DMLD and physical resilience of the system of systems

Figure 12 shows the GTST-DMLD of the system of systems depicted following the scheme of Figure 6 and on the basis of the graph of Figure 10. The goal function is the safety of the nuclear power plant assured by water inputs (i.e., the principal function) that can be provided by four different alternative paths (ζ_k^{Water} , $k = 1, \dots, 4$): the main feedwater system (ζ_1^{Water}), the high pressure coolant injection system (ζ_2^{Water}), the combination of low pressure coolant injection and depressurization systems (ζ_3^{Water}), the external water system (ζ_4^{Water}). The power coming from outside (Ext) or inside (Int) the plant is an auxiliary function to support the operation of most of the water systems. For the explanation of the logic gates, of dot- and square- dependencies, see Section 3.2.

It can be seen that the components among the systems MFW, HPCI, LPCI, EW, EE are connected in series for the presence of the “AND” gates. The systems IE, DS, R1, R2, R3, R4, R5, R6 and R7 are composed by only one component. Finally, the systems EE and IE are in parallel with respect to the LPCI system, as the roads R1 and R2 with reference to the components inside the nuclear power plant (“OR” gates).

Following the rules of the “AND” and “OR” gates, it is possible to compute the state and the mean time to recover the paths ζ_k^{Water} , $k = 1, \dots, 4$, and, then, the safety and the recovery of the nuclear power plant. For example, the mean time to recover ζ_k^{Water} , $k = 1$, is the maximum between the mean times to recover the MFW system and the EE system:

$$E[\zeta_1^{Water}] = \max(E[RT_{MFW}], E[RT_{EE}]),$$

where $E[RT_{MFW}]$ is the maximum expected value between the components of the MFW system and the minimum expected value of the two road accesses connected to them, and $E[RT_{EE}]$ is the maximum expected value between the components of the EE system and their road accesses:

$$E[RT_{MFW}] = \max(E[RT_{Pi2}], E[RT_{FWP}], E[RT_{Pi1}], E[RT_{Cond}], \min(E[RT_{R1}], E[RT_{R2}]))$$

$$E[RT_{EE}] = \max(E[RT_{Po2}], E[RT_{Po1}], E[RT_S], E[RT_{GS}], E[RT_{R7}], E[RT_{R6}], E[RT_{R5}], E[RT_{R4}])$$

In Table 4, for illustration purposes, the expected values of the time needed to recover the paths ξ_k^{Water} , $k = 1, \dots, 4$, into a marginal and healthy state are reported assuming that all the components are in state 1 (first two columns) and, then, that all of them are in state 2.

Table 4: Expected values of recovery time to turn the nuclear power plant into a healthy and marginal state assuming all the components in a risk state, in the first two columns, and all the components in a marginal state, in the last column.

	Risk \rightarrow Healthy	Risk \rightarrow Marginal	Marginal \rightarrow Healthy
$E[\xi_1^{Water}]$ [days]	75	52.5	22.5
$E[\xi_2^{Water}]$ [days]	75	52.5	22.5
$E[\xi_3^{Water}]$ [days]	75	52.5	22.5
$E[\xi_4^{Water}]$ [days]	5	3.5	1.5

The states at system-of-systems level depend on the degrees of achievement of the goal function (Section 2.2.2). Since in the present case study the goal function can be attained by four different alternative paths (ξ_1^{Water} , ξ_2^{Water} , ξ_3^{Water} and ξ_4^{Water}), their states identify the state of the nuclear power plant. We assume that to be in a healthy state at least one path among ξ_1^{Water} , ξ_2^{Water} and ξ_3^{Water} , (i.e. water from the main input or the designed internal barriers) should be in state 3, i.e., healthy, and another path, including also ξ_4^{Water} (water from the external support), should be at least in state 2, i.e., marginal or healthy. To be in a marginal state, it is necessary that at least one path among ξ_1^{Water} , ξ_2^{Water} , ξ_3^{Water} and ξ_4^{Water} is at least in state 2. All the other combinations lead the nuclear power plant into a risk state.

Table 5 reports the combination of the states of the possible paths ξ_k^{Water} , $k = 1, \dots, 4$, that bring the nuclear power plant into a healthy, marginal or risk state.

Table 5: Definition of risk, marginal and healthy states at system-of-systems level with respect to the states of the alternative paths ξ_k^{Water} , $k = 1, \dots, 4$, that can assure the safety of the nuclear power plant. In the empty space, any state is possible.

	ξ_1^{Water}	ξ_2^{Water}	ξ_3^{Water}	ξ_4^{Water}
Safe	3	3		
	3	2		
	3		3	
	3		2	
	3			2
		3	2	
	2	3		
		3		2
		3	3	
		2	3	
	2		3	
Marginal			3	2
	2	~3	~3	~3
	~3	2	~3	~3
	~3	~3	2	~3
Risk	~3	~3	~3	2
	1	1	1	1

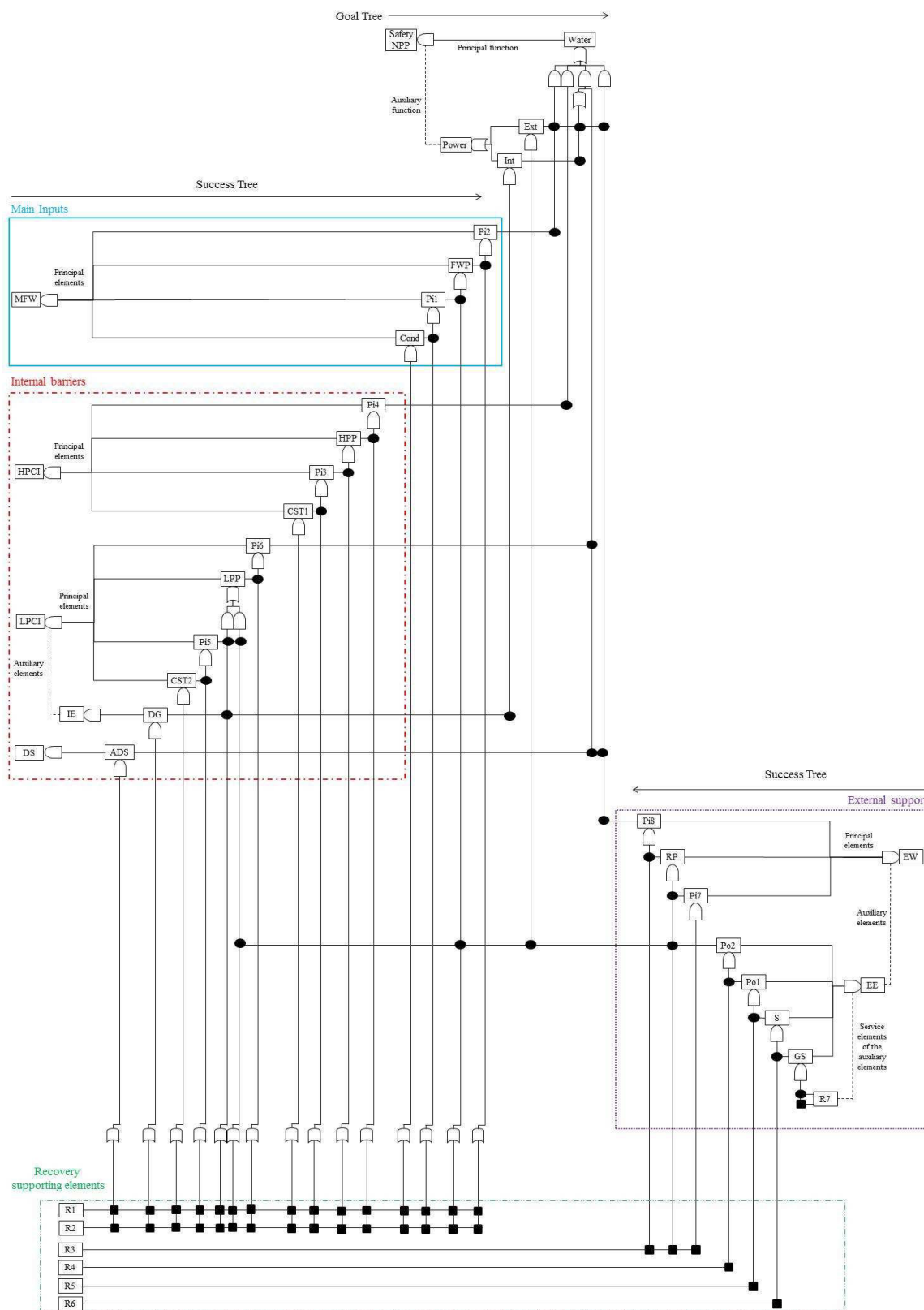


Figure 12: GTST – DMLD of the case study. MFW: Main Feedwater System; HPCI: High Pressure Coolant Injection System; LPCI: Low Pressure Coolant Injection System; IE: Internal Energy System; DS: Depressurization System; EW: External Water System; EE: Offsite power system; R: Road access; GS:

Generation Station, S: Substation, Po: Pole, Pi: Pipe, CST: Condensate Storage Tank, Cond: Condenser; RP: River Pump, HPP: High Pressure Pump; FWP: Feedwater Pump; LPP: Low Pressure Pump, ADS: Automatic Depressurization System; DG: Diesel Generator.

4.3. Results

The Monte Carlo simulation for Seismic Probabilistic Risk Assessment illustrated in Section 3.3 and Appendix C has been applied to the case study of Section 4.1 for an earthquake with moment magnitude equal to 5.5 at the epicenter of coordinates $(x, y) = (70, 70)$ (Figure 9, right). The number of earthquake simulations (N_T) is 2000 and the number of recovery time simulations (N_{RT}) for each components configuration that turns the nuclear power plant (NPP) into a risk or marginal state is 4000.

4.3.1. Safety

Figure 13 shows the comparison of the estimated mean probability that the NPP turns into the states 1 (risk), 2 (marginal) and 3 (healthy), considering multistate and binary state models for the components. As expected, the probability to enter into the risk state is similar for both models (equal to 0.332) and obviously the probability to turn into a marginal state is zero for the binary state model, since this state is not contemplated in such a model.

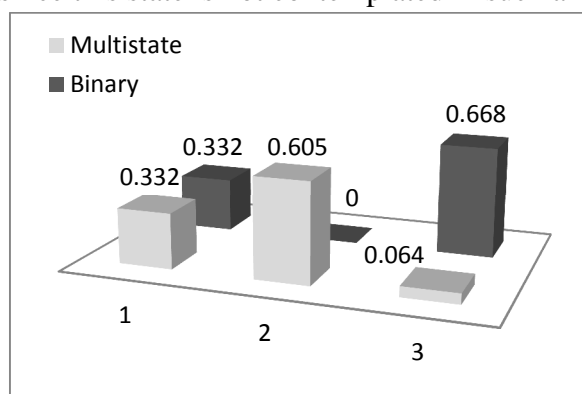


Figure 13: Estimate of the probability that the nuclear power plant reaches a risk (1), marginal (2) and healthy (3) state upon occurrence of an earthquake of moment magnitude equal to 5.5, in the case of multistate (grey) and binary state (black) models.

It can be noticed that the multistate model identifies a criticality in the safety of the NPP, since it shows that the NPP is mostly in a marginal state (0.605). This means that safety margins are not satisfied, and the NPP could be exposed to aftershocks. On the contrary, the binary state model considers these marginal situations as completely safe (healthy), thus underestimating these situations.

Figure 14 shows the same comparison as in Figure 13, except that, for each of the N_T configurations a sequence of aftershocks is simulated N_{RT} times. These values have been obtained by adding (and/or subtracting) to the values of Figure 13, the transition probabilities (Table 6, third column) to enter in (and/or to exit from) the states 1, 2 and 3. These are obtained by the multiplication of the probabilities that the NPP enters in a certain state after

the earthquake (values of Figure 13) and the conditional transition probabilities (Table 6, second column) that the NPP degrades into worse states upon the occurrence of aftershocks, given the state in which it entered after the earthquake.

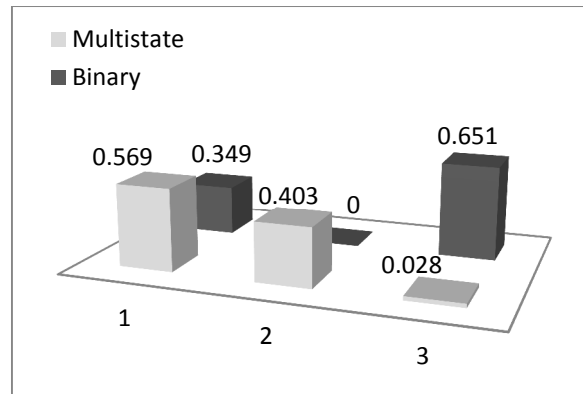


Figure 14: Estimate of the probability that the nuclear power plant reaches a risk (1), marginal (2) and healthy (3) state upon occurrence of an earthquake of moment magnitude equal to 5.5 and upon occurrence of subsequent aftershocks, in the case of multistate (grey) and binary state (black) models.

Table 6: Conditional transition probabilities, given that the NPP entered in a given state after an earthquake (second column), and transition probabilities that the NPP remains in the same state or turns into another (lower) one after the occurrence of a sequence of aftershocks (third column) for the multistate and binary state models. The transitions considered are reported in the first column.

	States transition (from -> to)	Conditional transition probability	Transition probability
Multistate	2 -> 1	0.3861	0.2334
	2 -> 2	0.6139	0.3711
	3 -> 1	0.0597	0.0038
	3 -> 2	0.4987	0.0317
	3 -> 3	0.4416	0.0280
	1 -> 1	1.0000	0.3320
Binary state	3 -> 1	0.0254	0.0170
	3 -> 3	0.9746	0.6510
	1 -> 1	1.0000	0.3320

From Figure 14, it can be seen that, after a sequence of aftershocks, the probability of the NPP to turn into a risk state is higher in the case of the multistate model (i.e., 0.569) than in the case of the binary state model (i.e., 0.349). This is due to the higher probability that the marginal state of the multistate model turns into a risk state (0.2334, in Table 6) with respect to the probability that the healthy state of the binary state model turns into a risk state (0.0170, in Table 6). The first result depends on the definition of marginal state at component and at system-of-systems levels: i) the components in state 2 are more fragile to withstand aftershocks (as explained in Section 2.2.1) and ii) in the present simulation, the configurations of the marginal state of the system of systems after the occurrence of the earthquake are composed mostly (with probability 0.6940) by only one path ζ_k^{Water} , $k = 1, \dots, 4$, in state 2 and

the others in state 1: thus, they are more exposed to the occurrence of aftershocks than configurations composed by all the paths ζ_k^{Water} , $k = 1, \dots, 4$, in state 2 (this situation occurs with probability equal to 0.007). Instead, the low probability value for the transition from healthy state to risk state for the binary state model is explained by the fact that, in this case, there is no distinction among structural and functional state, since they coincide. Therefore, when the NPP is a healthy state also the components are in a structural and functional healthy state.

4.3.2. Physical resilience

In the following, the results of evaluation of the physical resilience of the system of systems are reported. In particular, for the configurations that lead the NPP into a risk state, the recovery from a state 1 to a state 2 (Figure 15 a), from state 2 to state 3 (Figure 15 b), from state 1 to state 3, direct and total (Figure 15 c and d, respectively), is analyzed and, for the configurations that lead the NPP into a marginal state, the recovery from a state 2 to a state 3 (Figure 15 e) is considered.

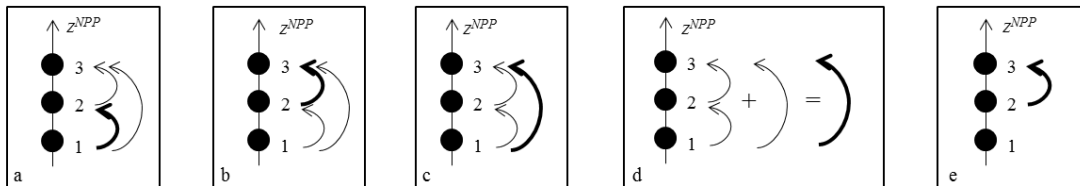


Figure 15: Illustration of the transitions considered (bold lines) for the analysis of the recovery time with respect to the functional state, z^{NPP} , of the nuclear power plant (NPP).

Figure 16 shows the probability density function (PDF) (on the left) and the respective cumulative distribution function (CDF) (on the right) of the time necessary to restore the marginal state of the nuclear power plant from a risk state. As illustrated in the Figure, the transition into a marginal state of the NPP depends on the transition of one of the alternative logic paths ζ_k^{Water} , $k = 1, \dots, 4$, into a state 2. The mean of the distribution is 2.6 days.

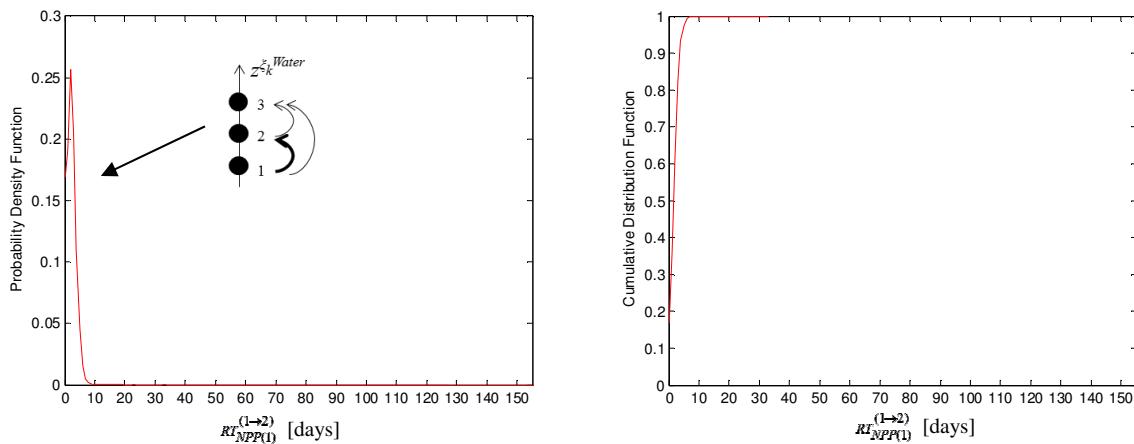


Figure 16: Probability density function (PDF) (on the left) and respective cumulative distribution function (CDF) (on the right) of the time (RT) necessary to restore the marginal state (2) of the nuclear power plant (NPP) from a risk state (1).

In Figure 17, the frequency of the paths ξ_k^{Water} , $k = 1, \dots, 4$, that perform the transition into the states 2 or 3 to lead the NPP in a marginal state are reported on the left, and the details of the frequency of the systems MFW, HPCI, LPCI, DS, IE, EW and EE to be in healthy, marginal or risk state are illustrated, on the right, with respect to Figure 16.

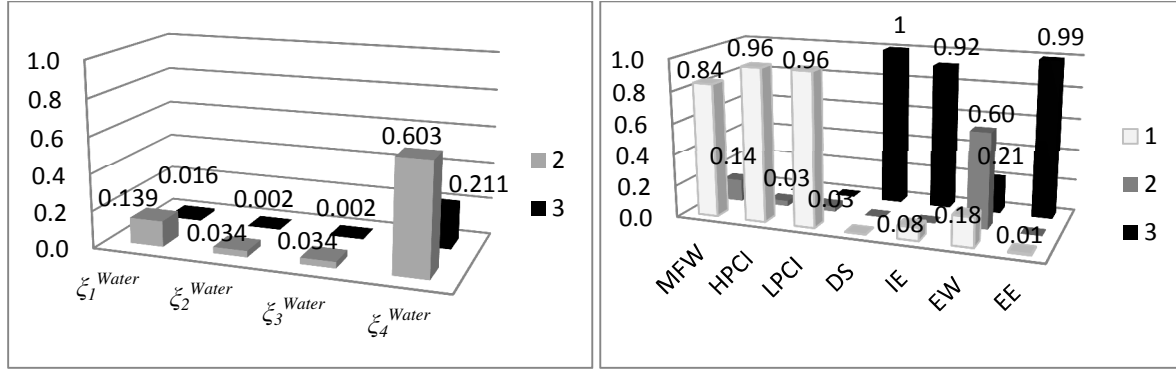


Figure 17: Left: frequency of the paths ξ_k^{Water} , $k = 1, \dots, 4$, that performing a transition into the states 2 or 3 turn the nuclear power plant into a marginal state with respect to Figure 16; Right: corresponding frequency of the Main Feedwater (MFW) system, High Pressure Coolant Injection (HPCI) system, Low Pressure Coolant Injection (LPCI) system, Depressurization System (DS), Internal Energy (IE) system, External Water (EW) system and offsite power (EE) system to be in risk (1), marginal (2) or healthy (3) state.

It can be seen that the transition from the state 1 to the state 2 is mainly due to the path ξ_k^{Water} , $k = 4$, that is formed by the external water system. This system can also turn directly into a state 3 with probability 0.21 (Figure 17, on the right).

Figure 18 shows the probability density function (on the left) and the respective cumulative distribution function (on the right) of the time necessary to restore the healthy state of the nuclear power plant from a marginal state given that the plant entered in a risk state after the occurrence of the earthquake, i.e., after the recovery from risk to marginal state. As shown in Table 5, the recovery of the healthy state requires that i) at least one path among ξ_k^{Water} , $k = 1, \dots, 3$, is in state 3, and ii) another one is in state 2, including also ξ_k^{Water} , $k = 4$.

From the recovery from state 1 to state 2, ξ_k^{Water} , $k = 4$, is in a state higher than 1 with probability equal to 0.814 (Figure 17, left), thus, the PDF of Figure 18 presents mainly the transition of the first condition, i.e., one path among ξ_k^{Water} , $k = 1, \dots, 3$, should turn into a state 3. The distribution presents three peaks: the first one with mean equal to 2.3 days can be due to i) the short recovery of some components, e.g., pipes, of the paths ξ_k^{Water} , $k = 1, \dots, 3$, from state 2 to state 3 or ii) the recovery of the path ξ_k^{Water} , $k = 4$, to state 2, when one of the ξ_k^{Water} , $k = 1, \dots, 3$, paths has previously entered in state 3 in the transition of the NPP from state 1 to state 2. The second peak with mean equal to 21 days is due to the recovery of one of the paths ξ_k^{Water} , $k = 1, \dots, 3$, that has entered previously in a state 2; and the third one, with mean equal to 70 days is due to the recovery of one path among ξ_k^{Water} , $k = 1, \dots, 3$, from state 1 to state 2, and then from state 2 to state 3 or directly from state 1 to state 3. Notice that with very low probability, i.e., around 10^{-5} , the recovery can take from 115 to 151 days to be carried out, as illustrated in the zoom in Figure 18. As explained in the following, this is due

to the presence of aftershocks that in few cases can have a strong impact on the system recovery.

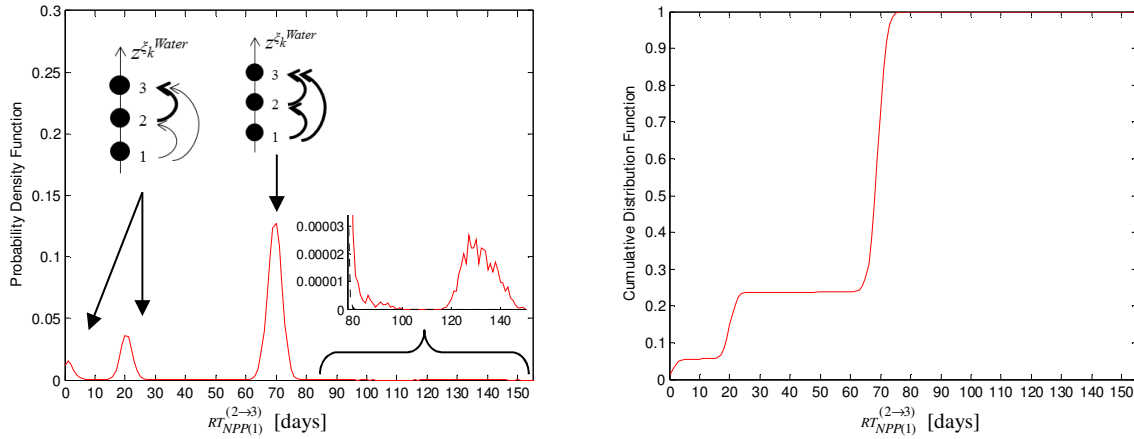


Figure 18: Probability density function (PDF) (on the left) and respective cumulative distribution function (CDF) (on the right) of the time (RT) necessary to restore the healthy state (3) of the nuclear power plant (NPP) from a marginal state (2) given that it entered in a risk state (1) after the earthquake occurrence.

In Figure 19, the frequency of the paths ζ_k^{Water} , $k = 1, \dots, 4$, that perform the transition into the states 2 and 3 to lead the NPP in a healthy state are reported, on the left, and the details of the frequency of the systems MFW, HPCI, LPCI, DS, IE, EW and EE to be in healthy, marginal or risk state are illustrated, on the right, with respect to Figure 18.

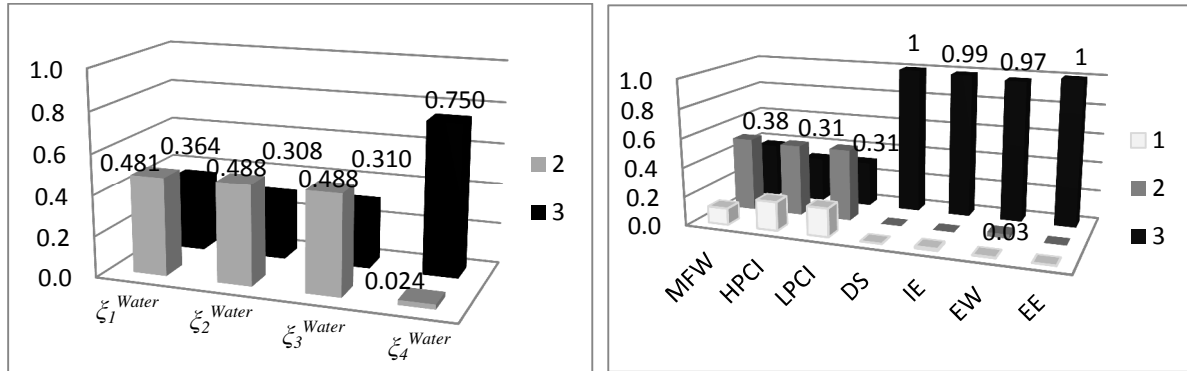


Figure 19: Left: frequency of the paths ζ_k^{Water} , $k = 1, \dots, 4$, that performing a transition into the states 2 or 3 turn the nuclear power plant into a healthy state with respect to Figure 18; Right: corresponding frequency of the Main Feedwater(MFW) system, High Pressure Coolant Injection (HPCI) system, Low Pressure Coolant Injection (LPCI) system, Depressurization System (DS), Internal Energy (IE) system, External Water (EW) system and offsite power (EE) system to be in risk (1), marginal (2) or healthy (3) state.

The external water system is in state 3 with probability 0.97 (Figure 19 on the right). Looking to the other three paths it can be seen that their contribution is similar, slightly higher for ζ_1^{Water} that has previously reached the state 2 with higher probability than ζ_2^{Water} and ζ_3^{Water} , as shown in Figure 17, on the left.

The direct transition of the nuclear power plant from state 1 to state 3 occurs with very low probability, i.e., 0.003 in this simulation, thus, the results of the recovery time are not reported here. However, they are included in Figure 22, where the probability density function and the respective cumulative distribution function of the total time necessary to restore the healthy state of the nuclear power plant, given that the plant entered in a risk state after the occurrence of the earthquake, is reported in comparison with the PDF and CDF obtained by a binary state model.

Figure 20 shows the probability density function (on the left) and the respective cumulative distribution function (on the right) of the time necessary to restore the healthy state of the nuclear power plant, given that the plant entered in a marginal state after the occurrence of the earthquake.

This distribution presents the same three peaks (with means equal to 2.6, 22.3 and 73.2) as the recovery from state 2 to 3 given that the NPP has entered in a state 1 after the earthquake (Figure 18). The explanation of the shape of the distribution is the same as that reported for Figure 18, since the initial state, i.e., the marginal state of the NPP, is the same for both the recovery. The difference in the probability values of the peaks (higher for the first two peaks and lower for the third one) depends on the initial configuration of the marginal state: in the case of Figure 18, the starting configuration before the transition is composed by just one path ζ_k^{Water} , $k = 1, \dots, 4$, in state 2 (or exceptionally in state 3, as illustrated in Figure 17) since it is obtained from the recovery of the NPP from state 1 to 2, whereas in the case of Figure 20, more configurations are possible, e.g., the configuration given by more than one path ζ_k^{Water} , $k = 1, \dots, 4$, in state 2 occurs with probability 0.306. Thus, the recovery can be shorter with higher probability.

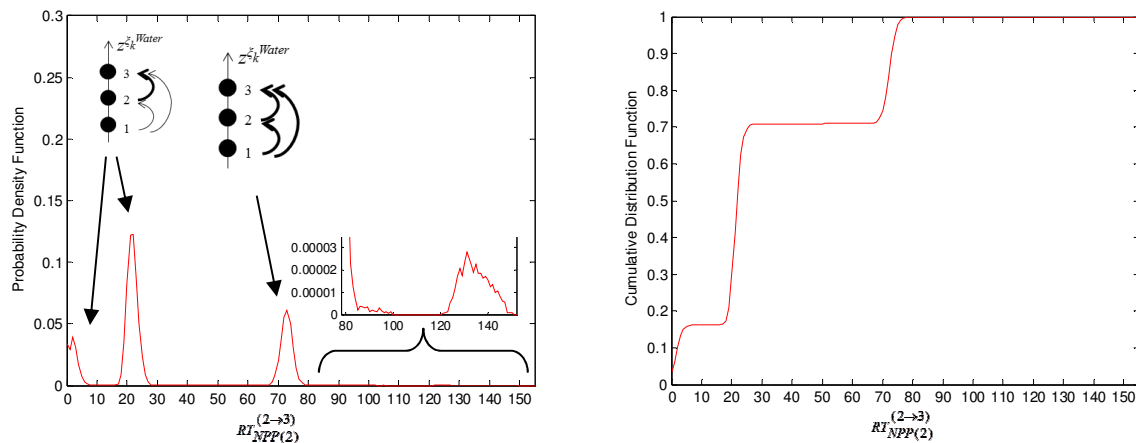


Figure 20: Probability density function (PDF) (on the left) and respective cumulative distribution function (CDF) (on the right) of the time (RT) necessary to restore the healthy state (3) of the nuclear power plant (NPP), given that it entered in a marginal state (2) after the earthquake occurrence.

In Figure 21, the frequency of the paths ζ_k^{Water} , $k = 1, \dots, 4$, that perform the transition into the states 2 and 3 to lead the NPP in a healthy state are reported on the left, and the details of the

frequency of the systems MFW, HPCI, LPCI, DS, IE, EW and EE to be in healthy, marginal or risk state are illustrated, on the right, with respect to Figure 20.

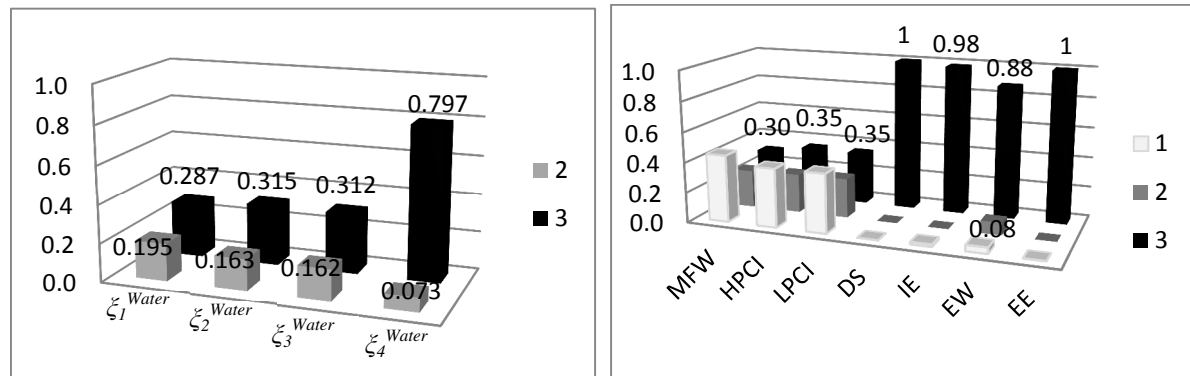


Figure 21: Left: frequency of the paths ζ_k^{Water} , $k = 1, \dots, 4$, that by performing a transition into the states 2 and 3 turn the nuclear power plant into a healthy state with respect to Figure 20; Right: frequency of the Main Feedwater (MFW) system, High Pressure Coolant Injection (HPCI) system, Low Pressure Coolant Injection (LPCI) system, Depressurization System (DS), Internal Energy (IE) system, External Water (EW) system and offsite power (EE) system to be in risk (1), marginal (2) or healthy (3) state.

The contribution of the paths ζ_k^{Water} , $k = 1, \dots, 3$, to turn the NPP into a healthy state is similar (frequency around 0.3).

Figure 22 shows the comparison among the probability density function (on the left) and the respective cumulative distribution function (on the right) of the time necessary to restore the healthy state of the nuclear power plant, given that the plant entered in a risk state after the occurrence of the earthquake, by multistate (solid line) and binary state (dashed line) models.

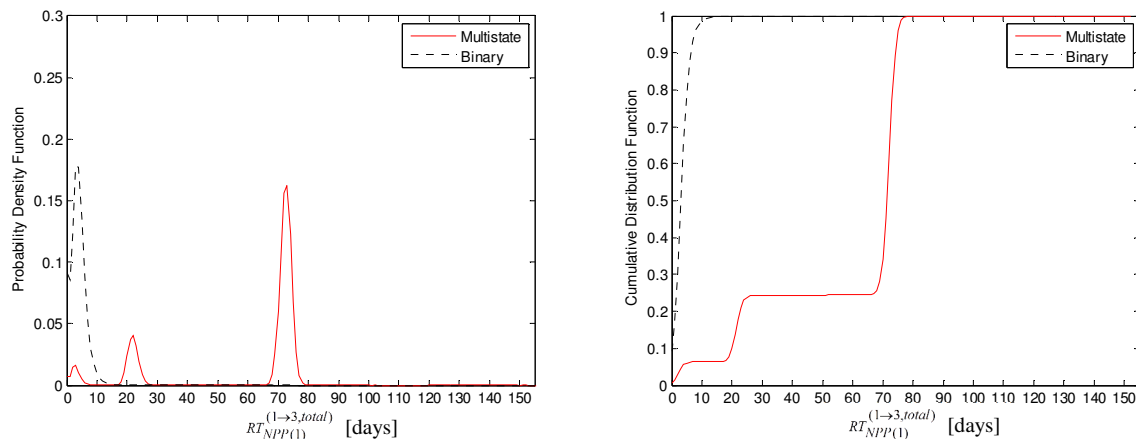


Figure 22: Comparison of the probability density function (PDF) (on the left) and respective cumulative distribution function (CDF) (on the right) of the time (RT) necessary to restore the healthy state (3) of the nuclear power plant (NPP) from a risk state (1), in the case of a multistate (solid line) and binary state (dashed line) model.

The PDF obtained by the binary state model is shifted at low values with mean equal to 4.31 days, whereas the PDF resulted from the multistate model presents three peaks with means 3.2, 22.4, 73 days, the peak with highest mean being widely dominating the other two in

probability mass terms. The binary state model results in a short time for the NPP to recover its full safety; the multistate model instead leads to a different conclusion, that is: the time to reach a healthy state is short with low probability (the first peak has probability mass equal to 0.06), due to few "lucky" configurations of failed components that can be easily recovered after the earthquake, but it is higher with large probability (a probability mass of 0.18 concentrated around the second peak of 21.4 days and a probability mass of 0.76 around 73 days).

Comparing the results obtained by the binary state model with those of the multistate model for the recovery of the marginal state (Figure 16), it can be seen that the time needed to recover the NPP to a marginal state (mean value equal to 2.6), is lower than that required by the binary state model to recover the healthy state. In conclusion, the above results show the importance of resorting to a multistate modelling framework, to capture the insight that safety is reached faster than as resulting from a simplistic binary state assumption, but on the other hand, it is recognized that such safety is not "complete" with respect to the required safety margins, for the achievement of which more time is needed.

From the recovery viewpoint, there is a slight difference between the results given by a multistate model considering and not considering aftershocks when short recovery from a risk state, e.g., from a risk to a marginal (or directly to a healthy) state, are considered, since the component in a risk state cannot degrade further if an aftershock occurs. On the contrary, the impact of the aftershocks in the recovery can be seen in the transition from a marginal to a risk state, as illustrated in Figure 23 where the comparison of the probability density functions (on the left) and the respective cumulative distribution functions (on the right) of the time necessary to restore the healthy state of the nuclear power plant, given that the plant entered in a marginal state after the occurrence of the earthquake considering (solid line) and not considering (dashed line) the occurrence of aftershocks, is illustrated.

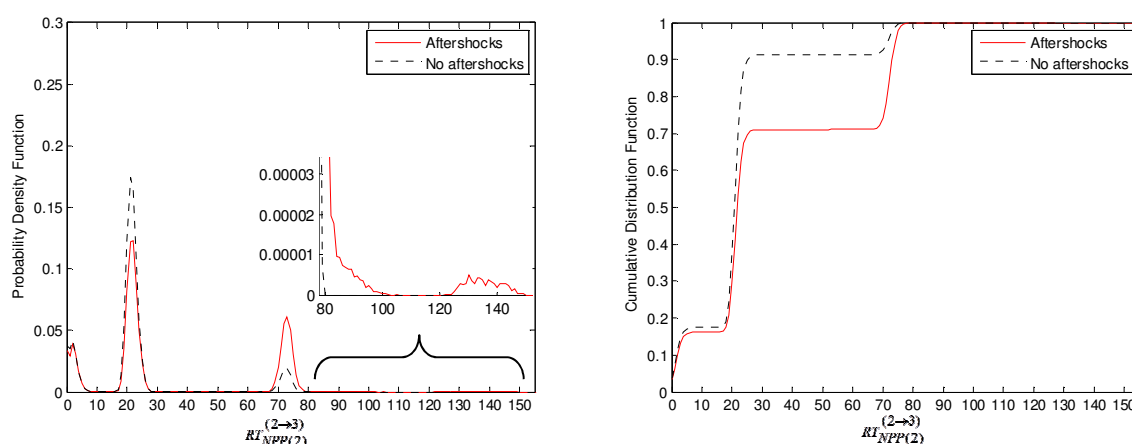


Figure 23: Comparison of the probability density functions (PDFs) (on the left) and respective cumulative distribution functions (CDFs) (on the right) of the time (RT) necessary to restore the healthy state (3) of the nuclear power plant (NPP), given that it entered in a marginal state (2) after the earthquake occurrence, considering (solid line) and not considering (dashed line) the occurrence of aftershocks.

The two probability density functions show the same peaks with mean around 2.5, 22.2, 73.2 days, but in the case with aftershocks the probability values are lower for the first two peaks and higher for the third one than in the case without aftershocks. Thus, in the case with aftershocks, the probability that the recovery needs more time is higher; in addition, there is a small probability, i.e., around 10^{-6} , that the recovery is carried out in more than 120 days, as illustrated in the zoom of Figure 23.

5. CONCLUSIONS

We have significantly extended a system-of-systems framework previously proposed by the authors for the analysis of the risk of a critical plant (e.g., a nuclear power plant) from natural external events (e.g., earthquakes).

We have explicitly modelled the different parts of the system-of-systems into i) main inputs, i.e., the infrastructure systems devoted to provide the main supply for the safety of the nuclear power plant, ii) internal barriers, i.e., the internal emergency devices designed to automatically activate in emergency conditions, iii) external supports, i.e., the redundant infrastructure systems that can replace the main inputs and the internal barriers when they do not function, iv) the recovery supporting elements, i.e., the infrastructure systems that can be a support in the actions to keep or restore the safety of the plant.

We have adopted a multistate model distinguishing structural damage and functional performance of the individual components, that reflects into a multistate model of the system of systems based on different degrees of safety (risk, marginal and healthy) of the nuclear power plant.

We have represented the system of systems with a Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) and we have used Monte Carlo simulation for the probabilistic evaluation of the safety of the nuclear power plant and its physical resilience, measured in terms of the time needed to restore the safety. In addition, we have included the impacts of aftershocks.

In particular, by exemplification of a case study concerning the seismic risk of a nuclear power plant, the following analyses have been carried out:

- a. a comparison between the probabilities that the nuclear power plant enters in risk, marginal and healthy states calculated by multistate and binary state models: as expected, the probability to enter in a risk state is the same for both models, whereas the probability to be in a healthy state is lower for the multistate model that identifies (marginal) configurations of the system of systems that present criticalities because not satisfying safety margins;
- b. a comparison of the previous probabilities (a.) considering also sequences of aftershocks that could further degrade the safety of the nuclear power plant. The multistate models evidences a higher probability that the nuclear power plant enters into a risk state (+ 0.2372) than the binary state model (+ 0.0170). Thus, it can capture the impact of the aftershocks that are almost neglected by the binary state model since

the structural healthy state of the components is characterized by fragilities that are not much sensitive to small ground motion levels produced by aftershocks. Actually, the increased probability of the risk state is mainly (0.2334) due to the degradation of the marginal state that is more exposed to aftershocks than the healthy state;

- c. a comparison of the probability density function (PDF) and the respective cumulative distribution function (CDF) of the time necessary to restore the healthy state of the nuclear power plant, given that the plant has entered in a marginal and risk state, and the recovery time of the marginal state given that the plant has entered in risk state, with the i) binary state model and ii) multistate model without considering the occurrence of the aftershocks:
 - i) From the first comparison, it can be seen that the binary state model is less conservative than the multistate model in that it identifies a mean time to recover the healthy state lower than the one identified by the multistate model, but higher than the one needed to recover a marginal state. On the contrary, the multistate model is capable of capturing the fact that a faster recovery to reach a safe condition is possible, but this condition is marginal with respect to the safety margins and a longer time is needed to arrive at a completely safe state, including the safety margins.
 - ii) From the second comparison, important differences cannot be seen in the recovery time distribution for fast recovery from risk states, e.g., from risk to marginal state, since, in this work, a component in risk state cannot further degrade into a worse state. A further development of the model will be done in the future to take into account the disturbance of the aftershocks for the components in risk state. On the contrary, the impact of aftershocks is evident in the recovery from a marginal state to healthy state since the components in state 2 can degrade to state 1 more than once during the total recovery. As a consequence, the time needed for the restoration of the healthy state increases considering the occurrence of aftershocks.

The results obtained, albeit performed on a simplified case study and under limiting assumptions, highlight that the multistate model is relevant to identify marginal conditions of safety of the critical plant that may turn into a risk state. This can be relevant for the decision making related to safety-critical issues when external events occur: a marginal condition may degrade to a risky one but this would not happen (or it would happen with very small probability, e.g, 0.0038 in the present case study) for a complete safe state that can mainly degrade to a marginal state. On the contrary, the binary state model does not allow these considerations since it does not distinguish different safety levels; in this case, a complete safe state can directly change into a risk state. However, this is not evident in the simulation: the healthy state turns into a risk state with probability 0.0170 (Section 4.3.1), as explained in the point b. above. Thus, the multistate model allows identifying criticalities that are hidden in a binary model and that can lead to an underestimation of the risk. The multistate model is more

conservative than the binary state one; this can be seen also from the results related to the system resilience characteristics, where the time necessary to restore the complete safety is longer than that needed with a binary state model for most of the cases. However, as explained before, the complete safety of the binary state model hides criticalities and it can be affected by aftershocks. The multistate model, instead, shows that restoration of the marginal safety can occur in a shorter time; the faster recovery is associated with the awareness that safety margins are not satisfied. These findings can help to improve the structural/functional responses of the critical elements of the alternative logic paths, for improving the global resilience of the system of systems so as to increase the safety of the critical plant. The multistate model is a valid support for achieving these goals, provided that the definition of the structural and functional limit states is carefully addressed.

Future work will be devoted to apply the framework of analysis presented to a critical networked infrastructure and to consider advanced simulation techniques in order to render more efficient the computation.

REFERENCES

- [1] Manyena SB. The concept of resilience revisited. *Disasters*. 2006; 30:434-450.
- [2] Cimellaro GP, Reinhorn AM, Bruneau M. Framework for analytical quantification of disaster resilience. *Engineering Structures*. 2010; 32: 3639-3649.
- [3] Zio E, Ferrario E. A framework for the system-of-systems analysis of the risk for a safety- critical plant exposed to external events. *Reliability Engineering and System Safety*. 2013; 114: 114-225.
- [4] Li Y, van de Lindt JW. Loss-based formulation for multiple hazards with application to residential buildings. *Engineering Structures*. 2012; 38: 123-133.
- [5] Selva J, Kakderi K, Alexoudi M, Pitilakis K. Seismic performance of a system of interdependent lifeline and infrastructure components. 8th International Conference on urban Earthquake Engineering, Tokyo Institute of Technology, Japan. 2011.
- [6] Billinton R, Karki R. Application of Monte Carlo simulation to generating system well-being analysis. *Power Systems, IEEE Transactions on*. 1999; 14(3):1172-1177.
- [7] Hu YS, Modarres M. Evaluating system behavior through Dynamic Master Logic Diagram (DMLD) modeling. *Reliability Engineering and System Safety*. 1999; 64:241-269.
- [8] Kalos MH, Whitlock PA. Monte Carlo methods. Vol. 1, Basics. New York: Wiley; 1986. 186 p.
- [9] Zio E. Computational methods for reliability and risk analysis. Series on Quality, Reliability and Engineering Statistics, Vol 14. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2009. Chapter 2, Monte Carlo simulations for reliability and availability analysis; p. 59-69.
- [10] Zio E. The Monte Carlo Simulation Method for System Reliability and Risk Analysis. London: Springer Series in Reliability Engineering; 2012.
- [11] USNRC Glossary <http://www.nrc.gov/reading-rm/basic-ref/glossary.html>. 2013
- [12] Brissaud F, Barros A, Bérenguer C, Charpentier D. Reliability analysis for new technology-based transmitters. *Reliability Engineering and System Safety*. 2011; 96: 299-313.
- [13] Modarres M, Kaminskiy M, Krivtsov V. Reliability engineering and risk analysis: a practical guide. New York: CRC Press; 1999
- [14] Huang YN, Whittaker AS, Luco N. A probabilistic seismic risk assessment procedure for nuclear power plants: (I) Methodology, *Nuclear Engineering and Design*. 2011; 241: 3996– 4003.
- [15] Federal Emergency Management Agency Multi-hazard Loss Estimation Methodology, Earthquake Model HAZUS^{MH} MR4. Technical Manual. 2003. www.fema.gov/plan/prevent/hazus.

- [16] Guerini P, Paci S. Appunti di impianti nucleari. Parte III: componenti. Facoltà di Ingegneria. Dipartimento di ingegneria meccanica, nucleare e della produzione. Università di Pisa. 1998.
- [17] Final Safety Analysis Report of the Kuosheng Nuclear Power Station. Taiwan Power Company. 1988.
- [18] Cimellaro GP, Reinhorn AM. Multidimensional Performance Limit State for Hazard Fragility Functions. *Journal of Engineering Mechanics*. 2010; 1:156.
- [19] Ferrario E, Zio E. Assessing nuclear power plant safety and recovery from earthquakes using a system-of-systems approach, submitted to ESREL 2012 Special Issue of Reliability Engineering and System Safety. 2013.
- [20] Seismic Probabilistic Risk Assessment Implementation Guide, EPRI, Palo Alto, CA: 2003. TR-1002989.
- [21] Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Expert. Main Report, Vol. 1. 1997, NUREG/CR-6372 UCRL-ID- 122160. Supported by U.S. Nuclear Regulatory Commission (NRC), the U.S. Department of Energy (DOE); and the Electric Power Research Institute (EPRI).
- [22] Sen TK. Fundamentals of seismic loading and structures. Singapore: John Wiley & Sons, Ltd; 2009. Chapter 7, Probabilistic Seismic Hazard Analysis; p. 181-218.
- [23] Gutenberg B, Richter CF. Frequency of earthquakes magnitude in California, *Bulletin of the Seismological Society of America*. 1944; 34:185-188.
- [24] Kanamori H. The energy release in great earthquakes. *Journal of Geophysical Research*. 1977; 82(20): 2981–2987.
- [25] Kanamori H. Magnitude scale and quantification of earthquakes. In: SJ. Duda and K. Aki Editors. *Quantification of Earthquakes*. Tectonophysics, 1983; 93: 185-199.
- [26] Kramer SL. *Geotechnical Earthquake Engineering*, Prentice Hall, New Jersey. 1996.
- [27] Weatherill GA, Burton PW. The application of multiple random earthquake simulations to probabilistic seismic hazard assessment in the Aegean region. *Firs European Conference on Earthquake Engineering and Seismology*. Geneva, Switzerland. 2006.
- [28] Ambraseys NN, Douglas J, SARMA SK, Smit PM. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering*. 2005; 3:1-53.
- [29] A Bayesian Network Methodology for Infrastructure Seismic Risk Assessment and Decision Support. PEER Report 2011/02. Pacific Earthquake Engineering Research Center College of Engineering, University of California, Berkeley.
- [30] Ryu H, Luco N, Uma SR, Liel AB. Developing fragilities for mainshock-damaged structures through incremental dynamic analysis. *Proceedings of the Ninth Pacific Conference on Earthquake Engineering Building an Earthquake-Resilient Society*, Auckland, New Zealand, Paper 225. 2011.
- [31] Réveillère A, Gehl P, Seyedi D, Modaressi H. Development of seismic fragility curves for mainshock-damaged reinforced-concrete structures. *15th World Conference on Earthquake Engineering*, Lisboa, Portugal, 2012.
- [32] Bath M. Lateral inhomogeneities in the upper mantle. *Tectonophysics*. 1965; 2:483-514.
- [33] Utsu T, Ogata Y, Matsuura RS. The centenary of the Omori formula for a decay law of aftershock activity. *J Phys Earth*. 1995. 43(1):1-33.
- [34] Shcherbakov R, Turcotte DL, Rundle JB. A generalized Omori's law for earthquake aftershock decay. *Geophys Res Lett*. 2004; 31(11):1613-1624.
- [35] Omori F. On the aftershocks of earthquakes. *J Coll Sci Imp Univ Tokyo*. 1984; 7:113-200.
- [36] Zhao J, Liu Y, Zhou Z, Zhao C. Spatio-temporal characteristics of strong aftershocks of the M_s8.0 Wenchuan earthquake. *Earthquake Science*. 2010; 23(3):215-221.

APPENDIX A. Qualities, parts and GTST-DMLD within a system-of-systems framework: an example

For illustration purpose, let us consider the main function F^* of a critical plant H , i.e., the critical element E , achieved through the success of two principal functions, F_1 and F_2 , where the former is in turn obtained by the combination of functions $F_{1,1}$ and $F_{1,2}$. In addition, we consider an auxiliary function F_3 that is not needed directly for achieving F^* , but it serves the function F_2 . In the hierarchy, the function F_3 is represented in a parallel branch connected to F^* by a dashed line (Figure A.1).

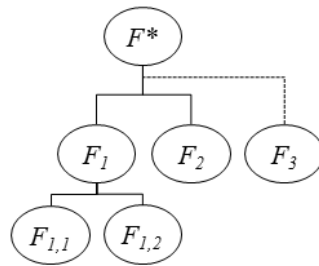


Figure A.1: Hierarchy of the qualities of the example proposed.

Figure A.2 represents the graph of the components of this example with respect to the safety levels of Figure 2. The links show the relationship among the components; they are directed from an element to another dependent on it. The safety of a critical element E (star) is assured by $A = 8$ systems divided into $n^{MI} = 1$ system of main inputs, $S^{(1)}$, $n^{IB} = 3$ internal barriers, $S^{(2)}$, $S^{(3)}$ and $S^{(4)}$, $n^{ES} = 2$ external supports, $S^{(5)}$ and $S^{(6)}$, $n^{RS} = 2$ recovery supporting elements, $S^{(7)}$ and $S^{(8)}$, represented in dashed oval shape. The components included in these systems are represented in solid oval shape. For example, the system $S^{(1)}$ is formed by 3 components ($S_1^{(1)}$, $S_2^{(1)}$, $S_3^{(1)}$), the system $S^{(2)}$ is composed by 1 component, $S_1^{(2)}$, and so on. Notice that there are some components that are directly connected to E , e.g., $S_3^{(1)}$ and $S_1^{(2)}$, and others that are connected to the components of other systems, e.g., $S_1^{(3)}$ is connected to $S_1^{(2)}$. The first type of components belongs to principal systems, whereas the latter one to the auxiliary systems, except for the recovery supporting elements that are considered apart from these systems for their role of recovery, as explained in Section 2.1. Each system $S^{(a)}$, $a = 1, \dots, 8$, can be represented in the form of a hierarchy as illustrated in Figure A.3.

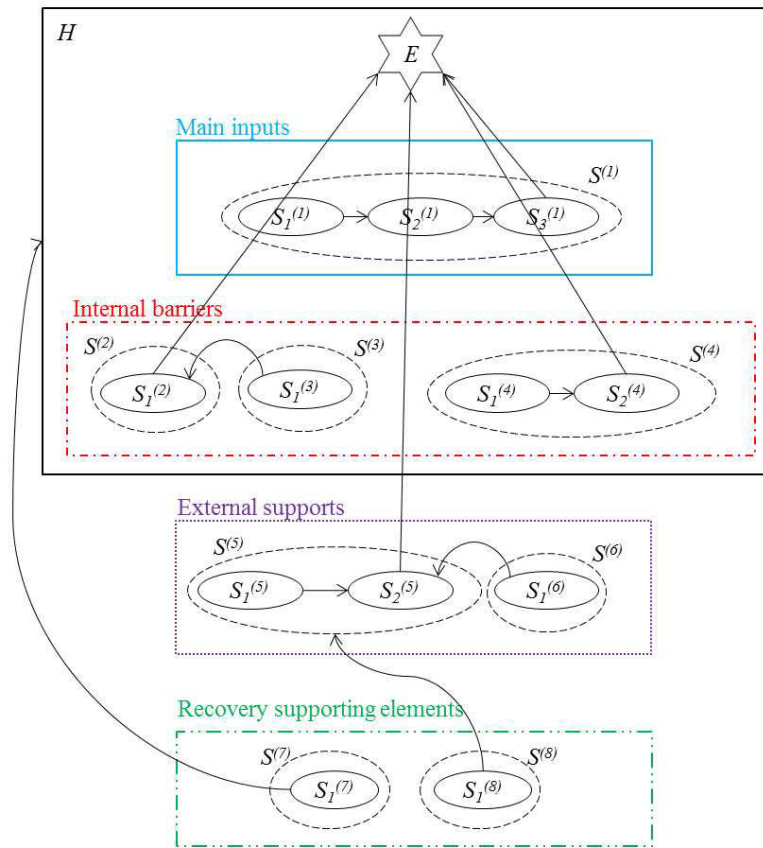


Figure A.2: Graph of the physical components for the example proposed.

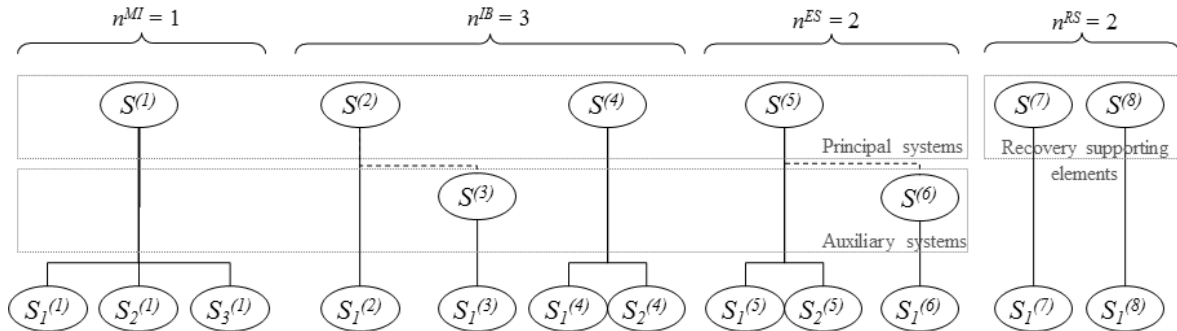


Figure A.3: Hierarchy of the parts of the example proposed; n^{MI} , n^{IB} , n^{ES} , n^{RS} refer to the number of main inputs, internal barriers, external supports and recovery supporting elements, respectively.

In Figure A.4, the GTST-DMLD of the example above is reported. The goal tree is the hierarchy of Figure A.1 and the success tree is composed by the hierarchies of Figure A.3. The dot- and square- dependencies detail the connections of the graph of Figure A.2 and connect the physical elements to the functions.

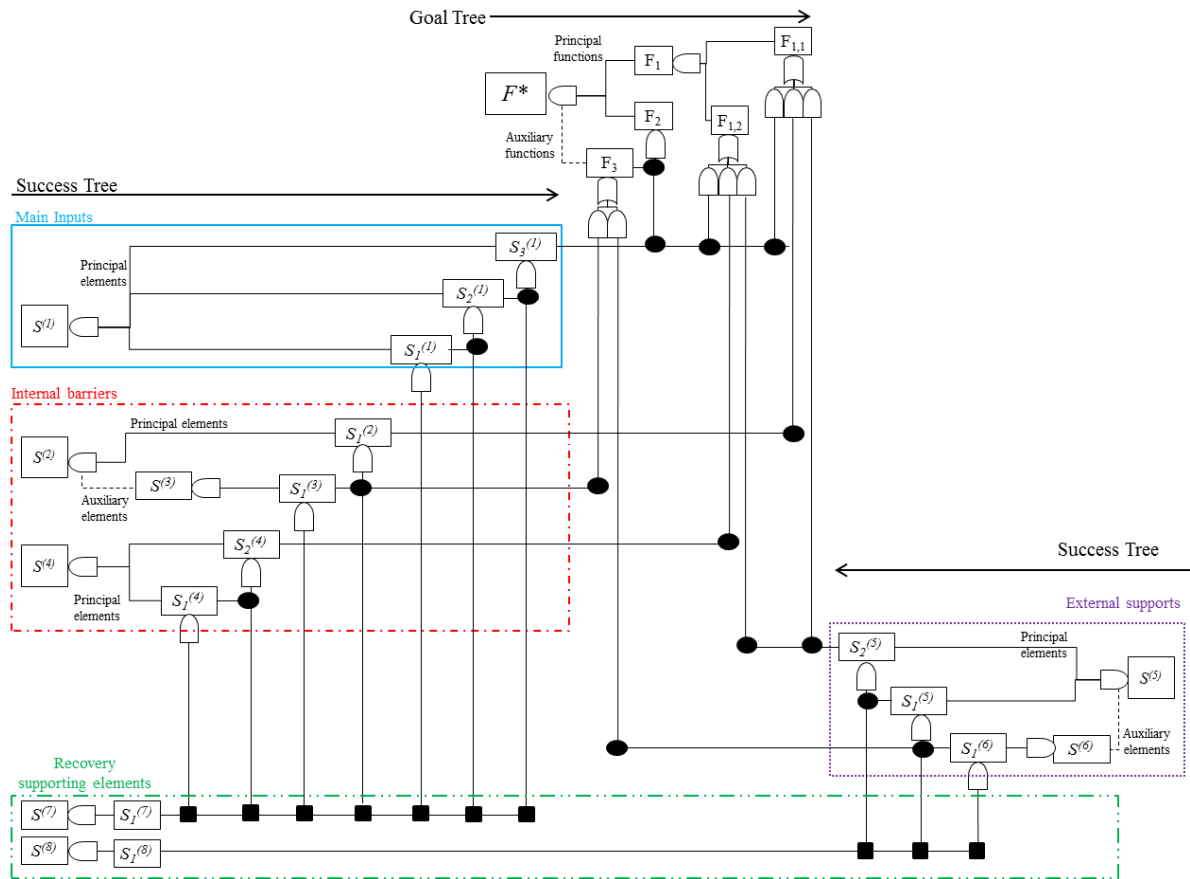


Figure A.4: GTST – DMLD of the example considered.

APPENDIX B. Seismic Probabilistic Risk Assessment

Since the exemplification of the modelling framework is done with reference to a nuclear power plant as critical plant and earthquakes as the external events, in Appendix B.1 some basic information on the procedure for Seismic Probabilistic Risk Assessment (SPRA) of a nuclear power plant is given; aftershocks are also considered (Appendix B.2).

B.1 Seismic risk

The risk on a system deriving from an earthquake (hereafter referred to as the main shock) is evaluated by a procedure of Seismic Probabilistic Risk Assessment (SPRA) that consists of three parts: i) Seismic Hazard Analysis, ii) Seismic Fragility Evaluation and iii) System Analysis [20].

The first part is aimed at computing the probabilities of occurrence of different levels of earthquake ground motion at a site of interest. It is traditionally developed as a Probabilistic Seismic Hazard Analysis (PSHA) consisting of four procedural steps [20], [21], [22]:

- 1) Identification and characterization of the earthquake source;

- 2) Definition of the earthquake recurrence relationship, i.e., the annual frequency of occurrence of a given magnitude event for each source, typically described by the Gutenberg-Richter law [23]:

$$\log(n^{etq}) = a - bm^{etq} \quad (B.1)$$

where n^{etq} is the number of earthquakes with magnitude¹ greater than m^{etq} , and a and b are parameters obtained by data regression analysis [20], [21], [22]. This relation implies a double truncated distribution for the magnitude [26], [27]:

$$F_M(m^{etq}) = \frac{1 - e^{-\beta(m^{etq} - m_{\min}^{etq})}}{1 - e^{-\beta(m_{\max}^{etq} - m_{\min}^{etq})}} \quad (B.2)$$

where β represents the relative frequency of smaller to larger events, and m_{\max}^{etq} and m_{\min}^{etq} are the upper and lower bounds of the magnitude, respectively, that avoid the high values which are unrealistic and the low values that are negligible.

- 3) Formulation of the ground motion attenuation relationship that identifies the ground motion value at the site of interest, e.g., the peak ground acceleration, given the source-to-site distance and the magnitude. The higher the distance from the source, the lower is the ground motion value. The following relationship described by [28] has been adopted in this paper:

$$\log_{10} z' = C_1 + C_2 m^{etq} + (C_3 + C_4 m^{etq}) * \log_{10} \sqrt{r^2 + C_5^2} + C_6 S_S + C_7 S_A + C_8 F_N + C_9 F_T + C_{10} F_O \quad (B.3)$$

where m^{etq} is the earthquake magnitude, r is the source-to-site distance, S_S and S_A represent the types of soil (soft, stiff or rock, when both variables are set to zero) and F_N , F_T and F_O describe the faulting mechanism (normal, thrust or odd). Equation B.3 has been derived by weighted regression analysis on a set of strong-motion records collected in Europe and in Middle Est [28].

- 4) Computation of the exceedance probability of ground motion in any time interval by analytical integration for each magnitude, distance and ground motion value.

The second part of the SPRA identifies the seismic capacity of a component in terms of its conditional probability of failure f' for any given ground motion level z' [20]:

$$f' = \Phi \left[\frac{\log(z' / A_m) + \beta_u \Phi^{-1}(Q)}{\beta_r} \right] \quad (B.4)$$

where Q is the subjective probability of not exceeding a fragility f' , A_m is the median acceleration capacity, β_r and β_u are the logarithmic standard deviation due to randomness and to uncertainty in the median capacity, respectively. Considering different damage states of a component, “failure” means generically “degree of damage”: thus, the fragility is the conditional probability of exceeding a level of damage for any given ground motion level [29]. The damage states are therefore identified by the fragility curves. A fragility evaluation

¹ The magnitude scale typically used is the moment magnitude defined by [24]. For medium size earthquakes it is similar to the Richter values [25].

is carried out to provide the parameter values (A_m , β_r and β_u) for the fragility model. This evaluation is performed for critical failure modes by considering safety margins inherent in capacity predictions, response analysis and equipment qualification [20]. Recent studies [30], [31] have been devoted to identifying methodologies for developing “aftershocks fragilities”, i.e., fragility curves for main shock-damaged structures that are initially in a given damage state due to the occurrence of an earthquake. Since the estimation of the fragility parameters is not the objective of the present work, in our evaluation we have assumed arbitrarily the parameter values to determine the damage states due to main shocks and aftershocks (see Section 4.1.1).

In the third part, the outputs of the hazard and fragility analyses are integrated to evaluate the impact of an external event to the system of interest [20]. In this work, we adopt a Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) representation for the analysis of the impact on the system and Monte Carlo simulation for the quantitative evaluation. In extreme synthesis, Monte Carlo simulation is used for determining the state of each component of the system as a result of the impact of the external event on the component given its fragility in terms of its probability of exceeding different damage states for a given ground motion level. Then, the GTST-DMLD accounts for the dependencies among all the components and their states for determining the state of the entire system due to the impact of the external event. This part is described in detail in Section 3 and Appendix C.

B.2 Aftershocks

Aftershocks, small earthquakes that occur naturally after the main shock, can further degrade the conditions of a component or a system. In this work, we compute their impacts on the system of interest by the same SPRA procedure explained above for the earthquake (Appendix B.1).

According to Bath’s law [32], the difference, Δ , between the magnitude of an earthquake, m^{eq} , and of its largest aftershock, m_{\max}^{af} , is a constant, independent on the earthquake magnitude, and typically approximated to 1.2:

$$\Delta = m^{eq} - m_{\max}^{af} = 1.2 \quad (B.5)$$

As for the earthquake, the recurrence relationship of aftershocks is described by the Gutenberg-Richter law (eq. B.1) and their magnitude, m^{af} , is still represented by the double truncated distribution of eq. B.2, computing the maximum magnitude, m_{\max}^{af} , from eq. B.5 and defining a minimum magnitude, m_{\min}^{af} , of interest.

The temporal decay activity of aftershocks follows the modified Omori’s law [33], [34]:

$$\lambda(t) = \frac{W}{(c+t)^p} \quad (B.6)$$

where $\lambda(t)$ is the occurrence rate of aftershocks with magnitude greater than the minimum magnitude of aftershocks considered, m_{\min}^{af} , t is the time passed from the earthquake and p , c and W are parameters which depend on the geophysics of the environment. Assuming $p = 1$,

as in the original formulation of the Omori's law [35], and fixing the value of the parameter c , e.g., $c = 0.05$ [36], it is possible to identify the parameter W comparing the integral of eq. B.6 in a time window $[0, T^*]$ (e.g., $[0, 365]$ days [36]) with the maximum number of aftershocks, n_{\max}^{af} , that can occur in one year [36]:

$$\int_0^{T^*} \frac{W}{(c+t)^p} dt = n_{\max}^{af}, \quad (\text{B.7})$$

where $n_{\max}^{af} = 10^{a-b*m_{\min}^{af}}$ from the Gutenberg-Richter law (eq. B.1).

Once that all the parameters of the occurrence rate $\lambda(t)$ are determined, the number of aftershocks in the intervals of time $[0, T_i]$, $T_i = 1, 2, \dots, T^*$, can be computed by solving the integral of eq. B.7. Normalizing these values with respect to the maximum number of aftershocks, n_{\max}^{af} , we can obtain the cumulative distribution function (CDF) of the occurrence time of aftershocks.

In this work we have considered the occurrence of an earthquake of magnitude, m^{eq} , equal to 5.5 on the moment magnitude scale, followed by a sequence of aftershocks whose minimum moment magnitude value, m_{\min}^{af} , is 3 (assumed) and the maximum, m_{\max}^{af} , is 4.3 (computed by eq. B.5). We have fixed the parameter b of the Gutenberg-Richter law to 1, since it can vary in the range 1 ± 0.3 [26], and we have computed the parameter a of the same law by assuming $n^{af} = 1$ with respect to the magnitude of the largest aftershock, m_{\max}^{af} , i.e., by assuming that an aftershock that has a magnitude equal to that of the largest aftershock can occur once in a year. Then, given the parameters a ($a = 4.3$) and b we have obtained $n_{\max}^{af} = 20$ from the Gutenberg-Richter law, considering the magnitude m_{\min}^{af} . Assuming the parameters $p = 1$ [35] and $c = 0.05$ [36], we have determined the value of W ($W = 2.25$) from eq. B.7 in a time window equal to $[0, 365]$ days. The CDF obtained is illustrated in Figure B.1.

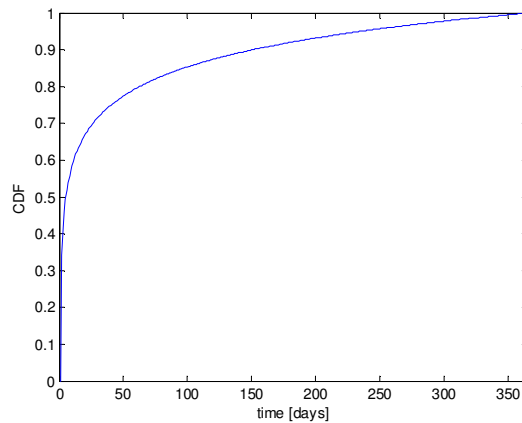


Figure B.1: Cumulative distribution function (CDF) of the occurrence time of aftershocks.

APPENDIX C. Monte Carlo simulation for Seismic Probabilistic Risk Assessment within a system-of-systems framework: operative steps

The simulation procedure consists of the following operative steps:

1. choose a value of earthquake magnitude and epicenter coordinates with respect to which the analysis is performed;
2. compute by eq. B.3 the ground acceleration value at each of the η , $\eta = 1, \dots, L$, components in the last levels of the physical hierarchies of the systems $S^{(a)}$, $a = 1, \dots, A$; L is the total number of components of the system of systems;
3. compute the fragilities, $\{f\}$, for all the components of the system of systems by eq. B.4; $\{f\}$ is a matrix of $2 \times L$ values (two for each component), representing the conditional probability of exceeding a marginal ($f_{1,\eta}$, $\eta = 1, \dots, L$) and risk ($f_{2,\eta}$, $\eta = 1, \dots, L$) threshold;
4. sample a matrix of uniform random numbers in $[0,1)$ $\{u_v^\eta\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, L$, where N_T is the number of simulations;
5. determine the structural multistate matrix $\{g_{j,v}^\eta\}$, $j \in \{1, 2, 3\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, L$, where j represents the structural state index, by comparing the matrix $\{u_v^\eta\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, L$ with the fragility $\{f\}$: if $u_v^\eta > f_{1,\eta}$, set $\{g_{j,v}^\eta: j = 3\}$; if $f_{2,\eta} < u_v^\eta < f_{1,\eta}$, set $\{g_{j,v}^\eta: j = 2\}$; otherwise if $u_v^\eta < f_{2,\eta}$, set $\{g_{j,v}^\eta: j = 1\}$ for $v = 1, \dots, N_T$ and $\eta = 1, \dots, L$. When $\{g_{j,v}^\eta: j = 1\}$, it means that in the v -th simulation the η -th component is strongly hit by the earthquake, i.e., it enters in a risk state; when $\{g_{j,v}^\eta: j = 2\}$, it means that in the v -th simulation the η -th component is slightly hit by the earthquake, i.e., it enters in a marginal state; otherwise, when $\{g_{j,v}^\eta: j = 3\}$, in the v -th simulation the η -th component survives the earthquake, i.e., it remains in a healthy structural state. Each row of the matrix g represents the states of the L system components in the v -th simulation;
6. determine the functional multistate matrix $\{z_{i,v}^\eta\}$, $i \in \{1, 2, 3\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, L$, where i represents the functional state index, on the basis of the relationships between the structural and functional states of component η ;
7. determine the state of the critical plant H by propagating through the GTST-DMLD the functional states at component level to the functional states at system-of-systems level. In doing so, the state of H is evaluated for each row of the matrix $\{g_{j,v}^\eta\}$, $j \in \{1, 2, 3\}$, $v = 1, \dots, N_T$, $\eta = 1, \dots, L$, i.e., for each configuration of the system sampled. A vector $\{h_v\}$ is then recorded, whose element h_v , $v = 1, \dots, N_T$, assumes value 1, 2 or 3 when the critical plant H is in a risk, marginal or healthy state, respectively;
8. estimate the probability of the critical plant H of being in a risk, marginal or healthy state by computing the sample average of the values of the elements of the N_T –dimensional vector $\{h_v\}$, $v = 1, \dots, N_T$;
9. for each v -th simulation of the system sampled that turns the critical plant H in an unsafe or marginal state, evaluate the recovery time (RT_H) by the following steps:
 - a. set the current time, t^{curr} , equal to zero in correspondence of the earthquake occurrence and initialize the counter q equal to 1;

- b. initialize the vectors of the time, t^H , and the functional state, z^H_i , of the critical plant H as $t^H(q) = t^{curr}$ and $\{z^H_i(q): i = h_v\}$, respectively;
- c. compute the number of aftershocks, n_{max}^{af} , that will occur with a magnitude higher than a given threshold, m_{min}^{af} , and lower than the maximum possible m_{max}^{af} (eq. B.5) by eq. B.1; sample their magnitude, m^{af} , from eq. B.2 and their time of occurrence from the cumulative distribution function of Figure B.1;
- d. sample a vector RT_η , $\eta = 1, \dots, L$, of recovery times of the components that are in state 1 or 2, from the respective probability density functions (PDFs) and set to infinite (i.e., a very large value) the recovery time of the components in state 3. If the component $\eta = 1, \dots, L$, is in state 1, it can reach both the state 2 and the state 3. In this case, sample the two recovery times and choose the lower. Save then a vector $g^{next}_j, j \in \{1, 2, 3\}$, $\eta = 1, \dots, L$, of structural states in which the components will enter if the recovery is carried out.
- e. While the critical plant H does not turn into a healthy state $\{z^H_i(q): i = 3\}$, perform the following steps:
 - i. evaluate the vector RT^{sum}_η , $\eta = 1, \dots, L$, that is equal to RT_η , $\eta = 1, \dots, L$, when the functional state of the road accesses to component η in state 1 and 2 is in a state 3, i.e., the accesses are available; whereas, it is the sum of the recovery times of the road accesses and of the component, when the road accesses are not available;
 - ii. identify the minimum recovery time, RT^{min} , of the vector RT^{sum}_η , $\eta = 1, \dots, L$;
 - iii. evaluate if aftershocks have occurred in the interval $t^{int} = [t^{curr}, t^{curr} + RT^{min}]$. If no, go to the following step iv.; otherwise, go to step v.;
 - iv. update the structural state vector $g^\eta_j, j \in \{1, 2, 3\}$, $\eta = 1, \dots, L$, for the component η that has performed the transition with the corresponding index j of the vector $g^{next}_j, j \in \{1, 2, 3\}$, $\eta = 1, \dots, L$. If the component η enters in a state 2, sample a new recovery time for η and update that value in the vector RT_η . For all other components, reduce the recovery time of the quantity equal to RT^{min} since the recovery of all the components proceeds at the same time. Then, update the functional state vector $\{z^\eta_i\}$, $i \in \{1, 2, 3\}$, $\eta = 1, \dots, L$, and evaluate the state of the critical plant H as in step 7., identifying the value h^{new} , $h^{new} \in \{1, 2, 3\}$. Set $q = q+1$, $t^H(q) = RT^{min}$ and $\{z^H_i(q): i = h^{new}\}$; Return to step e.
 - v. consider the first aftershock that occurs in the interval t^{int} and evaluate its impact on the structural states of the components η , $\eta = 1, \dots, L$, by steps 4. and 5. for the first row of the matrix u , i.e., for one simulation;
 - if the aftershock changes the state of one or more components, consider the new vectors of structural and functional state, $\{g^\eta_{j,v}\}$ and $\{z^\eta_{i,v}\}$, respectively, and update the vector RT_η sampling the

recovery time of the components η that have changed structural state. Update the vector g^{next}_{η} , $j \in \{1, 2, 3\}$, $\eta = 1, \dots, L$, with the new structural state in which the components will enter if their recovery is carried out. Set $q = q+1$, $t^H(q) = t^{af} - t^{curr}$ and set $t^{curr} = t^{af}$. Return to step e.i.;

- otherwise, perform again step e.v., evaluating the impact of the following aftershock that occurs in the interval t^{int} ; if there are no other aftershocks in the interval t^{int} , the recovery of the component η associated with the minimum recovery time RT^{min} (step e.ii.) is carried out. Return to step e.iv.;
- f. if the critical plant H was in state 1 ($h_v = 1$), save the time needed to recover the safety from state 1 to state 2 ($RT_{H(1)}^{(1 \rightarrow 2)}$), from state 2 to state 3 ($RT_{H(1)}^{(2 \rightarrow 3)}$) and from state 1 to state 3 ($RT_{H(1)}^{(1 \rightarrow 3)}$); if the critical plant H was in state 2, save the time needed to recover the safety from state 2 to state 3 ($RT_{H(2)}^{(2 \rightarrow 3)}$);
 - g. repeat the steps 9.a. – 9.g. N_{RT} number of times (e.g., $N_{RT} = 4000$);
10. save the recovery time for all the configurations from states 1 and 2, and obtain the empirical probability density functions and corresponding cumulative distribution functions.

Paper V

Analysis of the robustness and recovery of critical infrastructures within a multi-state system-of-systems framework, in presence of epistemic uncertainty

E. Ferrario, N. Pedroni and E. Zio

Submitted to ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering.

Analysis of the robustness and recovery of critical infrastructures within a multi-state system-of-systems framework, in presence of epistemic uncertainty

E. Ferrario^a, N. Pedroni^a and E. Zio^{a,b}

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supelec, France

enrico.zio@ecp.fr, enrico.zio@supelec.fr

^bDepartment of Energy, Politecnico di Milano, Italy

enrico.zio@polimi.it

Abstract

In this paper, we look at the robustness and recovery of connected critical infrastructures (CIs) under a system-of-systems (SoS) framework taking into account i) the dependencies among the components of an individual critical infrastructure and the interdependencies among different CIs; ii) the variability in component performance, by a multi-state model; iii) the epistemic uncertainty in the probabilities of transitions between different components states and in the mean values of the holding times distributions, by means of intervals. We adopt the Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) for system modelling and perform the quantitative assessment by Monte Carlo simulation. We illustrate the approach by way of a simplified case study consisting of two interdependent infrastructures (electric power system and gas network) and a supervisory control and data acquisition (SCADA) system connected to the gas network.

Keywords: critical infrastructures; electric power system, gas distribution network, SCADA, robustness; recovery time; multi-state; Goal Tree Success Tree – Dynamic Master Logic Diagram; Monte Carlo simulation; epistemic uncertainty; imprecise probability; interval analysis

1. INTRODUCTION

Critical infrastructures (CIs), e.g., transportation, electric power, water, gas, communication systems, interact on the basis of complex relationships that cross the single infrastructure boundary. This exposes to the risk that a failure in an infrastructure can have negative impacts on another interconnected one. For example, CIs are getting more and more dependent on information technologies that, on one hand, provide control and support their increasing efficiency, but, on the other hand, create new vulnerabilities [Nozick et al., 2005]. As additional example from the field, the widespread power electric blackout that occurred in the Midwest and Northeast of the United States and Ontario, Canada, on August 2003, affected the serviceability of the water system at Cleveland, OH, due to the lack of power needed to operate the water pumping stations [Adachi and Ellingwood, 2008]. Analyzing and understanding the interdependences existing among infrastructure systems is fundamental for the safe operation and control of these “systems of systems”.

Then, we adopt a system-of-systems (SoS) framework of analysis to evaluate the SoS robustness and recovery properties, considering the dependencies among the components of a critical infrastructure and the interdependencies among different CIs. For a more realistic representation, we utilize a multi-state model for consideration of the different degrees of damage that the individual components may experience [Ferrario and Zio, 2014]. Transitions between different states of damage occur stochastically (aleatory uncertainty) and epistemic uncertainty affects the associated transition probabilities due to insufficient knowledge and information on the components degradation behavior [Apostolakis, 1990; USNRC, 2009; NASA, 2010]. Indeed, safety-CIs are highly reliable and, thus, undergo few degradations to failure, so that it is difficult to estimate damage levels and transition probabilities [de Finetti, 1974; Bernardo and Smith, 1994; Coolen and Utkin, 2007; Aven and Zio, 2011; Sallak et al., 2013].

For illustration purpose, we adopt the framework of analysis to a case study proposed in [Nozick et al., 2005], in which the system considered consists of two interdependent infrastructures (gas and electric power networks), and a supervisory control and data acquisition (SCADA) system connected to the gas network. To measure the robustness and recovery capacity of the system, we look at the steady-state probability distributions of the supply of gas and electricity at the demand nodes and the time needed to recover the SoS from the worst scenario to a level in which all the demand nodes are satisfied, respectively.

We propose a hierarchical model description of the system logic and functionality by Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) [Hu and Modarres, 1999], extending its representation characteristics to evaluate the physical flows of gas and electricity through the interdependent infrastructures. We adopt intervals to describe the epistemic uncertainty in the probabilities of transition between different components states and in the mean values of the holding time distributions [Kuznetsov, 1991; Walley, 1991;

Kozine and Utkin, 2002; Lindley, 2006; Beer and Ferson, 2013; Beer et al., 2013; Blockley, 2013; Crespo et al., 2013; Jalal-Kamali and Kreinovich, 2013; Mehl, 2013] and we use interval analysis to calculate the (uncertain) probabilities of the states of all the components of the CIs [Ferson and Ginzburg, 1996; Buckley, 2004; Ferson and Hajagos, 2004; Ferson and Tucker, 2006; Ferson et al., 2007; Ferson et al., 2010]. Finally, we employ Monte Carlo simulation [Kalos and Whitlock, 1986; Zio, 2013] for the probabilistic evaluation of the SoS performance.

The reminder of the paper is organized as follows. In Section 2, the case study is presented; in Section 3, the SoS modelling by GTST-DMLD is illustrated; in Section 4, details of the procedural steps to evaluate the SoS performance under epistemic uncertainty are given; in Section 5, the results of the analysis are shown and commented; in Section 6, conclusions are provided. Finally in the Appendix a brief overview of imprecise probabilities is given.

2. CASE STUDY

The case study is taken from [Nozick et al., 2005] and it deals with two interconnected infrastructures, i.e., a natural gas distribution network and an electricity generation/distribution network (Figure 1, solid and dashed lines, respectively). The gas distribution network is supported by a SCADA system (Figure 1, dotted lines). The objective of this interconnected system of systems (SoS) is to provide the necessary amount of gas and electricity (hereafter also called “product”) to four demand nodes (end-nodes), namely D1 and D2 (gas) and L1 and L2 (electricity).

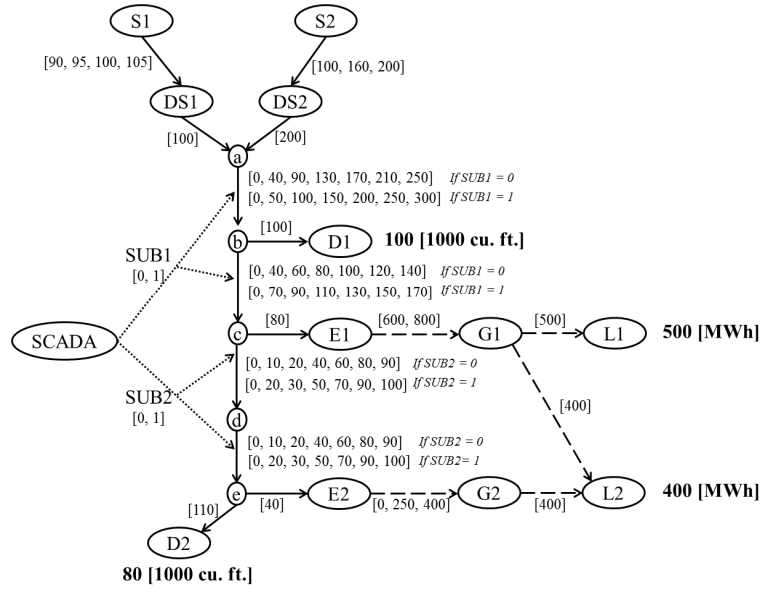


Figure 1: Interdependent gas (solid lines) and electric (dashed lines) infrastructures and SCADA system (dotted lines) [Nozick et al., 2005]. The possible states of the arcs are given in square brackets; the quantities demanded by the end-nodes D1, D2, L1, L2 are reported in bold.

The gas distribution network, supplied by two sources of gas (namely, S1 and S2, that are connected to the network by arcs S1_DS1 and S2_DS2, respectively), provides gas to the end-nodes D1 and D2 and to two nodes of the electricity network (E1 and E2). Once the gas enters into nodes E1 and E2, it is transformed into electrical energy that flows through arcs E1_G1 and E2_G2 (representing the electric power generation stations) to supply the end-nodes of electricity (L1 and L2); notice that the demand L2 can be supplied by both electrical generations E1_G1 and E2_G2. The assumption is made that the gas-electricity transformation occurs with a constant coefficient, i.e., 100 cu. ft. of natural gas produces 1 MWh of electricity [Nozick et al., 2005].

A SCADA system controls the gas flow through arcs a_b, b_c, c_d and d_e. It is assumed that: i) the SCADA has two core subsystems controlling different sets of arcs (in particular, the first one – SUB1 – refers to links a_b and b_c, whereas the second one – SUB2 – controls arcs c_d and d_e); ii) the SCADA is always provided with electric power [Nozick et al., 2005].

The capacities of the arcs of the gas and electricity networks (determining the maximum flows of gas or electricity supported by the arcs) can be deterministic (i.e., fixed constant values) or stochastic (i.e., randomly evolving in time) (Figure 1, values in the square brackets). The stochastic capacities give rise to a multi-state model that reflects the possibly different degrees of damage of the arcs. On the contrary, the SCADA system state is defined by a binary random variable, whose values 1 and 0 represent its complete and partial functioning, respectively. For example, when the state of the SCADA subsystem SUB1 (controlling arcs a_b and b_c) is 0, the capacity of these arcs decreases because of the incorrect information provided by the SCADA subsystem (even if the arcs are not subject to a

direct damage). On the basis of the two states of the SCADA subsystems, two different vectors of capacities are identified for each arc a_b, b_c, c_d and d_e: as illustrated in Figure 1, the first vector is used when the corresponding SCADA subsystem is in state 0, whereas the second one is employed when the SCADA subsystem is in state 1.

In the following, we generically denote the value of the state of a component (i.e., the capacity of the arcs) as $\zeta_{comp,i}$, $i \in \{1, 2, \dots, S_{comp}\}$, where the subscript ‘comp’ indicates the component of interest and i the state number (when $i = 1$, the component is in the worst state, whereas when $i = S_{comp}$, it is in the best state); S_{comp} is the total number of states for that component. For example, component S1_DS1 has $S_{S1_DS1} = 4$ possible states: $\zeta_{S1_DS1,1} = 90$ [1000 cu. ft.], $\zeta_{S1_DS1,2} = 95$ [1000 cu. ft.], $\zeta_{S1_DS1,3} = 100$ [1000 cu. ft.], $\zeta_{S1_DS1,4} = 105$ [1000 cu. ft.]. The total number of components in the SoS is referred to as N_{comp} .

Changes in the arc capacities are due to random failures or recovery actions. The state transitions over time are modeled by Markov and semi-Markov processes as in [Nozick et al., 2005]. Semi-Markov processes are adopted to represent the evolution of the capacities of the gas supply links (S1_DS1 and S2_DS2), whereas Markov processes are used for all the others arcs. Both Markov and semi-Markov processes for a generic component ‘comp’ are defined by a transition probability matrix $\underline{P}_{comp} = \{p_{ij} : i, j = 1, 2, \dots, S_{comp}\}$, where p_{ij} is the one-step probability of transition from state i to state j . In addition, the semi-Markov processes are characterized by continuous probability distributions for the holding time T_{comp}^{ij} , i.e., for the time of residence in state i before performing a transition to state j . The total number of components in the SoS described by the semi-Markov processes is referred to as N_{compSM} .

Differently from [Nozick et al., 2005], we take into account the epistemic uncertainty affecting the transition probabilities and the holding time distributions of the Markov and semi-Markov processes, respectively. In particular, intervals $[\underline{p}_{ij}, \bar{p}_{ij}]$, $i, j = 1, \dots, S_{comp}$, (instead of fixed constant values) are used to describe the state transition probabilities for both Markov and semi-Markov processes (matrices \underline{P}_{comp} , $comp = S1_DS1, S2_DS2, a_b, b_c, c_d, d_e, SCADA, E1_G1$ and $E2_G2$, in Figure 2 with respect to the states defined in Figure 1) [MSSP, 2013; Muscolino and Sofi, 2013; Pannier et al., 2013; Reid, 2013; Sankararaman and Mahadevan, 2013; Zhang et al., 2013]. The holding time distributions for the components modeled by the semi-Markov processes are considered normal with epistemically-uncertain mean (described by an interval) and fixed standard deviation (matrices \underline{T}_{comp} , $comp = S1_DS1, S2_DS2$, in Figure 2); this level-2 hierarchical representation produces a family of Normal probability distributions characterized by the same standard deviation, but different mean values: such a bundle of distributions is often referred to as distributional probability-box (p-box) [Moller et al., 2003; Ferson, 2005; Karanki et al., 2009; Limbourg and de Rocquigny, 2010; Pedroni and Zio, 2012; Pedroni et al., 2013].

In the present work, the demand nodes are not given the same importance: in particular, D1 is more important than L1; on its turn, L1 is more important than both D2 and L2 (which instead

are equally important). These assumptions are made to illustrate and motivate the logical repartition of electricity and gas flows in the network and its representation in the GTST-DMLD given in the next Section 3.

The objectives of the analysis are to determine the cumulative distribution functions of i) the product delivered to the demand nodes (i.e., D1, D2, L1, L2) at the steady state and ii) the time needed to recover the SoS from the worst scenario. Since the state transition probabilities of the network components are affected by epistemic uncertainty and are described by intervals, $[\underline{p}_{ij}, \bar{p}_{ij}]$, $i, j = 1, \dots, S_{comp}$, the corresponding component steady-state probabilities are also affected by epistemic uncertainty and represented by intervals of possible values, $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, \dots, S_{comp}$. As a consequence, a set of cumulative distribution functions corresponding to the set of possible steady-state probabilities within the intervals $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, \dots, S_{comp}$, is obtained for each demand node. For the same reason (i.e., for the presence of the epistemic uncertainty in the state transition probabilities and in the mean of the components holding time distributions) a set of cumulative distribution functions for the recovery time of the system is obtained in correspondence of the set of possible state transition probabilities.

comp = S1_DS1 (Semi-Markov)					comp = S2_DS2 (Semi-Markov)				
$\underline{P}_{comp} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ [0.002; 0.008] & [0.002; 0.008] & 0 & [0.998; 1] \\ [0.002; 0.008] & [0.002; 0.008] & [0.998; 1] & 0 \end{bmatrix}$					$\underline{P}_{comp} = \begin{bmatrix} 0 & 1 & 0 \\ [0; 0.02] & 0 & [0.98; 1] \\ [0; 0.02] & [0.98; 1] & 0 \end{bmatrix}$				
$\underline{T}_{comp} = \begin{bmatrix} - & N([2; 6], 1) & - & - \\ - & - & N([2; 6], 1) & - \\ N([7; 13], 3) & N([7; 13], 3) & - & N([17; 23], 2) \\ N([7; 13], 3) & N([7; 13], 3) & N([17; 23], 2) & - \end{bmatrix}$					$\underline{T}_{comp} = \begin{bmatrix} - & N([2; 6], 1) & - \\ N([7; 13], 3) & - & N([2; 6], 1) \\ N([7; 13], 3) & N([17; 23], 2) & - \end{bmatrix}$				
comp = a_b, b_c, c_d, d_e (Markov)									
$\underline{P}_{comp} = \begin{bmatrix} [0.04; 0.06] & [0.04; 0.06] & [0.04; 0.06] & [0.04; 0.06] & [0.1; 0.3] & [0.1; 0.3] & [0.3; 0.5] \\ [0.04; 0.06] & [0.04; 0.06] & [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.1; 0.3] & [0.3; 0.5] \\ [0.04; 0.06] & [0.04; 0.06] & [0.04; 0.06] & [0.1; 0.3] & [0.1; 0.3] & [0.04; 0.06] & [0.3; 0.5] \\ [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.04; 0.06] & [0.3; 0.5] \\ [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.04; 0.06] & [0.3; 0.5] \\ [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.1; 0.3] & [0.04; 0.06] & [0.04; 0.06] & [0.3; 0.5] \\ [0.0005; 0.0015] & [0.0005; 0.0015] & [0.0005; 0.0015] & [0.0015; 0.0025] & [0.0015; 0.0025] & [0.002; 0.004] & [0.985; 0.995] \end{bmatrix}$									
comp = SCADA (Markov) (states of SUB1 and SUB2: 0 0, 0 1, 1 0, 1 1)					comp = E2_G2 (Markov)				
$\underline{P}_{comp} = \begin{bmatrix} [0.7; 0.9] & [0.03; 0.05] & [0.03; 0.05] & [0.02; 0.22] \\ [0.05; 0.15] & [0.3; 0.5] & [0.2; 0.4] & [0.1; 0.3] \\ [0.05; 0.15] & [0.2; 0.4] & [0.3; 0.5] & [0.1; 0.3] \\ [0.0005; 0.0007] & [0.0001; 0.0003] & [0.0001; 0.0003] & [0.998; 1] \end{bmatrix}$					$\underline{P}_{comp} = \begin{bmatrix} [0.1; 0.3] & [0.05; 0.15] & [0.6; 0.8] \\ 0 & [0.1; 0.3] & [0.7; 0.9] \\ [0.0004; 0.0006] & [0.0004; 0.0006] & [0.998; 1] \end{bmatrix}$				
					comp = E1_G1 (Markov)				
					$\underline{P}_{comp} = \begin{bmatrix} [0.05; 0.15] & [0.89; 0.91] \\ [0; 0.002] & [0.998; 1] \end{bmatrix}$				

Figure 2: Holding time distributions (matrices \underline{T}_{comp}) for the arcs described by semi-Markov processes: each element of the matrix represents a Normal distribution with uncertain (interval) mean and fixed standard deviation. State transition probability matrices (\underline{P}_{comp}) for the arcs described by Markov and semi-Markov processes: each element of the matrix represents an interval for the corresponding transition probability.

3. SYSTEM-OF-SYSTEMS MODELLING

3.1. Goal Tree Success Tree – Dynamic Master Logic Diagram: basic concepts

The Goal Tree Success Tree – Dynamic Master Logic Diagram (GTST-DMLD) is a goal-oriented method based on a hierarchical framework [Hu and Modarres, 1999]. It gives a comprehensive description of the systems in terms of functions (qualities), objects (parts) and their relationships (interactions). The first description is provided by the Goal Tree (GT), the second by the Success Tree (ST) and the third by the DMLD [Hu and Modarres, 1999].

The GT identifies the hierarchy of the qualities of the system composing the objective of the analysis, i.e., the goal, organizing them in functions that are in turn subdivided into other functions and so on. The hierarchy is built by answering questions on “how” the subfunctions can attain the parent functions (looking at the hierarchy from top to bottom) and on “why” the functions are needed (looking at the hierarchy from bottom to top). Two types of qualities, i.e., main and support functions, are considered: the former directly contribute to achieving the goal, whereas, the latter support the realization of the former [Brissaud et al., 2011].

The ST represents the hierarchy of the objects of the system, from the entire system to the parts necessary to attain the last levels of the GT. This hierarchy is built identifying the elements that are “part of” the parent objects. As for the GT, two types of objects are distinguished also in the ST: main and support. The former are directly contributing to achievement of the main functions, whereas the latter are needed for the operation of the former [Brissaud et al., 2011].

The DMLD is an extension of the Master Logic Diagram (MLD) [Hu and Modarres, 1999] introduced to model the dynamic behavior of a physical system. It describes the interactions between parts, functions and parts and functions, in the form of a dependency matrix, and it includes the dynamics by means of time-dependent fuzzy logic rules [Hu and Modarres, 1999].

A conceptual sketch of GTST-DMLD is given in Figure 3.

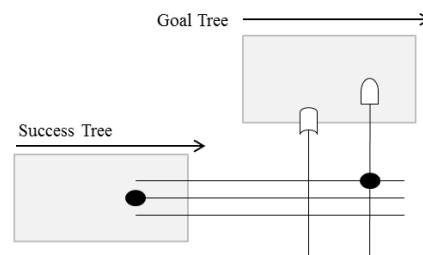


Figure 3: Conceptual sketch of GTST-DMLD: the filled dots indicate the possible dependencies between the objects (filled dot on the left) and between the objects and functions (filled dot on the right), the logic gates indicate how a given function depends on the input values.

The GT is drawn at the top, the ST tree on the left and the DMLD is represented by filled dots at the intersections between vertical and horizontal lines, to indicate the possible dependencies between the elements on the left and on the top. Several types of logic gates can be used to represent the time-dependent fuzzy logic rules, and different dependency-matrix

nodes to describe the probabilities and degrees of truth in the relationships [Hu and Modarres, 1999]. Figure 4 gives an example of dependency of an element C on two elements A and B by the “AND” gate in a DMLD [Hu and Modarres, 1999]. In this case, the output value of the element C is the minimum value between the inputs A and B. Replacing the “AND” gate with an “OR” gate, the output value will be the maximum between the input values.

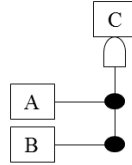


Figure 4: Example of an element C that depends on two elements A and B by an “AND” gate.

Further details on the GTST-DMLD modeling are not given here for brevity sake: the interested reader is referred to the cited literature [Hu and Modarres, 1999; Brissaud et al., 2011]. In the next Section 3.2, the adaptation of the GTST-DMLD for modeling interconnected networked infrastructures is illustrated.

3.2. Goal Tree Success Tree – Dynamic Master Logic Diagram for interconnected networked infrastructures

In this Section, we adapt the GTST-DMLD presented in Section 3.1, in general terms, for an adequate representation of interconnected networked infrastructures, and in particular of the ones making the SoS of our case study of Section 2. Specifically, we introduce new concepts in order to model in the diagram not only the dependency relations between the components, but also the ways in which the flows of gas and electricity are partitioned into the network on the basis of i) the importance of the demand nodes, ii) the amount of product necessary to satisfy each demand, iii) the constraints of the arc capacities and iv) the information provided by the SCADA system. In the following, first we explain the notation adopted in the GTST-DMLD and, then, we apply it to the case study of interest.

In the present work, we distinguish between three main types of dependency: *direct*, *indirect* and *constraint-based* dependencies, as illustrated in Figures 5 and 6. The former, pictorially represented by a dot and hereafter called “*dot-dependencies*”, express the fact that the product of the element on the bottom of the dot passes straightly into the element on the top. The indirect dependencies, represented by a hexagon and called hereafter “*hexagon-dependencies*”, are instead needed for the optimal allocation of the product in the network: for example, they are used to describe those cases where the flow exceedance in an arc can be better partitioned into another arc that is not directly connected to it but that shares one of the inputs (see the example of Figure 5 b). Finally, the constraint-based dependencies, depicted by a triangle and hereafter called “*triangle-dependencies*”, are employed to take into account some physical constraints posed by the problem, like the maximum flow required by a demand node.

It is worth mentioning that since in the present case we are interested in analyzing the flows

passing through the network (and not just the dependency relations), the inputs of an arc are flows and the output is (generally) the sum of the flow inputs. For this reason, in this context the “AND” gate assumes a different meaning with respect [Hu and Modarres, 1999] (see the previous Section 3.1): in particular, the output value is the sum of the input values and it is represented by a “+” in the middle of the gate, as shown in the following examples (Figures 5 and 6).

For clarity of illustration, in Figure 5, examples of two types of dot- and hexagon-dependencies are given, with respect to different graph representations. Figure 5 a. shows the dependence of arc C on two input arcs A and B: arc C receives all the input products from A and B (e.g., if the flows in arcs A and B are 50 and 70 units, respectively, the flow in arc C is 120 units); this complete direct dependence is depicted by a *black dot*. Figures 5 b. and c. describe the same "physical" situation (i.e., an input arc A and two output arcs B and C), but with different relative importance of the arcs. Two different cases are illustrated. In the first case (Figure 5 b.) arc B is more important than C: thus, in this situation, the flow from A supplies first arc B until its demand is satisfied, and then arc C, e.g., if the flows in arc A is 100 units and both arcs B and C need 80 units, arc B will receive 80 units – demand fully satisfied – and arc C the rest, i.e., 20 units, – demand partially satisfied. In the second case (Figure 5 c.), arcs B and C are equally important: thus, the input flow (A) is divided into equal parts on the basis of the number of output arcs (i.e., two in this example); with respect to the numeric example above, both arcs B and C will receive 50 units – demands partially satisfied. In the case of Figure 5 b., the flow that enters in C is given by the difference between the entire flow from A and the flow given to B; to represent and compute this difference in the DMLD, the hexagon-dependency is adopted to correct the black dot-dependency from arc A to arc C (in fact, it is impossible that the entire flow of A enters at the same time in the arcs B and C, as expressed by the black dot-dependency). The *white hexagon* assumes the value of the flow in B with a negative sign; this value is, then, summed to the initial flow of A to obtain the flux to C. The flow given to B can be the entire flow of A or a lower value depending on the constraints and arc capacity (see the following example in Figure 6). In the case of Figure 5 c., the flow from A is divided into equal parts: this condition is represented by a *grey dot*. However, this equal partition of the flow may not represent the optimal one, since some output arcs may require less flow than the one allocated according to this criterion, e.g., if the flows in arc A is 100 units and arcs B and C need 80 and 20 units, respectively, giving 50 units to both arcs is not a good allocation of the resource since B is partially satisfied and some product (i.e., 30 units) given to arc C is wasted. Thus, to optimize the repartition of the flow, hexagon-dependencies are adopted: they are directed from an output arc to all the other output arcs that share the same input. In this case, the “surplus flow” is a positive quantity and it is represented by a *grey hexagon* (to distinguish it from the “negative” white hexagon of the example in Figure 5 b).

Notice that the graph representation of Figures 5 b. and 5 c. are identical; however, the partition of the flux from A is completely different in the two cases: this means that the graph

representation alone cannot be used to describe the repartition of the flows in the network according to different criteria. On the contrary, the DMLD can capture and represent this aspect, which is useful in the quantitative evaluation of system performance.

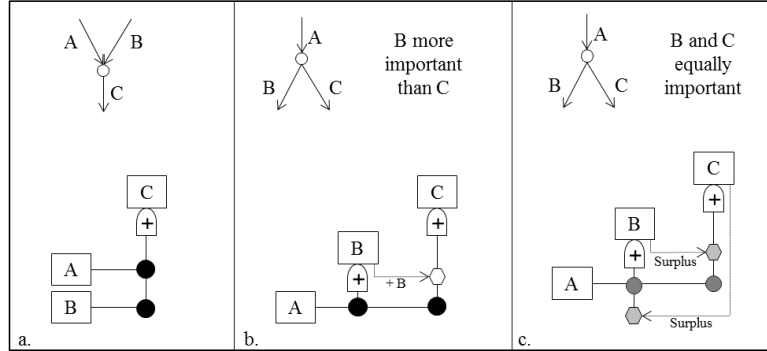


Figure 5: Examples of dot- and hexagon-dependencies with respect to possible graph representations.

In Figure 6, examples of two types of triangle-dependencies are given, with respect to different possible graph representations. Figure 6 a. depicts the same situation as Figure 5 a., with an additional arc D whose behavior impacts on the state of arc C (however, notice that D is not an input to C). This dependency is represented by a grey triangle and it means that the output of C can be modified on the basis of the state of arc D. In the present case study, this constraint-based dependency is used to model the SCADA system that can decrease the actual flow of the controlled arc if it is in a damage state. Figure 6 b. represents the same situation of Figure 5 c. with the addition of another arc (D) sequential to arc C. In this case, the capacity (or the demand) of arc D can limit the amount of flow in input to arc C, e.g., if the flows in arc A is 100 units, the capacity of arc C is 50 units and arcs B and D need 80 and 20 units, respectively, the repartition of the flow is as follows: first 100 units from A are equally divided into arcs B and C (50 units each) and the surplus (if there is) is partitioned into arcs B and C, then the triangle constraints is considered (i.e., arc D needs 20 units) and the new surplus is given to arc B (i.e., the exceedance of 30 units from arc C is directed to arc B). This constraint is represented in the DMLD by a black triangle and it is needed to control the input flow partitioned in different arcs and guarantee that it is not higher than necessary.

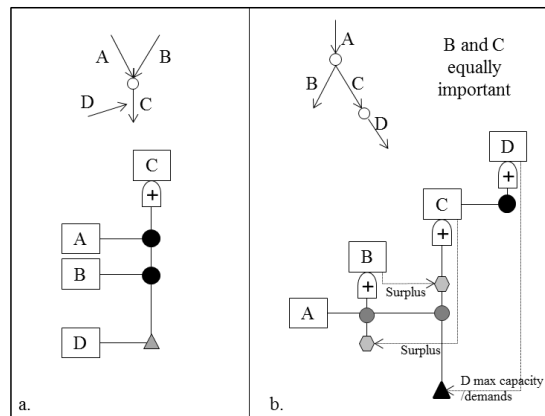


Figure 6: Examples of triangle-dependencies with respect to possible graph representations.

Finally, another type of constraint is taken into account, i.e., the one related to the capacity of the arcs: when the flow in input to an arc is higher than the capacity of the arc itself, the output flow will be equal to the capacity of the arc. The arc capacity can be deterministic or stochastic and in the GTST-DMLD it is represented by a grey or dot-filled rectangular, respectively (see Figure 7).

In Figure 7, the GTST-DMLD of the case study of Section 2 is shown.

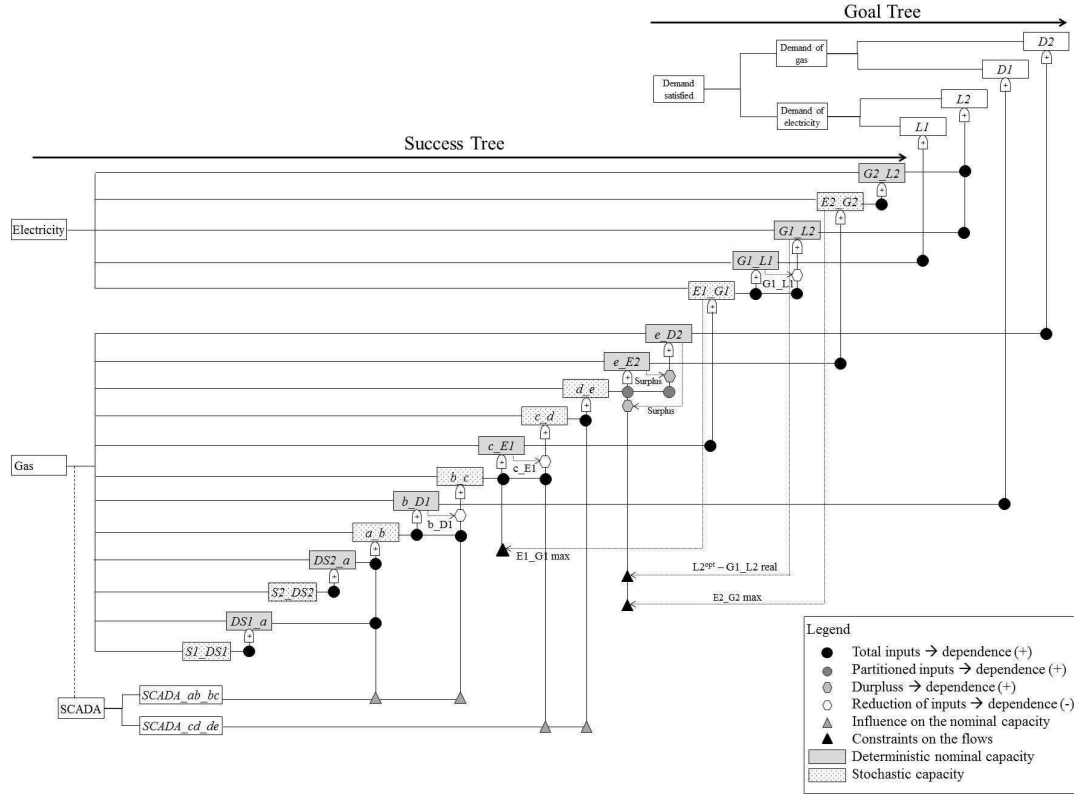


Figure 7: GTST-DMLD of the case study of Section 2 corresponding to the graph of Figure 1.

The GT on the top represents the main goal of the system of systems (SoS), related to the supply of the demands of gas and electricity: the objective is achieved if the corresponding nodes D1, D2, L1 and L2 receive the required amount of gas and electricity, respectively. In the present case study, we limit the analysis to the last level of the GT, i.e., we analyze the performance of each demand, without investigating a global indicator of the SoS.

The ST is composed by the main hierarchies of the gas and electricity networks (that directly provide the demand nodes with gas and electricity to achieve the goal function) and by the support hierarchy of the SCADA system (that is needed for the control of the gas network and, therefore, it is not directly involved in the achievement of the goal function); given its support role, it is represented in a parallel dashed branch connected to the gas hierarchy.

The DMLD is represented by the relationships between objects of the ST or between objects of the ST and functions of the GT. It allows determining the goal function by the evaluation of all the dependencies from the bottom to the top of the diagram, following the rules

explained above for the dot-, hexagon- and triangle- dependencies. For example, arc a_b depends on two arcs, $DS1_a$ and $DS2_b$, connected by black dot-dependencies (Figure 7). Thus, the output of a_b is given by the sum of the corresponding input values, i.e., $DS1_a + DS2_b$. This value may, then, be modified by the triangle constraint of the SCADA system and by the (stochastic) capacity of arc a_b itself.

4. EVALUATION OF THE SYSTEM-OF-SYSTEMS PERFORMANCE

In this Section, we illustrate the evaluation of the performance of the system of systems (SoS) described in Section 2, in the presence of epistemic uncertainties (represented by intervals) affecting the components state transition probabilities and the mean values of the holding time distributions. As already mentioned in Section 2, the system performance is quantified in terms of i) robustness, measured by the steady-state probability distributions of the product delivered at the demand nodes (see Section 4.1) and ii) recovery capacity, measured by the time needed to recover the SoS from the worst scenario (see Section 4.2).

4.1. Robustness

To compute the steady-state probability distributions of the product delivered at the demand nodes the following three main steps are carried out:

1. Processing the epistemic uncertainties by interval analysis: this step leads to the evaluation of the intervals of the steady-state probabilities, $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, 2, \dots, S_{comp}$, for the states of each component ($comp = 1, 2, \dots, N_{comp}$) of the SoS.
2. Evaluation of the SoS performance (i.e, robustness) by Monte Carlo simulation: this step leads to the determination of a set of cumulative distribution functions (CDFs) of the product delivered at each demand node at steady state, one for each possible combination of steady-state probabilities ranging within the intervals $[\Pi_{\min}^{comp,i}, \Pi_{\max}^{comp,i}]$, $i = 1, 2, \dots, S_{comp}$, (found at step 1. above).
3. Post-processing the results obtained at the previous step 2: this step leads to the identification of two extreme upper and lower CDFs that bound the set of CDFs produced at step 2. above.

In more details:

1. Solve the following optimization problems for the lower (resp., upper) bounds $\Pi_{\min}^{comp,i}$ (resp., $\Pi_{\max}^{comp,i}$), $i = 1, 2, \dots, S_{comp}$, for all the N_{comp} components of the SoS:

$$\Pi_{\min}^{comp,i} = \min_{p_{ij}, i, j=1,2,\dots,S_{comp}} \{\Pi^{comp,i}\}, \quad \forall i=1,2,\dots,S_{comp}, \quad comp = 1, 2, \dots, N_{comp} \quad (1)$$

$$\Pi_{\max}^{comp,i} = \max_{p_{ij}, i, j=1,2,\dots,S_{comp}} \{\Pi^{comp,i}\}, \quad \forall i=1,2,\dots,S_{comp}, \quad comp = 1, 2, \dots, N_{comp}$$

such that:

$$p_{ij} \in [\underline{p}_{ij}, \bar{p}_{ij}] \quad (2)$$

$$\sum_{j=1}^{S_{comp}} p_{ij} = 1 \quad (3)$$

$$\underline{\Pi}^{comp} = \underline{\Pi}^{comp} \cdot \underline{P}_{=comp} \quad (4)$$

The constraint of eq. (2) means that the transition probability from state i to state j is not known precisely and can take values in the interval of probabilities $[\underline{p}_{ij}, \bar{p}_{ij}]$ [Buckley, 2004]; the constraint of eq. (3) refers to a fundamental property of Markov and semi-Markov processes, i.e., that the states for each component are assumed exhaustive [Zio, 2009]; finally, eq. (4) reports the definition of steady-state probability for a Markov process [Zio, 2009]. In the case of a semi-Markov process, eq. (4) is weighted by the expected time of residence, τ^i , in a given state, i , before performing a transition [Barry, 1995]: $\xi^{comp,i} = \Pi^{comp,i} \cdot \tau^i / \sum_{j=1}^{S_{comp}} \Pi^{comp,j} \cdot \tau^j$ for $i = 1, \dots, S_{comp}$.

Notice that the optimization problems (1) can be solved by performing an exhaustive greedy search within the probability intervals $[\underline{p}_{ij}, \bar{p}_{ij}]$, if the dimensions of the corresponding transition probability matrices are relatively small (e.g., below 4 x 4), otherwise, alternative intelligent techniques should be sought, e.g., meta-heuristic methods like Genetic Algorithms (GAs) [Buckley, 2004]. In this work, we resort to GAs for arcs a_b, b_c, c_d, d_e (whose transition probability matrices are 7 x 7), whereas we perform an exhaustive search for all the other arcs.

2. Identify the CDFs of the product delivered at each demand node at steady state for all the possible combinations of components steady-state probabilities found at step 1. above:
 - a. For each component $comp$, let the steady-state probabilities, $\Pi^{comp,i}$, $i = 1, 2, \dots, S_{comp}$, range within the corresponding interval $[\Pi_{min}^{comp,i}, \Pi_{max}^{comp,i}]$, $i = 1, 2, \dots, S_{comp}$, to obtain a set of Q_{comp} vectors of steady-state probabilities, $\{\underline{\Pi}^{comp,1}, \underline{\Pi}^{comp,2}, \dots, \underline{\Pi}^{comp,q}, \dots, \underline{\Pi}^{comp,Q_{comp}} : q = 1, \dots, Q_{comp}\}$, such that $\sum_{i=1}^{S_{comp}} \Pi^{comp,q,i} = 1$, $q = 1, \dots, Q_{comp}$. Notice that this gives rise to $\prod_{comp=1}^{N_{comp}} Q_{comp} = N_{tot}$ possible combinations of steady-state probability vectors of the system components, i.e., to N_{tot} steady-state probability vectors for the entire system.
 - b. For all the N_{comp} components, select one steady-state probability vector among the set $\underline{\Pi}^{comp,q}$, $q \in \{1, \dots, Q_{comp}\}$ (generated at step a. above); in other words, this amounts to selecting one of the $N_{tot} = \prod_{comp=1}^{N_{comp}} Q_{comp}$ steady-state probability vectors for the entire SoS.
 - c. Fixing the SoS steady-state probability vector selected in b., randomly sample the states $\zeta_{comp,i}$ (i.e., the capacities), $i \in \{1, \dots, S_{comp}\}$, of all the components of the system (i.e., arcs). Then, compute the product delivered at the demand nodes propagating the flow in each component of the SoS through the GTST-DMLD (see Section 3.2).

- d. Repeat step c. a large number of times (e.g., 1000 in this work) and obtain the CDF for the product delivered at each demand node.
- e. Repeat steps c.-d. for another combination of the steady-state probability vectors, $\underline{\Pi}^{comp,q}$, $q \in \{1, \dots, Q_{comp}\}$, of all the N_{comp} components, until all the N_{tot} possible combinations of the steady-state probability vectors of the SoS are explored.

At the end of steps a.-e., an ensemble of CDFs for each demand nodes is obtained, one for each of the N_{tot} possible combinations of steady-state probabilities of the entire SoS.

3. Identify the extreme minimum and maximum CDFs of the product delivered at the demand nodes that bound the set of CDFs produced at step 2. above.

4.2. Recovery time

The time needed to recover the SoS from the worst scenario (i.e., the one characterized by components in the worst state) to a level in which all the demand nodes are satisfied, is carried out by three main steps:

1. Processing the epistemic uncertainties by interval analysis: this step leads to the identification of K_{comp} transition probability matrices \underline{P}_{comp}^k , $k = 1, 2, \dots, K_{comp}$, for each component ($comp = 1, 2, \dots, N_{comp}$) of the SoS.
2. Evaluation of the SoS performance (i.e., recovery capacity) by Monte Carlo simulation: this step leads to the determination of a set of cumulative distribution functions (CDFs) of the time needed to recover the SoS, one for each possible combination of state probability matrices sampled.
3. Post-processing the results obtained at the previous step 2: this step leads to the identification of two extreme upper and lower CDFs that bound the set of CDFs produced at step 2. above.

In more details, step 1. is performed as follows:

- a. Select a component $comp$ and a row i of the matrix \underline{P}_{comp} and let the probability p_{ij} , $j = 1, 2, \dots, S_{comp}$, vary within the corresponding interval $[\underline{p}_{ij}, \bar{p}_{ij}]$, in order to identify $C_{comp,i}$ combinations of probabilities such that $\sum_{j=1}^{S_{comp}} p_{ij} = 1$ (by the assumption that the states are exhaustive, as for the previous eq. 3). If the component $comp$ is described by a semi-Markov process, select also a row i of the matrix \underline{T}_{comp} and let the mean, μ_{ij} , $j = 1, 2, \dots, S_{comp}$, of the holding time distributions vary within the corresponding interval $[\underline{\mu}_{ij}, \bar{\mu}_{ij}]$ to obtain $M_{comp,i}$ vectors of combinations of mean values for the row i . Repeat this step 1. a. for all the rows $i = 1, 2, \dots, S_{comp}$, of the matrices \underline{P}_{comp} and \underline{T}_{comp} .

- b. Combine the $\sum_{i=1}^{S_{comp}} C_{comp,i}$, vectors of probabilities for all the components ($comp = 1, 2, \dots, N_{comp}$) to obtain K_{comp} transition probability matrices $\underline{\underline{P}}_{comp}^k$, $k = 1, 2, \dots, K_{comp}$, for each component. If the component $comp$ is described by a semi-Markov process, combine also the $\sum_{i=1}^{S_{comp}} M_{comp,i}$ vectors of mean values to obtain H_{comp} matrices $\underline{\underline{Mu}}_{comp}^h$, $h = 1, 2, \dots, H_{comp}$, of the mean values of the holding time distribution.
- c. Repeat steps a.-b. for each component ($comp = 1, 2, \dots, N_{comp}$) of the SoS. All the N_{comp} components are, then, associated with a set of transition probabilities matrices $\underline{\underline{P}}_{comp}^k$, $k = 1, 2, \dots, K_{comp}$; in addition, those components described by a semi-Markov process (i.e., N_{compSM} components) are also associated with a set of matrices, $\underline{\underline{Mu}}_{comp}^h$, $h = 1, 2, \dots, H_{comp}$, containing the mean values of the corresponding holding time distributions.

Step 2. is carried out as follows:

- a. Randomly select N_{comp} matrices $\underline{\underline{P}}_k^{comp}$, $k \in \{1, 2, \dots, K_{comp}\}$, $comp = 1, 2, \dots, N_{comp}$, for all the components of the SoS and N_{compSM} matrices $\underline{\underline{Mu}}_h^{comp}$, $h \in \{1, 2, \dots, H_{comp}\}$ for the components described by a semi-Markov process.
- b. Set $u = 1$ (counter of the number of simulations).
- c. Initialize the state of the components at the worst state ($\zeta_{comp,i}$, $i = 1, comp = 1, 2, \dots, N_{comp}$): in this state configuration of the SoS, the product delivered to the demand nodes is lower than the optimum required.
- d. Initialize the following time variables: system simulation time $t = 0$, starting time of the simulation: this variable represent the current simulation time and is needed to compute the recovery time of the SoS; set $t^{comp} = \Delta t$, $comp = 1, 2, \dots, N_{comp}$, where Δt is the time step of the simulation ($\Delta t = 1$ in arbitrary units, in this work): these time variables are needed to determine if the component $comp$ can perform a state transition at a given time step (they are set to 1 since at this time step all the components perform the first state transition).
- e. Set $t = t + \Delta t$: if $t = t^{comp}$, then the component $comp$ performs a state transition: then, randomly sample its new state from the matrix $\underline{\underline{P}}_{comp}^k$ selected at step 2. a. and update the variable t^{comp} as follows:
 - ✓ If $comp$ is described by a Markov process, $t^{comp} = t^{comp} + \Delta t$, since a state transition occurs at each time step.
 - ✓ If $comp$ is described by a semi-Markov process, $t^{comp} = t^{comp} + t^*$, where t^* is the time of next transition that is sampled from the corresponding holding time

distribution with mean value taken from the matrix $\underline{\underline{\text{Mu}}}_h^{comp}$ selected at the previous step 2. a. The sampled value t^* is rounded to the nearest integer except when it is zero; in this case, the value is rounded to 1.

- f. Evaluate the product delivered to the demand nodes by adopting the GTST-DMLD (see Section 3.2).
- g. Repeat steps e.-f. until the product delivered to the demand nodes is equal to, or higher than, the optimum required: the corresponding value of recovery time (t_{RT}^u) is then recorded for the simulation u .
- h. Set $u = u + 1$ and repeat steps c.-g. a large number of times (e.g., 1000 in this work).
- i. A cumulative distribution function of the recovery time of the SoS is identified for a combination of state probability matrices $\underline{\underline{\text{P}}}_{comp}^k$, $k \in \{1, 2, \dots, K_{comp}\}$, selected at step 2. a.
- j. Repeat the entire procedure (steps a.-i.) a large number of times (e.g., 10000 in this work) to explore many different combinations of probability matrices $\underline{\underline{\text{P}}}_{comp}^k$, $k \in \{1, 2, \dots, K_{comp}\}$.

At the end of the procedure, a set of cumulative distribution functions of the recovery time of the performance of the SoS is obtained.

The results are processed at step 3., where the minimum and maximum CDFs of the recovery time that bound the set of CDFs obtained at step 2. above are identified and the 99th percentiles of the distributions are computed as a measure of the recovery time.

5. RESULTS

Figure 8 shows the lower (dotted line) and upper (solid line) cumulative distribution functions of the gas and the electricity delivered at steady state to the demand nodes D1, D2 and L1, L2, respectively, in steady state, obtained by the procedure illustrated in Section 4.1. Table 1 reports the corresponding (upper and lower) probabilities that the product delivered to the demand nodes, D1, D2, L1 and L2, exceeds the following threshold values: $d_1^* = 95$ [1000 cu. ft.], $d_2^* = 75$ [1000 cu. ft.], $l_1^* = 475$ [MWh] and $l_2^* = 375$ [MWh] (i.e., the probabilities that the corresponding demands are satisfied).

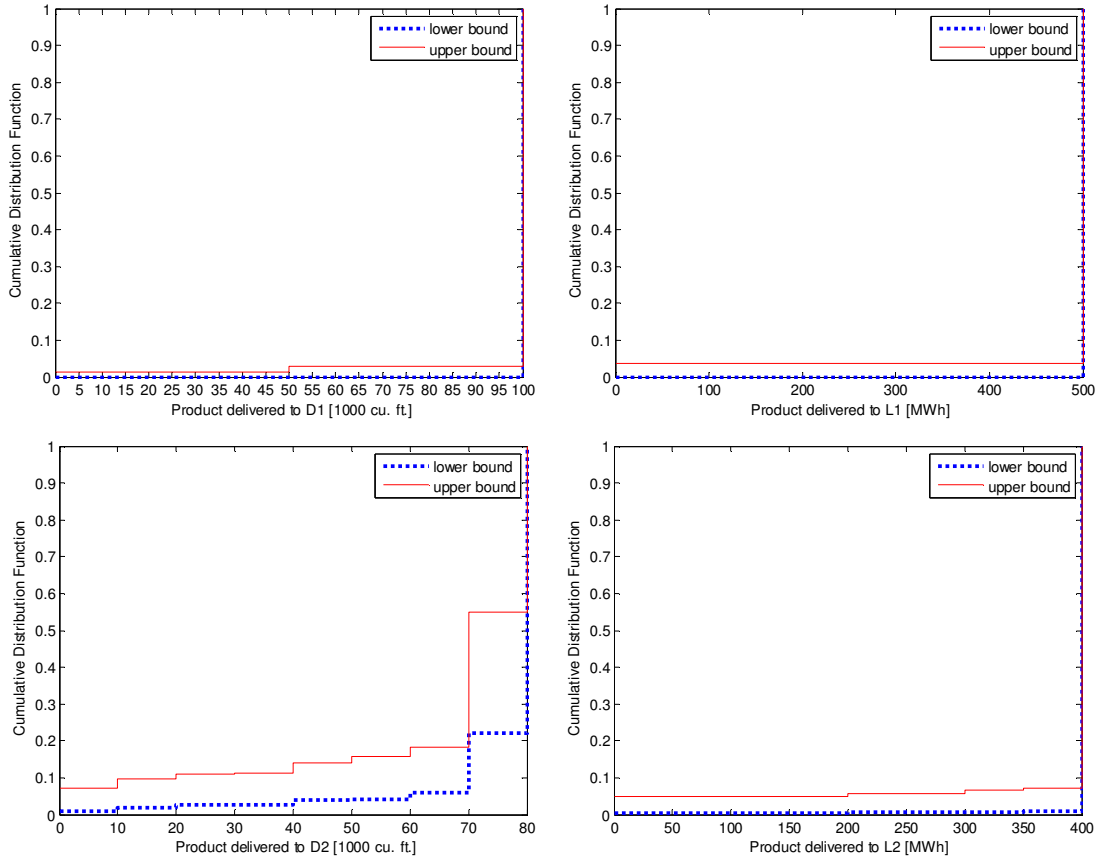


Figure 8: lower (dotted line) and upper (solid line) cumulative distribution functions of the product delivered the nodes D1, D2, L1 and L2 at steady state.

Table 1: upper and lower probabilities that the product delivered to the demand nodes (D1, D2, L1 and L2) exceeds the corresponding requested threshold value

$D1 \geq d_1^* = 95$ [1000 cu. ft.] [lower, upper]	$D2 \geq d_2^* = 75$ [1000 cu. ft.] [lower, upper]	$L1 \geq l_1^* = 475$ [MWh] [lower, upper]	$L2 \geq l_2^* = 375$ [MWh] [lower, upper]
[0.971, 1]	[0.450, 0.780]	[0.963, 1]	[0.929, 0.992]

It can be seen that in general the probability of satisfying demand nodes D1 and L1 is higher than for nodes D2 and L2: their threshold values are satisfied, in the worst case, with probability equal to 0.971 and 0.963, respectively. On the contrary, node D2 is the least supplied: the upper and lower probabilities that the product delivered to it exceeds the corresponding threshold value are low, i.e., 0.450 and 0.780, respectively. This is due to the fact that node D2 can be satisfied by only one path that presents high epistemic uncertainty in the arc capacities (a_b , b_c , c_d and d_e). On the contrary, node L2 is satisfied with probability between 0.929 and 0.992 even if it is the farthest node from the input sources (and, thus, more affected by uncertainty due to the uncertainties in the arc capacities): this is due to the presence of two redundant paths that allow its supply by arcs E1_G1 and E2_G2.

Figure 9 illustrates the lower (dotted line) and upper (solid line) cumulative distribution functions of the time needed to restore the SoS to a level in which all the demand nodes are satisfied, starting from the worst scenario.

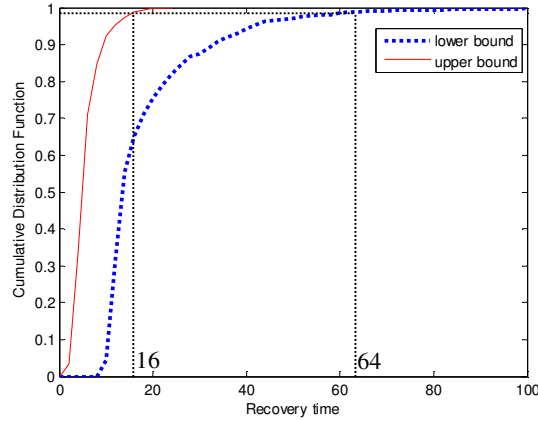


Figure 9: Upper and lower cumulative distribution functions of the recovery time of the supply of the demand nodes, starting from the worst scenario.

The gap between the CDFs reflects the epistemic uncertainty in the transition probability values. In the Figure, the 99th percentile of the CDFs is also reported as a measure of the recovery time.

6. CONCLUSIONS

In this paper, we have introduced a system-of-systems (SoS) framework for the analysis of the robustness and recovery of critical infrastructures (CIs). The analysis by such framework builds on the construction of a GTST-DMLD for system modeling and Monte Carlo simulation for the quantitative evaluation of the system performance at steady state. The development of the framework in practice has been shown considering two interdependent infrastructures, gas and electric power networks, and a SCADA system connected to the gas network.

The framework has shown the capability of representing, modeling and quantitatively accounting for i) the dependencies and interdependencies among the components of a critical infrastructure and between different CIs, respectively, ii) the variability in the states of the components (by adopting a multistate model), and iii) the epistemic uncertainty in the transition probabilities between different components states (by interval analysis).

The results and insights obtained can help to improve the global SoS performance by improving the structural response of specific arcs that more easily turn into damage states or by developing a more redundant network that allows the supply of the product from different paths.

REFERENCES

- Adachi, T., and Ellingwood, B. R. (2008). "Serviceability of earthquake-damaged water systems: Effects of electrical power availability and power backup systems on system vulnerability." *Reliability Engineering & System Safety*, 93(1), 78-88.
- Apostolakis, G. (1990). "The Concept of Probability in Safety Assessments of Technological Systems." *Science*, 250(4986), 1359-1364.
- Aven, T., and Zio, E. (2011). "Some considerations on the treatment of uncertainties in risk assessment for practical decision making." *Reliability Engineering & System Safety*, 96(1), 64-74.
- Barry, L. N. (1995). *Stochastic modeling: analysis and simulation*, McGraw-Hill, New York.
- Beer, M., and Ferson, S. (2013). "Special issue of Mechanical Systems and Signal Processing 'Imprecise probabilities - What can they add to engineering analyses?'" *Mechanical Systems and Signal Processing*, 37(1-2), 1-3.
- Beer, M., Ferson, S., and Kreinovich, V. (2013). "Imprecise probabilities in engineering analyses." *Mechanical Systems and Signal Processing*, 37(1-2), 4-29.
- Bernardo, J. M., and Smith, A. F. M. (1994). *Bayesian theory*, Wiley, Chichester.
- Blockley, D. (2013). "Analysing uncertainties: Towards comparing Bayesian and interval probabilities." *Mechanical Systems and Signal Processing*, 37(1-2), 30-42.
- Brissaud, F., Barros, A., Bérenguer, C., and Charpentier, D. (2011). "Reliability analysis for new technology-based transmitters." *Reliability Engineering & System Safety*, 96(2), 299-313.
- Buckley, J. (2004). "Fuzzy Markov Chains." *Fuzzy probabilities and fuzzy sets for web planning*, Springer, Berlin, 35-43.
- Coolen, F. P. A., and Utkin, L. V. "Imprecise probability: a concise overview." *Risk, reliability and societal safety: proceedings of the European safety and reliability conference (ESREL)*, Stavanger, Norway, 1959-66.
- Crespo, L. G., Kenny, S. P., and Giesy, D. P. (2013). "Reliability analysis of polynomial systems subject to p-box uncertainties." *Mechanical Systems and Signal Processing*, 37(1-2), 121-136.
- de Finetti, B. (1974). *Theory of Probability*, Wiley, New York.
- Ferrario, E., and Zio, E. (2014). "Goal Tree Success Tree-Dynamic Master Logic Diagram and Monte Carlo simulation for the safety and resilience assessment of a multistate system of systems." *Engineering Structures*, 59, 411-433.
- Ferson, S. (2005). "Bayesian methods in risk assessment." www.ramas.com/bayes.pdf.
- Ferson, S., and Ginzburg, L. R. (1996). "Different methods are needed to propagate ignorance and variability." *Reliability Engineering & System Safety*, 54(2-3), 133-144.
- Ferson, S., and Hajagos, J. G. (2004). "Arithmetic with uncertain numbers: rigorous and (often) best possible answers." *Reliability Engineering & System Safety*, 85(1-3), 135-152.
- Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., and Ginzburg, L. (2007). "Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty." Sandia National Laboratories, SAND2007-0939, Setauket, New York 11733.
- Ferson, S., Moore, D. R. J., Van den Brink, P. J., Estes, T. L., Gallagher, K., Connor, R. O., and Verdonck, F. (2010). "Bounding Uncertainty Analyses." *Application of Uncertainty Analysis to Ecological Risks of Pesticides*, CRC Press, 89-122.
- Ferson, S., and Tucker, W. T. (2006). "Sensitivity in risk analyses with uncertain numbers." Sandia National Laboratories, SAND2006-2801, Setauket, New York 11733.
- Hu, Y. S., and Modarres, M. (1999). "Evaluating system behavior through Dynamic Master Logic Diagram (DMLD) modeling." *Reliability Engineering & System Safety*, 64(2), 241-269.
- Jalal-Kamali, A., and Kreinovich, V. (2013). "Estimating correlation under interval uncertainty." *Mechanical Systems and Signal Processing*, 37(1-2), 43-53.
- Kalos, M. H., and Whitlock, P. A. (1986). *Monte Carlo methods. Volume: Basics*, Wiley, New York, NY.
- Karanki, D. R., Kushwaha, H. S., Verma, A. K., and Ajit, S. (2009). "Uncertainty Analysis Based on Probability Bounds (P-Box) Approach in Probabilistic Safety Assessment." *Risk Analysis*, 29(5), 662-675.
- Kozine, I. O., and Utkin, L. V. (2002). "Processing unreliable judgements with an imprecise hierarchical model." *Risk, Decision and Policy*, 7(03), 325-339.
- Kuznetsov, V. P. (1991). *Interval statistical models (in Russian)*, Radio i Svyaz, Moscow.
- Limbourg, P., and de Rocquigny, E. (2010). "Uncertainty analysis using evidence theory - confronting level-1 and level-2 approaches with data availability and computational constraints." *Reliability Engineering & System Safety*, 95(5), 550-564.

- Lindley, D. V. (2006). *Understanding uncertainty*, Wiley, Hoboken, NJ.
- Mehl, C. H. (2013). "P-boxes for cost uncertainty analysis." *Mechanical Systems and Signal Processing*, 37(1-2), 253-263.
- Moller, B., Graf, W., and Beer, M. (2003). "Safety assessment of structures in view of fuzzy randomness." *Computers & Structures*, 81(15), 1567-1582.
- MSSP. (2013). "Special issue of Mechanical Systems and Signal Processing "Imprecise probabilities-What can they add to engineering analyses?"." *Mechanical Systems and Signal Processing*, 37(1-2), 1-263.
- Muscolino, G., and Sofi, A. (2013). "Bounds for the stationary stochastic response of truss structures with uncertain-but-bounded parameters." *Mechanical Systems and Signal Processing*, 37(1-2), 163-181.
- NASA. (2010). "Risk-Informed Decision Making Handbook." NASA/SP-2010-576 - Version 1.0.
- Nozick, L. K., Turnquist, M. A., Jones, D. A., Davis, J. R., and Lawton, C. R. (2005). "Assessing the performance of interdependent infrastructures and optimising investments " *International Journal of Critical Infrastructures*, 1(2-3), 144-154.
- Pannier, S., Waurick, M., Graf, W., and Kaliske, M. (2013). "Solutions to problems with imprecise data" "An engineering perspective to generalized uncertainty models." *Mechanical Systems and Signal Processing*, 37(1-2), 105-120.
- Pedroni, N., and Zio, E. (2012). "Empirical Comparison of Methods for the Hierarchical Propagation of Hybrid Uncertainty in Risk Assessment, in Presence of Dependences." *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 20(4), 509-557.
- Pedroni, N., Zio, E., Ferrario, E., Pasanisi, A., and Couplet, M. (2013). "Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model." *Computers & Structures*, 126, 199-213.
- Reid, S. G. (2013). "Probabilistic confidence for decisions based on uncertain reliability estimates." *Mechanical Systems and Signal Processing*, 37(1-2), 229-239.
- Sallak, M., Schon, W., and Aguirre, F. (2013). "Reliability assessment for multi-state systems under uncertainties based on the DempsterShafer theory." *IIE Transactions*, 45(9), 995-1007.
- Sankararaman, S., and Mahadevan, S. (2013). "Distribution type uncertainty due to sparse and imprecise data." *Mechanical Systems and Signal Processing*, 37(1-2), 182-198.
- USNRC. (2009). "Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making." NUREG-1855. US Nuclear Regulatory Commission, Washington, DC.
- Walley, P. (1991). *Statistical reasoning with imprecise probabilities*, Chapman and Hall, New York.
- Zhang, H., Dai, H., Beer, M., and Wang, W. (2013). "Structural reliability analysis on the basis of small samples: An interval quasi-Monte Carlo method." *Mechanical Systems and Signal Processing*, 37(1-2), 137-151.
- Zio, E. (2009). *Computational methods for reliability and risk analysis*, World Scientific Publishing Co. Pte. Ltd., Singapore.
- Zio, E. (2013). *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, Springer, London.

APPENDIX: IMPRECISE (INTERVAL) PROBABILITIES

To explain the meaning of imprecise probabilities (or interval probabilities) consider an event A . Then uncertainty is represented by a lower probability $\underline{P}(A)$ and an upper probability $\bar{P}(A)$, giving rise to a probability interval $[\underline{P}(A), \bar{P}(A)]$, where $0 \leq \underline{P}(A) \leq \bar{P}(A) \leq 1$. The difference $\Delta P(A) = \bar{P}(A) - \underline{P}(A)$ is called the *imprecision* in the representation of the event A . Single-valued probabilities are a special case of no imprecision and the lower and upper probabilities coincide.

Peter M. Williams developed a mathematical framework for imprecise probabilities, based on de Finetti's betting interpretation of probability [de Finetti, 1974]. This foundation was further developed independently by Vladimir P. Kuznetsov and Peter Walley (the former only published in Russian), see [Kuznetsov, 1991] and [Walley, 1991]. Following de Finetti's betting interpretation, the lower probability is interpreted as the maximum price for which one would be willing to buy a bet which pays 1 if A occurs and 0 if not, and the upper probability

as the minimum price for which one would be willing to sell the same bet. If the upper and lower values are equal, the interval is reduced to a precise probability. These references, and [Walley, 1991] in particular, provide an in-depth analysis of imprecise probabilities and their interpretations, with a link to applications to probabilistic reasoning, statistical inference and decisions.

It is however also possible to interpret the lower and upper probabilities using the reference to a standard interpretation of a subjective probability $P(A)$: such an interpretation is indicated by [Lindley, 2006], p. 36. Consider the subjective probability $P(A)$ and say that the analyst states that his/her assigned degree of belief is greater than the urn chance of 0.10 (the degree of belief of drawing one particular ball from an urn which include 10 balls) and less than the urn chance of 0.5. The analyst is not willing to make any further judgement. Then, the interval $[0.10, 0.50]$ can be considered an imprecision interval for the probability $P(A)$.

Of course, even if the assessor assigns a probability $P(A) = 0.3$, one may interpret this probability as having an imprecision interval $[0.26, 0.34]$ (as a number in this interval is equal to 0.3 when displaying one digit only), interpreted analogously to the $[0.1, 0.5]$ interval. Hence imprecision is always an issue in a practical uncertainty analysis context. This imprecision is commonly viewed as a result of measurement problems. Lindley argues that the use of interval probabilities confuses the *concept* of measurement with the *practice* of measurement [Lindley, 2006]. The reference to the urn lottery provides a norm, and measurement problems may make the assessor unable to behave according to it. See also discussion in [Bernardo and Smith, 1994], p. 32.

However, other researcher and analysts have a more positive view on the need for such intervals, see discussions in [Ferson and Ginzburg, 1996; Ferson and Hajagos, 2004; Ferson and Tucker, 2006; Ferson et al., 2007; Ferson et al., 2010; Aven and Zio, 2011]: impression intervals are required to reflect phenomena as discussed above, for example when experts are not willing to express their knowledge more precisely than by using probability intervals.

Imprecise probabilities are also linked to the relative frequency interpretation of probability [Coolen and Utkin, 2007]. The simplest case reflects that the “true” frequentist probability p is in the interval $[\underline{P}(A), \overline{P}(A)]$ with certainty. More generally and in line with the above interpretations of imprecision intervals based on subjective probabilities $P(\cdot)$, a two-level uncertainty characterization can be formulated (see, e.g., [Kozine and Utkin, 2002]): $[\underline{P}(A), \overline{P}(A)]$ is an imprecision interval for the subjective probability $P(a \leq p \leq b)$ where a and b are constants. In the special case that $\underline{P}(A) = \overline{P}(A)$ ($= q$, say) we are led to the special case of a q -100% credibility interval for p (i.e., with subjective probability q , the true value of p is in the interval $[a, b]$). For further details, the reader is referred to the recent Special Issue on

imprecise probabilities appeared on the Journal of Mechanical Systems and Signal Processing [MSSP, 2013].

Paper VI

Hierarchical Graph representations for the evaluation of the robustness of critical infrastructures within a multi-state system-of-systems framework

E. Ferrario, N. Pedroni and E. Zio

Under submission.

Hierarchical Graph representations for the evaluation of the robustness of critical infrastructures within a multi-state system-of-systems framework

E. Ferrario^a, N. Pedroni^a and E. Zio^{a,b}

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy - Electricité de France, at École Centrale Paris - Supelec, France

enrico.zio@ecp.fr, enrico.zio@supelec.fr

^bDepartment of Energy, Politecnico di Milano, Italy

enrico.zio@polimi.it

Abstract

In this paper, we propose a Hierarchical Graph representation to evaluate the robustness of interdependent critical infrastructures (CIs) under a system-of-systems (SoS) framework, taking into account possibly different priorities in the partitioning of the product to the demand nodes. For a more realistic representation, we adopt a multi-state model where different degrees of damage of the individual components are contemplated: the transitions between these different states of damage happen stochastically. The quantitative robustness evaluation is performed by Monte Carlo simulation. We illustrate the approach by way of two case studies: the first one is characterized by small-sized gas and electricity networks and a supervisory control and data acquisition (SCADA) system; the second one is represented by a moderately large power distribution network, adapted from the IEEE 123 node test feeders. Due to the size of the second case study, the robustness analysis is supported by hierarchical clustering.

Keywords: Hierarchical Graph, hierarchical clustering, multi-state model, system of systems, Monte Carlo simulation, Markov and semi-Markov processes.

1. INTRODUCTION

Critical infrastructures (CIs) are essential in providing goods (such as energy, water, data) and services (such as transportation, banking and health care) across local, regional and national boundaries [Kröger and Zio, 2011]. They are getting more and more advanced, i.e., more automated and strongly interconnected due to their extension on large scales and the progressive developments in information technology. However, if on one hand these advances have increased their efficiency (e.g., they adopt more powerful control schemes), on the other hand, they have created new vulnerabilities to component failures, natural and manmade events [Gheorghe and Schlapfer, 2006]. Actually, in the last decades an increased number of disruptive events (natural external events, malicious acts, large scale blackouts) affecting CIs have occurred, e.g., the World Trade Center attack (New York, 2001), the North American blackout (Eastern USA and Canada, 2003), and a M_w 9.0 earthquake and subsequent tsunami (Japan, 2011). Understanding the behavior of interconnected CIs under a system-of-systems (SoS) framework is thus fundamental to their well-functioning.

Traditionally, in risk assessment three steps are performed to evaluate the performance of CIs: (i) the system has to be *represented* to highlight the structural, logical and possible functional connections between the elements; (ii) a *mathematical model* of the system is built to quantitatively describe the system functioning; (iii) the behavior of the system has to be *simulated* under different operational and accidental conditions.

With respect to the system representation (i), several types of approaches exist in literature and many of them rely on a hierarchy or graph structure. Hierarchical modeling has been often adopted to represent and model complex systems, since many organizational and technology-based systems are hierarchical in nature [Haimes, 2012]. Hierarchical functional models include Goal Tree Success Tree (GTST) – also combined with Master Logic Diagram (MLD) [Hu and Modarres, 2000] – and Multilevel Flow Models (MFM) [Lind, 2011a; Lind, 2011b]. In risk analysis, common representation techniques are hierarchical trees that are possibly used to identify i) the initiating causes of a pre-specified, undesired event or ii) the accident sequences that can generate from a single initiating event through the development of structured logic trees (i.e., fault and event trees, respectively) [Zio, 2007]. Recently, also networks have been represented by hierarchical modeling [Gómez et al., 2013].

In complex network theory approaches, instead, complex systems are represented by networks where the nodes stand for the components and the links describe the physical and relational connections among them. Network-based approaches model interdependent CIs on the basis of their topologies or flow patterns [Ouyang, 2014]. Also probabilistic methods (e.g., Petri nets, Bayesian networks and flowgraphs) are based on graph representations.

In this paper, we embrace a SoS framework of analysis and we propose a Hierarchical Graph representation to evaluate the robustness of interdependent CIs, measured as its capability to deliver the required amount of product (e.g., energy, water, etc.) to the demand nodes of the infrastructure. In doing so, we take into account the fact that the demand nodes may have

different importance, which establishes possibly different priorities in the partitioning of the product through the connections and elements of the CI. The representation consists of a graph structured in hierarchical levels that allows highlighting critical arcs and supporting the quantitative robustness evaluation by assigning different priorities to the demand nodes. For a more realistic representation, we adopt a multi-state model where different degrees of damage of the individual components are contemplated [Ferrario and Zio, 2014]; the transitions between these different states of damage happen stochastically and are modeled as Markov and semi-Markov processes.

For illustration purpose, we consider two case studies: the first one is characterized by small-sized interconnected gas and electricity networks and a supervisory control and data acquisition (SCADA) system [Nozick et al., 2005]; the second one is adapted from the IEEE 123 node test feeders [IEEE, 2000] and includes a large electricity distribution network. As a measure of the robustness of the system, we evaluate the steady-state probability distributions of the product (e.g., gas and/or electricity) delivered to the demand nodes.

The quantitative evaluation of the system robustness is performed by Monte Carlo (MC) simulation; in the case study of larger dimension, an unsupervised spectral clustering algorithm is also employed to make the size of the CI manageable and reduce the computational burden related to the analysis [Fang and Zio, 2013].

The remainder of the paper is organized as follows. In Section 2, the Hierarchical Graph representation is introduced; in Section 3, the procedural steps to evaluate the robustness of interconnected CIs by Hierarchical Graph and MC simulation are provided and the combination of Hierarchical Graph and clustering analysis is discussed; Section 4 contains the description of the two case studies, the representation of the corresponding systems and the results obtained; in Section 5, some conclusions are provided. Finally, in the Appendix we present the hierarchical clustering method adopted in the present work, i.e., the unsupervised clustering algorithm.

2. HIERARCHICAL GRAPH REPRESENTATION OF SYSTEMS OF SYSTEMS

The proposed representation technique requires that the critical infrastructure (CI) of interest is first modeled by a directed graph without loops and composed by nodes and arcs; notice that the arcs may represent the components of an infrastructure or the connections between different infrastructures. We then need to distinguish between input, demand (load) and transmission arcs: the “input arcs” connect the sources of product to the network, the “demand arcs” terminate with nodes that require a given amount of product, whereas the “transmission arcs” transfer the product to other components in the network. Notice that the transmission and the demand arcs may coincide: for example, an arc may be needed to supply the connected node and in addition it may be required to transmit the product to other arcs/nodes.

In the Hierarchical Graph representation, the adjective “hierarchical” does not imply a “decomposition of the system into different level of details”, as in other hierarchical models (e.g., Goal Tree Success Tree – Dynamic Master Logic Diagram [Hu and Modarres, 2000] and hierarchical clustering [Gómez et al., 2013]), but it simply means that the graph of interconnected CIs is structured in hierarchical levels. In extreme synthesis, the representation is built as follows: at the bottom of the graph, the inputs (i.e., the arcs through which the product is injected into the networks) are identified; at the top, the goals (i.e., the demand nodes that have to be satisfied) are reported; in the middle, all the other arcs (transmission and/or load arcs) that provide product to the demand nodes are organized in hierarchical levels. These levels are numbered on the basis of the number of demand nodes that are served by the corresponding arcs: the higher the number of demands supplied by an arc, the higher the hierarchical level of that arc. For example, all the arcs that are required to supply LV demand nodes are “placed” at hierarchical level LV .

Formally, let us consider A interconnected infrastructure systems $S^{(a)}$, $a = 1, \dots, A$, constituting the overall system of systems (SoS), numbered in order in such a way that the first q exchange physical product (e.g., energy or material) and the last $(A - q)$ exchange information and are useful for the operation and control of the connected systems (e.g., a supervisory control and data acquisition - SCADA - system). The total number of components (arcs) transmitting physical flow is referred to as N .

For illustration purposes, refer to Figure 1 in which the graph of a SoS (top) and its corresponding Hierarchical Graph (bottom) are reported. The SoS in the example is composed by $A = 4$ systems, where the first two, i.e., $S^{(1)}$ and $S^{(2)}$, exchange physical product (solid links in Figure 1, top) and the last two, i.e., $S^{(3)}$ and $S^{(4)}$, support system $S^{(1)}$ (dotted links in Figure 1, top). The total number of components (arcs) is $N = 8$.

As described above, the Hierarchical Graph depicts the inputs at the bottom of the representation, i.e., in this case, arc $S_1^{(1)}_S S_2^{(1)}$ in Figure 1 (bottom); also, it shows the goals (i.e., the demand nodes) at the top: in this case, the demand nodes are represented by all the nodes of systems $S^{(1)}$ and $S^{(2)}$, except $S_1^{(1)}$, which is the source of product. Finally, it organizes the arcs in different hierarchical levels according to the number of demand nodes they supply: for example, in this case arc $S_1^{(2)}_S S_2^{(2)}$ is at hierarchical level 4 since it provides product to four demand nodes, i.e., $S_2^{(2)}$, $S_3^{(2)}$, $S_4^{(2)}$ and $S_5^{(2)}$. The quantity of product required by the demand nodes is referred to as D_{dem} , where the subscript ‘*dem*’ is the indicator of a given demand node among the N components.

Notice that the arcs referred to the $(A - q)$ control and information systems (which are not contributing to the flow of product, but influence the state of the other arcs) do not appear in the hierarchical structure: instead, they are reported in a trapezoidal frame under the corresponding arc that they affect.

The squares located between the hierarchical levels mean that the product at that level has to be partitioned among the corresponding demand nodes.

This representation allows highlighting all the paths going from the input sources to the end nodes: for example, in Figure 2, the path from input $S_1^{(1)}$ to node $S_3^{(2)}$ is highlighted. In

addition, the representation is able to put in evidence the critical arcs as those located at higher hierarchical levels, since their interruption or degradation affects more demand nodes: for example, in Figure 2 arc $S_2^{(1)} - S_1^{(2)}$ is more critical than arc $S_2^{(1)} - S_3^{(1)}$ since the first one is required to supply five demand nodes (i.e., $S_1^{(2)}$, $S_2^{(2)}$, $S_3^{(2)}$, $S_4^{(2)}$ and $S_5^{(2)}$), whereas the second one is necessary just for node $S_3^{(1)}$.

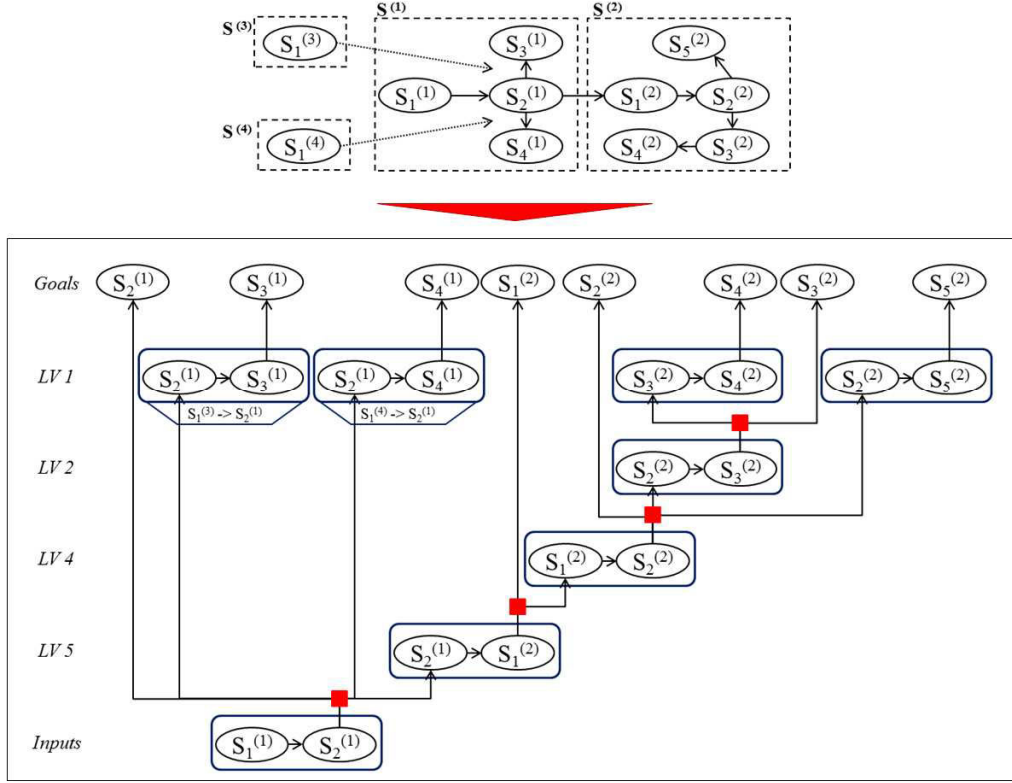


Figure 1: Top: graph of the components of the system of systems; the links represent the exchange of physical product (solid lines) and influence/support relationships (dotted lined). Bottom: corresponding Hierarchical Graph; LV: Level.

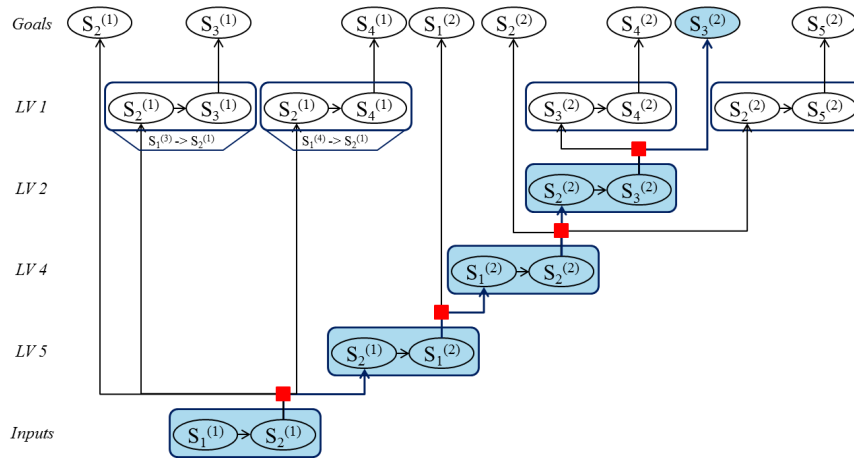


Figure 2: Hierarchical Graph of the system of systems in Figure 1, highlighting the path from the input to demand node $S_3^{(2)}$; LV: Level.

This representation has been introduced to analyze the robustness of interdependent CIs taking into account possibly different priorities in the partitioning of the product to the demand nodes according to their importance: for example, in the case of a malfunctioning in the electrical transmission line higher importance may be given to critical buildings, such as hospitals or industries, with respect to common residential areas. However, these importance criteria are not explicitly shown in the representation, which instead is more focused in highlighting the hard, physical constraints that affect the product partitioning. Three different importance criteria are considered in this work, namely, sequential, proportional and equal; such criteria are explained hereafter with respect to a simple example consisting of an input of 50 units and two demand nodes S_1 and S_2 that require 40 and 100 units, respectively (i.e., $D_{S1} = 40$ and $D_{S2} = 100$).

Sequential importance consists in ranking the demand nodes sequentially on the basis of a chosen “ranking criterion” (e.g., according to their distance from the source node: the closer the node to the source, the higher the importance). In this case, the nodes classified as more important are given the priority; with respect to the example above, if S_1 is more important than S_2 , the input product is given first to S_1 until it is completely supplied, and the rest (i.e., $50 - 40 = 10$ units) to S_2 . Vice versa if S_2 is more important than S_1 , the input is given entirely to S_2 and there is no product left to supply S_1 .

Proportional importance orders the demand nodes on the basis of the quantity of product they need: the higher their demand, the higher their importance. Then, the product is partitioned into the network according to *ratios of importance* associated to each demand node, computed as the ratio between the quantity of product required by a node and the sum of all the demands of the entire system. With respect to the example above, S_2 is automatically more important than S_1 because it requires a higher quantity of product. The ratios of importance are $D_{S2}/(D_{S1} + D_{S2}) = 0.7$ for S_2 and $D_{S1}/(D_{S1} + D_{S2}) = 0.3$ for S_1 . Then, the input is partitioned as follows: the 30% (i.e., 15 product units) is given to S_1 and the 70% (i.e., 35 product units) to S_2 .

Equal importance considers the demand nodes equal, even if their demands are different. Thus, the input is partitioned into equal parts on the basis of the number of the demand nodes. With respect to the example above the input is divided in two equal parts: the 50% is given to S_1 and the 50% to S_2 .

The detailed operative steps of the algorithm for partitioning the product according to these criteria is given in Section 3.1.

3. MONTE CARLO METHOD AND HIERARCHICAL GRAPHS FOR THE SIMULATION OF CRITICAL INFRASTRUCTURES AND THE EVALUATION OF THEIR ROBUSTNESS WITHIN A MULTI-STATE SYSTEM-OF-SYSTEMS FRAMEWORK

Within a multi-state system-of-systems (SoS) analysis framework, we wish to evaluate the performance of critical infrastructures (CIs) in terms of *robustness*, measured by the steady-state probability distribution of the product delivered at the demand nodes of the system. The

quantitative evaluation is carried out by combining the Hierarchical Graph representation of Section 2 and Monte Carlo (MC) simulation.

In Section 3.1, the operative steps of the basic procedure are presented. Then, a modification of the basic procedure is proposed in Section 3.2 to deal with CIs of large size: in particular, a clustering algorithm is adopted to pre-process the CI in order to make its size manageable and reduce the computational burden associated to the analysis; details about the unsupervised spectral clustering technique adopted are given in the Appendix.

3.1. Operative simulation steps combining Monte Carlo method and Hierarchical Graph system-of-systems representations for robustness evaluation

We generically denote the state of a component of the CI (i.e., the capacity of the arcs) as $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, $comp = 1, \dots, N$, where N is the total number of components in the SoS, the subscript ‘*comp*’ indicates the component of interest, identified by its name or by an integer number from 1 to N , NS_{comp} is the total number of states for component *comp*, and i is the state identification number (when $i = 1$, the component is in the worst state, whereas when $i = NS_{comp}$, it is in the best state). For example, supposing that component $S_1^{(2)}_S_2^{(2)}$ can enter three possible states, namely 0, 10 and 20, we denote the total number of states for the component as $NS_{S1(2)_S2(2)} = 3$, and the corresponding states as $\zeta_{S1(2)_S2(2),1} = 0$, $\zeta_{S1(2)_S2(2),2} = 10$, and $\zeta_{S1(2)_S2(2),3} = 20$.

The quantity of product requested by the demand nodes is indicated by the vector $\{D_{dem}\}$, $dem \in \{1, \dots, N\}$, where the subscript ‘*dem*’ identifies the demand nodes.

In what follows, we describe an algorithm combining the Monte Carlo method and Hierarchical Graph representations for the evaluation of the robustness of CIs within a multi-state SoS framework; as mentioned before, the robustness is quantified in terms of the steady-state probability distribution of the product delivered to the demand nodes.

In extreme synthesis, the algorithm requires as inputs:

- the Hierarchical Graph that allows representing the origin-destination paths and the corresponding arcs in hierarchical levels (see Section 2);
- the steady state probabilities of transition between the different arc states (i.e., capacities) $\zeta_{comp,i}$, $i = \{1, 2, \dots, NS_{comp}\}$, $comp = \{1, 2, \dots, N\}$;
- the vector $\{D_{dem}\}$, $dem \in \{1, \dots, N\}$, of product required by demand nodes;
- the importance of the demand nodes (see Section 2).

The output of the algorithm is represented by the steady state probability distributions of the product delivered to the demand nodes. For clarification purposes, we describe the procedure with reference to the simple example of Figure 3, where two interconnected systems, $S^{(1)}$ and $S^{(2)}$, are shown. The $N = 5$ components are: $S_1^{(1)}_S_2^{(1)}$, $S_2^{(1)}_S_3^{(1)}$, $S_2^{(1)}_S_1^{(2)}$, $S_1^{(2)}_S_2^{(2)}$ and $S_2^{(2)}_S_3^{(2)}$. The input component is arc $S_1^{(1)}_S_2^{(1)}$ that serves five demand nodes (i.e., the goals), $S_2^{(1)}$, $S_3^{(1)}$, $S_1^{(2)}$, $S_2^{(2)}$ and $S_3^{(2)}$, explicitly represented at the top of the diagram.

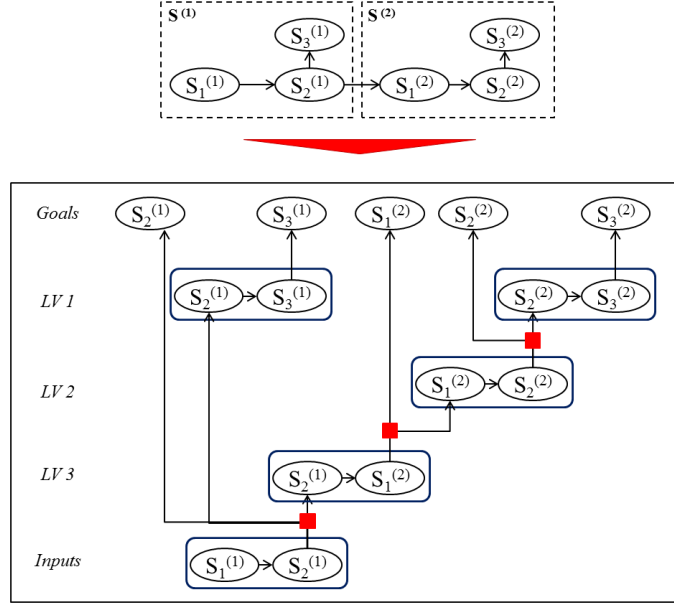


Figure 3: Hierarchical Graph of a generic example taken as reference to illustrate the algorithm; LV: Level.

The evaluation is carried out from the bottom to the top of the hierarchy and consists of the following steps:

- 1) Determine one possible system configuration by sampling the capacity of the arcs, $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, $comp = \{1, \dots, N\}$, from the corresponding steady state probability distributions;
- 2) Identify the minimum arc capacity ($mpath_{dem}$, $dem \in \{1, \dots, N\}$) for each origin-destination path: this capacity corresponds to the maximum product that can be delivered to the corresponding demand node dem , $dem \in \{1, \dots, N\}$; for example, in Figure 4 the minimum arc capacity for the path from $S_1^{(1)}$ to $S_3^{(1)}$ is the minimum among the capacities of arcs $S_1^{(1)}_S_2^{(1)}$ and $S_2^{(1)}_S_3^{(1)}$, connecting $S_1^{(1)}$ and $S_3^{(1)}$;
- 3) Set the input (inp) to the network equal to the capacity of the input arc, i.e., $inp = \zeta_{comp,i}$, where $i \in \{1, 2, \dots, NS_{comp}\}$ and $comp$ is the index of the input arc (in the example of Figure 3, the input arc is $S_1^{(1)}_S_2^{(1)}$);
- 4) If the input is zero ($inp = 0$), no product can be delivered to the demand nodes: $EP_{dem} = 0$ for all $dem \in \{1, \dots, N\}$; otherwise, estimate the optimal flows $\{EP_{dem}\}$ that can be delivered to the demand nodes by the following steps:
 - a. Estimate the vector $\{EP_{dem}\}$ of optimal flows to the demand nodes taking into account i) the importance of the demand nodes and ii) the minimum capacity of each path ($mpath_{dem}$, $dem \in \{1, \dots, N\}$) that limits the quantity of product that can be delivered to the demand nodes (Figure 5, top). For example, referring to Figure 3, let us consider a proportional importance of the demand nodes (see Section 2) and assume that the ratio of importance of $S_2^{(2)}$ and $S_3^{(2)}$ is 0.2 and 0.3, respectively. According to the importance criterion considered and assuming a total input of 100 units, we assign 20 units to $S_2^{(2)}$ and 30 units

to $S_3^{(2)}$. On the contrary, if the minimum capacity of the path to $S_3^{(2)}$ was lower than 30 units (say, $mpath_{S_3^{(2)}} = 10$ units), $S_3^{(2)}$ would receive at most a quantity equal to $mpath_{S_3^{(2)}}$ (i.e., 10 units) and the surplus quantity in exceedance (i.e., 20 units) would be distributed to other nodes (see the following steps 4 b. and 4 c.).

- b. Initialize an auxiliary variable *surp* to zero (i.e., $surp = 0$). This variable is used to quantify the surplus, i.e., the amount of product that cannot be allocated in the network due to arc capacity constraints (i.e., due to the bottlenecks of the infrastructure).
- c. Check if the capacities of the links, $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, can support the sum of the estimated optimal products to the corresponding demand nodes, (*dem*) computed at the previous step 4 a. Such evaluation is performed from the bottom to the top of the diagram. If the sum of the estimated optimal product to the nodes served by a link is higher than its capacity, save the exceeding amount (Δ) in the auxiliary variable *surp* (i.e., $surp = surp + \Delta$) and compute the optimal partition just for the nodes that are supplied by that link, considering as input the corresponding arc capacity $inp = \zeta_{comp,i}$, where $i \in \{1, 2, \dots, NS_{comp}\}$ and *comp* is the link under analysis (Figure 5, middle).
- d. Create a "new" graph, where the "new" capacities of all the arcs are updated on the basis of the quantity of product, $\{EP_{dem}\}$, that has been effectively allocated at step 4 c. In particular, the arc capacities are reduced by the total quantity of product that they have already supplied to the corresponding demand nodes (Figure 5, bottom).
- e. Compute again the minimum arc capacity for each path of the "new" graph (as in step 2) to evaluate the new maximum product that can reach the corresponding demand nodes (Figure 5, bottom).
- f. Update the demands $\{D_{dem}\}$, $dem \in \{1, \dots, N\}$, reducing them by the quantity $\{EP_{dem}\}$ that has been already allocated at step 4.c (Figure 5, bottom).
- g. Set the input *inp* equal to the auxiliary variable *surp* ($inp = surp$) and repeat step 4 until $surp > 0$ and the minimum ("new") arc capacity for at least one path is not zero. When one of these conditions is verified, the final vector $\{EP_{dem}\}$ of the optimal product that can be delivered to the demand nodes is determined (Figure 5, bottom).

The procedure above is repeated a large number of times (e.g., 10000) for many different MC-sampled values of the arc capacities and the probability distribution of the product delivered at steady state to each demand node is obtained.

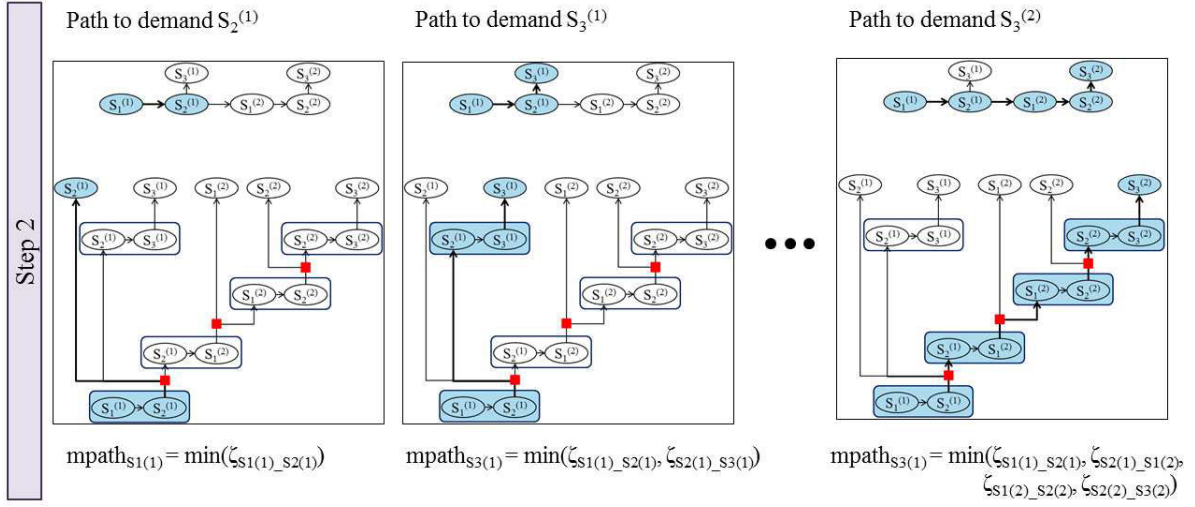


Figure 4: Exemplification of step 2 of the algorithm with respect to the example proposed in Figure 3.

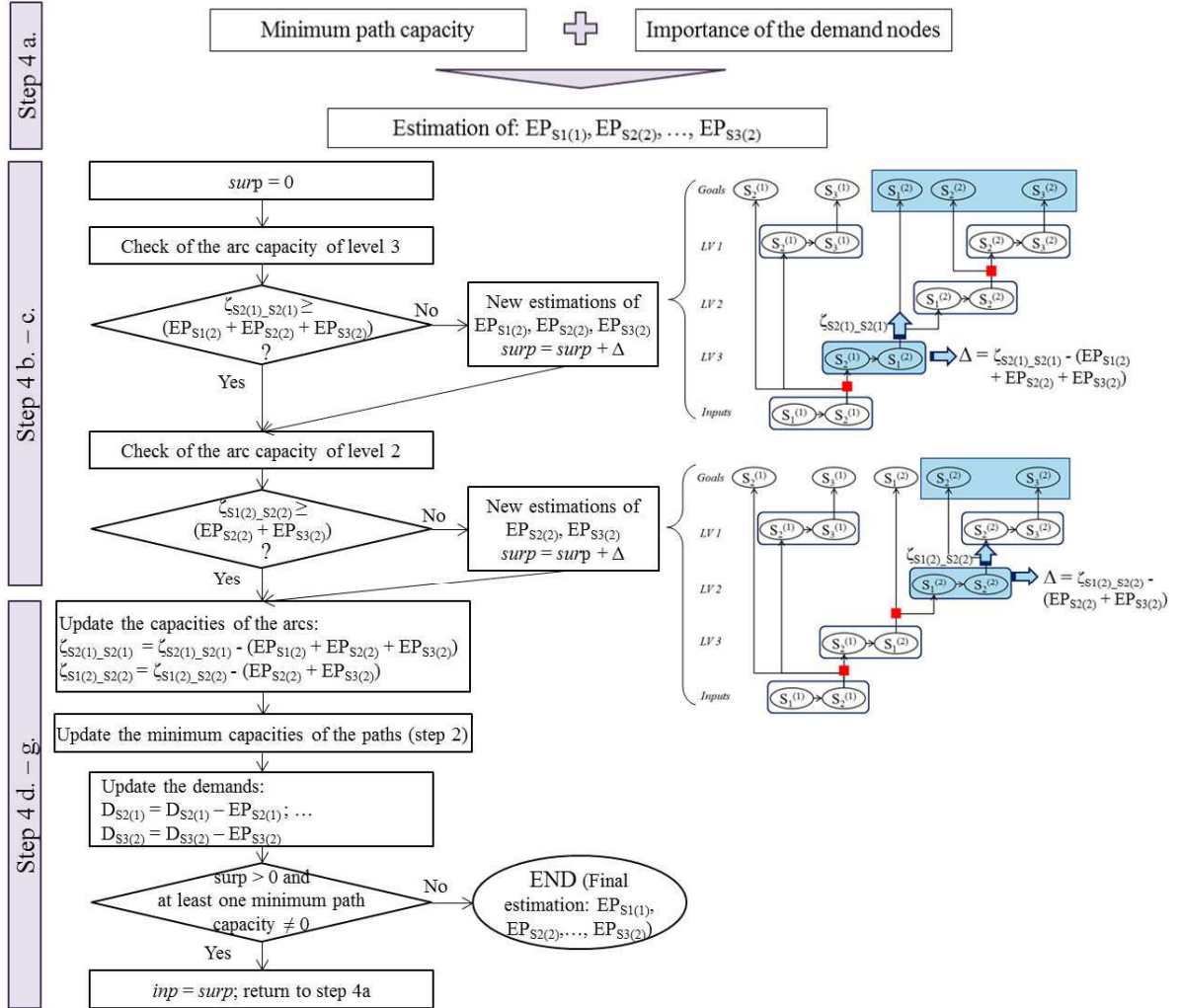


Figure 5: Exemplification of step 4 of the algorithm with respect to the example proposed in Figure 3.

It is worth noting that the procedure proposed is based on several iterative estimations of the vector $\{EP_{dem}\}$, obtained by repeating steps 4 a. – g. from the bottom to the top of the hierarchy: in the very first iteration, the system configuration is the one sampled at step 1. and the input product corresponds to the capacity of the input arc; then, at each loop a “new” graph is considered where (i) the new product input value is represented by the surplus (*surp*), i.e., the amount of product that has not been allocated in the network at the previous iteration, (ii) the “new arc capacities are reduced by the total amount of product they have already supplied at the previous iteration and (iii) the “new” demands are scaled by the quantity already allocated in the previous iteration.

Finally, it is worth mentioning that a drawback of the Hierarchical Graph representation proposed may be represented by its difficult applicability to large networks since *all* the origin-destination paths have to be identified and the bottlenecks of each path have to be spotted out. To overcome this limitation, we propose to pre-process the infrastructure system by means of a clustering algorithm to reduce the systems dimension by “collapsing” many components in few representative clusters and then apply the Hierarchical Graph to the “clustered” infrastructure (details about the particular clustering technique adopted in the present work, namely unsupervised spectral clustering algorithm, are reported in the Appendix). The general concepts underlying the pre-processing phase based on clustering is discussed in the following Section 3.2.

3.2. Combination of the Hierarchical Graph representation and a clustering algorithm for managing large-sized critical infrastructures

In order to manage large-sized CIs, it may be useful to resort to clustering techniques to reduce the complexity and dimension of the system. For illustration purpose, refer to the simple example of Figure 6, left, where the original components of a network, namely $S_1^{(I)}$, $S_2^{(I)}$, ..., $S_{16}^{(I)}$, are reported. According to some features of interest (e.g., proximity), such components can be clustered in groups of “similar characteristics”: in the example proposed, four clusters, C_1 , ..., C_4 , are identified (dotted oval shape in Figure 6, left). Then, a less complex analysis can be performed on the new fictitious, artificial (i.e., clustered) network, composed just by the identified clusters (Figure 6, right).

The cluster analysis can be carried out at different level of details: an artificial network with a high number of clusters is closer to the original one and, thus, it is more detailed (i.e., it carries more information) than one with a small number of clusters. The system can be clustered at different levels of details, which allows building a hierarchical¹ clustering representation where the different hierarchical levels correspond to the different levels of detail of the analysis.

¹ Notice that in this case the term “hierarchical” refers to the level of detail of the clustering and not to the levels of the Hierarchical Graph representation that instead correspond to the number of demands served by a given arc of the network.

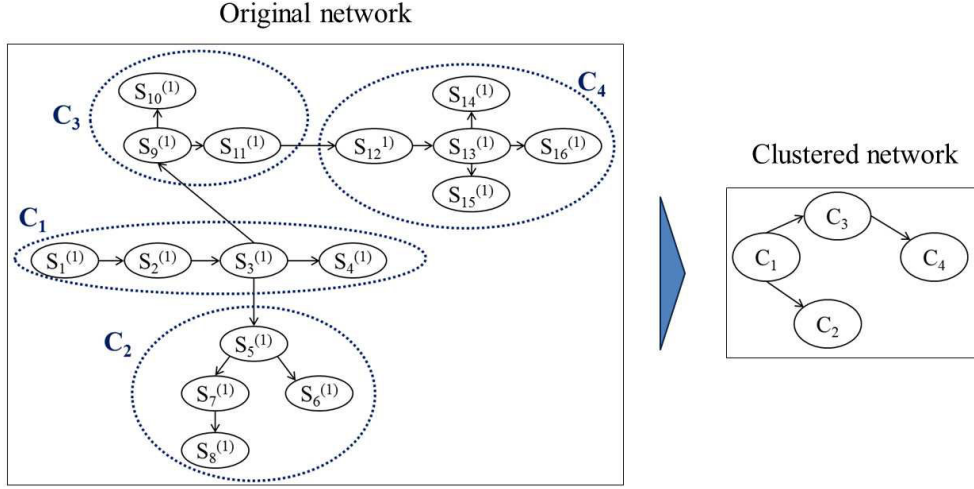


Figure 6: Exemplification of the clustering procedure.

In order to reduce the size of the infrastructure under analysis, the SoS is clustered (possibly at different hierarchical levels of detail) according to the procedure described in the Appendix: an artificial (fictitious) network composed by k_L clusters, $C_1^{(L)}, \dots, C_{k_L}^{(L)}$, is produced at each (clustering) hierarchical level L . Notice that the last level of the clustering hierarchy coincides with the real SoS, i.e., the corresponding clusters coincide with the actual/original/real nodes of the SoS. The clustering is performed on the entire network except for the input nodes that are left out (only one generation node is considered in the application of the present work). For illustration purposes, Figure 7 depicts a sketch of the decomposition in five (clustering) hierarchical levels of a SoS with one input node, $S_I^{(1)}$; level 1 of the hierarchy is then composed by two nodes: the input, $S_I^{(1)}$, and the rest of the system “condensed” in cluster $C_1^{(1)}$. The clustering algorithm allows a new analysis at hierarchical level 2 and it decomposes cluster $C_1^{(1)}$ of hierarchical level 1 into two clusters $C_1^{(2)}$ and $C_2^{(2)}$. At this point, if we want to increase the level of refinement of the analysis we can use the algorithm to further split clusters $C_1^{(2)}$ and $C_2^{(2)}$. In the example of Figure 7, this results in the decomposition of cluster $C_1^{(2)}$ into three clusters ($C_1^{(3)}$, $C_2^{(3)}$ and $C_3^{(3)}$) and cluster $C_2^{(2)}$ into five clusters ($C_4^{(3)}$, $C_5^{(3)}$, $C_6^{(3)}$, $C_7^{(3)}$ and $C_8^{(3)}$). The Hierarchical Graph representation of the decomposed system at level 3 is also shown on the right.

A cluster k is characterized by its demand D_k that is the sum of the demands of the real nodes at its inside: for example, cluster $C_1^{(4)}$ of Figure 7 has demand equal to the sum of the demands of nodes $S_3^{(1)}$, $S_6^{(1)}$ and $S_7^{(1)}$.

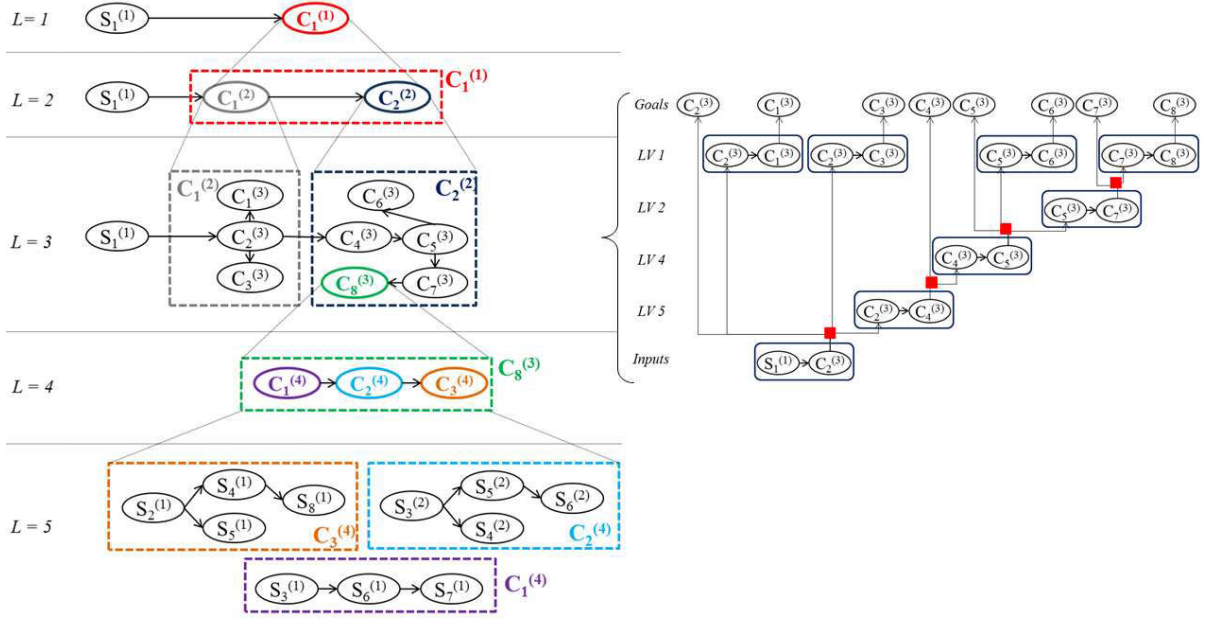


Figure 7: Left: sketch of the decomposition of a system in five hierarchical levels (L) where the last one ($L = 5$) coincides with the actual nodes of the system; right: Hierarchical Graph of the corresponding hierarchical level 3; LV: Level of the Hierarchical Graph.

For a given clustering hierarchical level L the quantitative evaluation of the performance of the "artificial" clustered system is carried out as illustrated in Section 3.1 by means of an indicator that represent the "global" state of the clusters $C_1^{(L)}, \dots, C_{k_L}^{(L)}$ of level L as a "synthesis" of the real capacity $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, $comp = 1, \dots, N$, of the arcs contained in the cluster itself. Actually, a measure of the cluster state is needed to approximately estimate the quantity of product that a cluster can receive and deliver to other clusters.

To represent the state of a cluster k (i.e., its performance) we consider an indicator id_k based on the ratio of the expected capacity of cluster k at current and at nominal (optimal) conditions as follows:

$$id_k = \frac{\sum_{comp=1}^{n_k} w_{comp} * \zeta_{comp,i}}{\sum_{comp=1}^{n_k} w_{comp} * \zeta_{comp,NS_{comp}}}, \quad (1)$$

where $comp$ indicates the component (arc) of the original network, n_k is the number of arcs inside cluster k , $\zeta_{comp,i}$, $i \in \{1, 2, \dots, NS_{comp}\}$, is the current (i.e., actual / sampled) state of the component $comp$, $\zeta_{comp,NS_{comp}}$ is the maximum capacity of the arc $comp$, and w_i is the weight associated to the capacity of the arc $comp$. The weight w_{comp} is computed as the ratio between the capacity of the arc $comp$ and the sum of the maximum capacities of all the arcs of the network, i.e., $w_{comp} = \zeta_{comp,i} / \sum_{comp=1}^N \zeta_{comp,NS_{comp}}$, $i \in \{1, 2, \dots, NS_{comp}\}$ and gives an idea of the weight of the arc in the entire network. The index of the cluster state (id_k), takes value between 0 and 1.

Notice that the state of a cluster affects the cluster itself and the connected clusters, since the cluster is both a *fictitious load node* (which should provide itself with the required amount of

product) and a *fictitious transmission node* (which should transmit the product to the other connected clusters). The top of Figure 8 shows two clusters, C_1 and C_2 , supplied by the input source $S_1^{(I)}$: cluster C_2 is both a load and a transmission node, since on one side it contains five demand nodes ($S_2^{(I)}$, $S_3^{(I)}$, $S_4^{(I)}$, $S_5^{(I)}$ and $S_6^{(I)}$ in Figure 8, bottom) and on the other side it is required to transmit the product to cluster C_2 . In particular, the product from input source $S_1^{(I)}$ has to pass through two arcs ($S_2^{(I)} \rightarrow S_5^{(I)}$ and $S_5^{(I)} \rightarrow S_6^{(I)}$) contained in C_1 to reach cluster C_2 : if their capacities decrease, then the flow to nodes $S_5^{(I)}$ and $S_6^{(I)}$ (i.e., to the cluster C_1) and to nodes $S_7^{(I)}$, $S_8^{(I)}$ and $S_9^{(I)}$ (i.e., to the cluster C_2) is reduced.

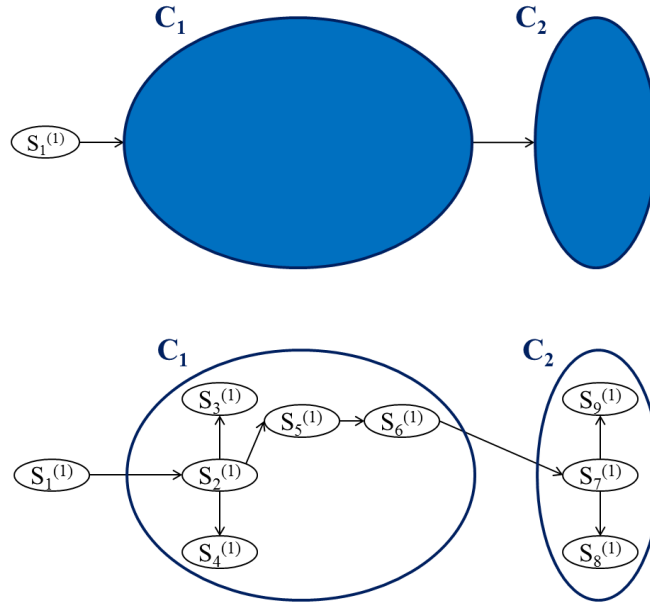


Figure 8: Top: artificial system composed by two clusters C_1 and C_2 supplied by one input node $S_1^{(I)}$. Bottom: illustration of the real nodes inside the fictitious clusters: two arcs of C_1 are needed to supply C_2 .

Thus, when the "capacity" of a cluster decreases, the consequence is twofold: the cluster cannot satisfy itself (i.e., the demand nodes at its inside) and also the connected clusters at best. In order to take into account the "twofold" reduction of performance, we "artificially reduce" the amount of product that can be given to the cluster itself and that can be delivered to the connected clusters by multiplying i) the maximum demand D_k that it requires and ii) the maximum capacities, $\zeta_{comp, NS_{comp}}$, of the arcs (*comp*) that link the output clusters, respectively, by the indicator of the state of the cluster, id_k .

4. APPLICATIONS

In this Section, we apply the proposed Hierarchical Graph representation to two case studies (hereafter referred to as "A" and "B"): case study A (Section 4.1) consists of two interdependent infrastructures (gas and electric power networks) and a supervisory control and data acquisition (SCADA) system connected to the gas network; case study B (Section 4.2) considers an electric power distribution network adapted from the IEEE 123 node test feeders. In both cases, we adopt a multi-state model to account for different degrees of

damage of the components and we describe state transitions (random failures) by Markov and semi-Markov processes.

4.1. Case study A

The case study is taken from [Nozick et al., 2005] and deals with two interconnected infrastructures, i.e., a natural gas distribution network and an electricity generation/distribution network (Figure 9, solid and dashed lines, respectively). The gas distribution network is supported by a SCADA system (Figure 9, dotted lines). The objective of this interconnected system of systems (SoS) is to provide the necessary amount of gas and electricity (hereafter also called “product”) to four demand nodes (end-nodes), namely *D1* and *D2* (gas) and *L1* and *L2* (electricity).

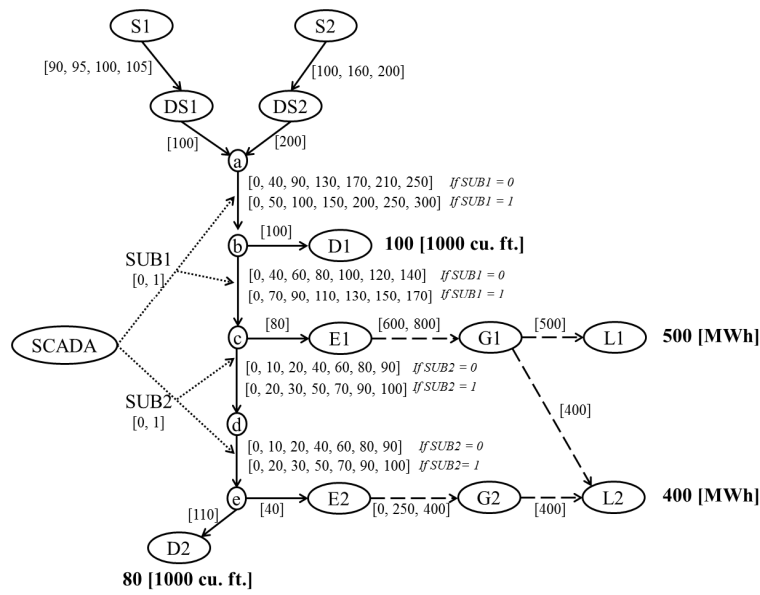


Figure 9: Interdependent gas (solid lines) and electric (dashed lines) infrastructures and SCADA system (dotted lines) [Nozick et al., 2005]. The possible states (i.e., capacities) of the arcs are given in square brackets; the quantities of product demanded by end-nodes *D1*, *D2*, *L1*, *L2* are reported in bold.

The gas distribution network, supplied by two sources of gas (namely, *S1* and *S2*, connected to the network by arcs *S1_DS1* and *S2_DS2*, respectively), provides gas to the end-nodes *D1* and *D2* and to two nodes of the electricity network (*E1* and *E2*). Once the gas enters into nodes *E1* and *E2*, it is transformed into electrical energy that flows through arcs *E1_G1* and *E2_G2* (representing the electric power generation stations) to supply the end-nodes of electricity (*L1* and *L2*); notice that demand *L2* can be supplied by both electrical generations *E1_G1* and *E2_G2*. The assumption is made that the gas-electricity transformation occurs with a constant coefficient, i.e., 100 cu. ft. of natural gas produces 1 MWh of electricity [Nozick et al., 2005].

A SCADA system controls the gas flow through arcs *a_b*, *b_c*, *c_d* and *d_e*. It is assumed that: i) the SCADA has two core subsystems controlling different sets of arcs (in particular, the first one – SUB1 – refers to links *a_b* and *b_c*, whereas the second one – SUB2 – controls

arcs c_d and d_e); ii) the SCADA is always provided with electric power [Nozick et al., 2005].

The capacities of the arcs of the gas and electricity networks (determining the maximum flows of gas or electricity supported by the arcs) can be deterministic (i.e., fixed constant values) or stochastic (i.e., randomly evolving in time) (Figure 9, values in the square brackets). The stochastic capacities give rise to a multi-state model that reflects the possibly different degrees of damage of the arcs. On the contrary, the SCADA system state is defined by a binary random variable, whose values 1 and 0 represent its complete and partial functioning, respectively. For example, when the state of the SCADA subsystem SUB1 (controlling arcs a_b and b_c) is 0, the capacity of these arcs decreases because of the incorrect information provided by the SCADA subsystem (even if the arcs are not subject to a direct damage). On the basis of the two states of the SCADA subsystems, two different vectors of capacities are identified for each arc a_b , b_c , c_d and d_e : as illustrated in Figure 9, the first vector is used when the corresponding SCADA subsystem is in state 0, whereas the second one is employed when the SCADA subsystem is in state 1.

Changes in the arc capacities are due to random failures or recovery actions. The state transitions over time are modeled by Markov and semi-Markov processes as in [Nozick et al., 2005]. Semi-Markov processes are adopted to represent the evolution of the capacities of the gas supply links ($S1_DS1$ and $S2_DS2$), whereas Markov processes are used for all the others arcs. Both Markov and semi-Markov processes for a generic component ‘*comp*’ are defined by a transition probability matrix $\underline{P}_{comp} = \{p_{ij} : i, j = 1, 2, \dots, S_{comp}\}$, where p_{ij} is the one-step probability of transition from state i to state j . In addition, the semi-Markov processes are characterized by continuous probability distributions for the holding time T_{comp}^{ij} , i.e., for the time of residence of a component in state i before performing a transition to state j .

The steady-state probability vectors for a generic component ‘*comp*’ described by a Markov process is computed as $\underline{\Pi}^{comp} = \underline{\Pi}^{comp} \cdot \underline{P}^{comp}$ [Zio, 2009]. For a semi-Markov process, this equation is weighted by the expected time of residence τ^i in a given state i before performing a transition, i.e., as $\xi^{comp,i} = \Pi^{comp,i} \cdot \tau^i / \sum_{j=1}^{S_{comp}} \Pi^{comp,j} \cdot \tau^j$, $i = 1, \dots, NS_{comp}$ [Barry, 1995].

Figure 10 reports the steady-state probability vectors of the arcs of the system of Figure 9 assuming the state transition probabilities given in [Nozick et al., 2005].

$$\begin{array}{l}
 \Pi^{S1_DS1} = \begin{array}{c|c} 1 & 0.0001 \\ 2 & 0.0002 \\ 3 & 0.5001 \\ 4 & 0.4996 \end{array} \quad \Pi^{S2_DS2} = \begin{array}{c|c} 1 & 0.0033 \\ 2 & 0.1703 \\ 3 & 0.8264 \end{array} \quad \Pi^{SCADA} = \begin{array}{c|c} 1 & 0.0042 \\ 2 & 0.0012 \\ 3 & 0.0012 \\ 4 & 0.9934 \end{array} \\
 \\
 \Pi^{a_b, b_c, c_d, d_e} = \begin{array}{c|c} 1 & 0.0022 \\ 2 & 0.0045 \\ 3 & 0.0022 \\ 4 & 0.0065 \\ 5 & 0.0038 \\ 6 & 0.0052 \\ 7 & 0.9756 \end{array} \quad \Pi^{E2_G2} = \begin{array}{c|c} 1 & 0.0006 \\ 2 & 0.0007 \\ 3 & 0.9987 \end{array} \quad \Pi^{E1_G1} = \begin{array}{c|c} 1 & 0.0011 \\ 2 & 0.9989 \end{array}
 \end{array}$$

Figure 10: Steady state probability vectors for the states of the components of the case study A (comp = S1_DS1, S2_DS2, a_b, b_c, c_d, d_e, E2_G2, E1_G1, SCADA).

4.1.1. Hierarchical Graph representation for the system of case study A

In Figure 11 the Hierarchical Graph of the system of case study A (Section 4.1) is illustrated. The injection of product (i.e., gas) in the SoS is made through arcs *S1_DS1* and *S2_DS2*, located at the bottom of the diagram (Section 2). Since both arcs enter in node *a*, also the following links *DS1_a* and *DS2_a* are considered part of the inputs and reported at the bottom of the hierarchy. Four demand nodes, i.e., *D1*, *D2* (gas) and *L1*, *L2* (electricity), represent the goals of the analysis and they are explicitly located at the top of the diagram. The graph presents four hierarchical levels: in level 4, arc *a_b*, is reported since it supplies all the four demand nodes; in level 3, arc *b_c* is depicted, since it serves three demand nodes (i.e., *L1*, *L2* and *D2*); in level 2, arcs *c_E1*, *E1_G1*, *c_d* and *d_e* are considered, since they supply two demand nodes: in particular, arcs *c_E1* and *E1_G1* supply *L1* and *L2*, whereas arcs *c_d* and *d_e* serve *L2* and *D2*; in level 1, there are the remaining arcs that are related just to one demand node: for example, *e_E2* serves only node *L2*. The influence of the SCADA subsystem SUB1 on the arcs *a_b* and *b_c* and of the SCADA subsystem SUB2 on the arcs *c_d* and *d_e* is illustrated in the trapezoidal frames under the corresponding arcs.

For illustration purposes, three different importance criteria for the demand nodes are considered (see Section 2):

- sequential importance: the demand nodes are ranked on the basis of the proximity to the sources: 1) *D1* (the most important), 2) *L1*, 3) *L2* and 4) *D2* (the least important).
- proportional importance: the demand nodes are satisfied on the basis of their demands (the nodes that require more product are given the priority). Since *D1* and *D2* require 100 and 80 cu. ft. of gas and *L1* and *L2* need 500 and 400 MWh (equivalent to 50 and 40 cu. ft.), the resulting importance ranking of the demand nodes is: 1) *D1* (the most important), 2) *D2*, 3) *L1* and 4) *L2* (the least important).
- equal importance: the product is divided equally among four demand nodes.

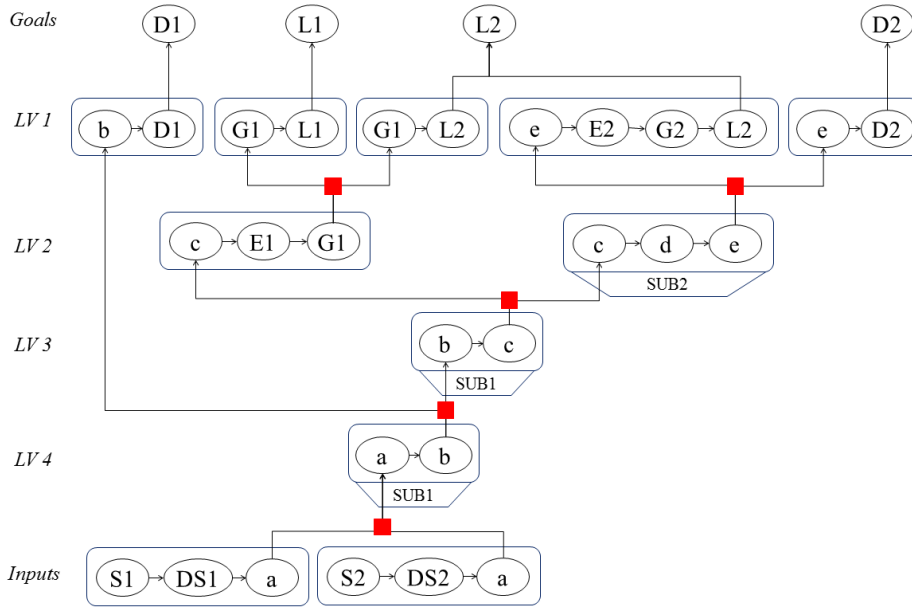


Figure 11: Hierarchical Graph of the system of systems depicted in Figure 9 (case study A); LV: Level.

4.1.2. Results of case study A

Figures 12 – 14 show the steady state probability distributions of the product delivered to *D1* (top, left), *L1* (top, right), *L2* (bottom, left) and *D2* (bottom, right), considering sequential, proportional and equal importance of the demand nodes, respectively, obtained by the procedure illustrated in Section 3.1.

Table 1 reports the steady state probabilities of (i) delivering the (maximum, optimal) required product to the demand nodes (top) and (ii) delivering a quantity of product exceeding the 90% of the corresponding demands (bottom).

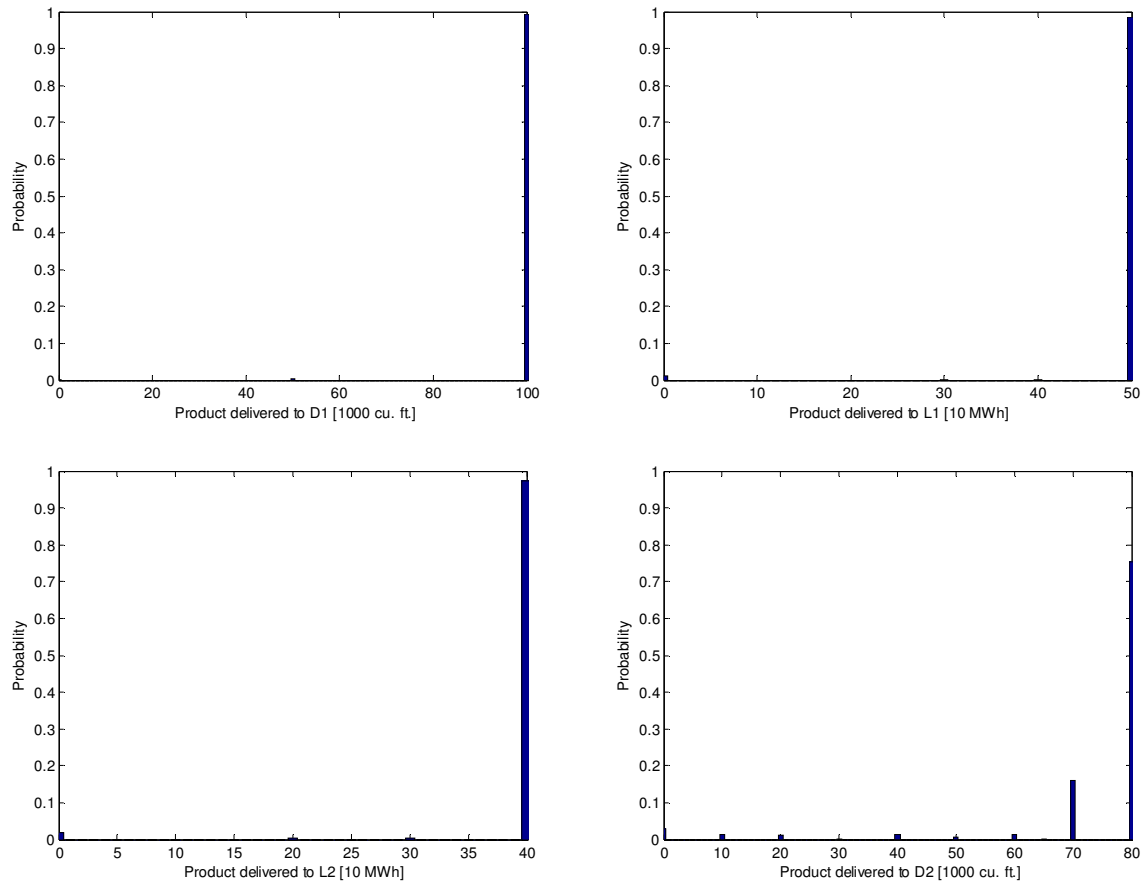
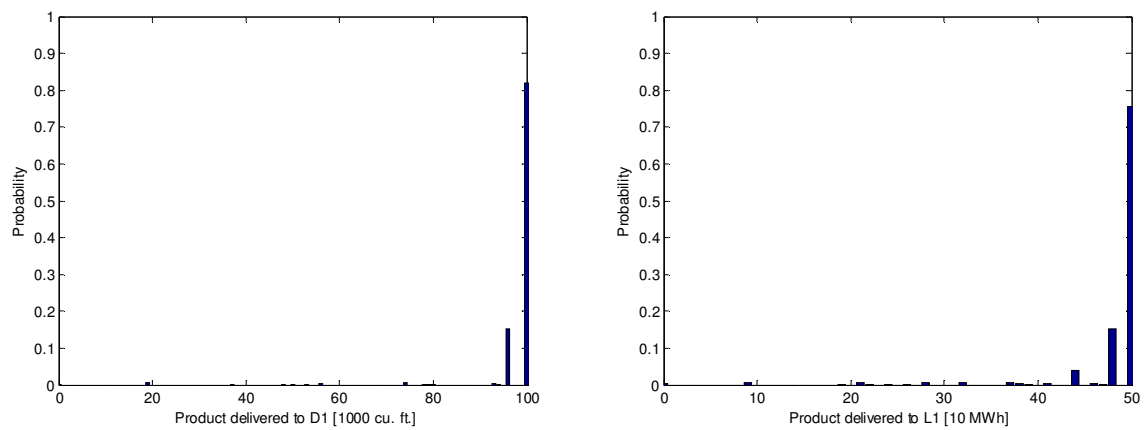


Figure 12: Steady state probability of the product delivered to the demand nodes D1 (on the top, left), L1 (on the top, right), L2 (on the bottom, left) and D2 (on the bottom, right), considering a sequential importance of the demand nodes.



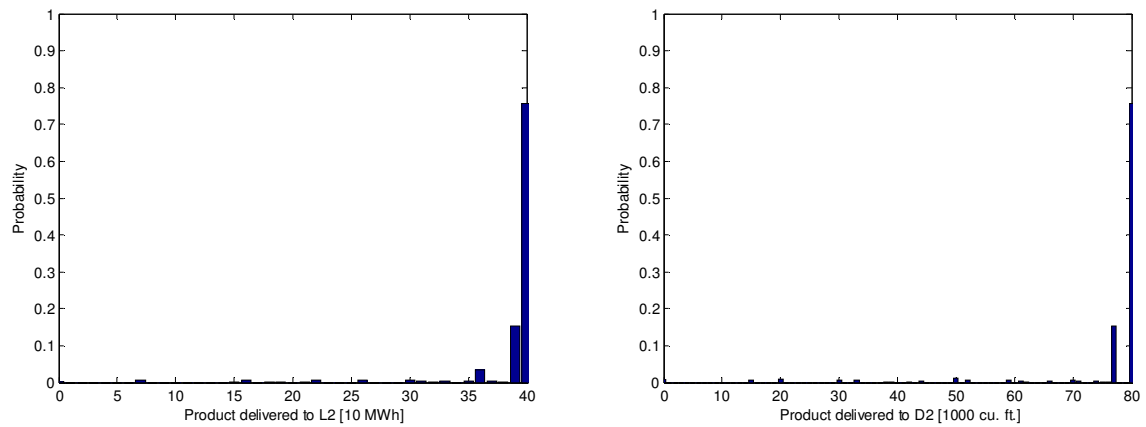


Figure 13: Steady state probability of the product delivered to the demand nodes D1 (on the top, left), L1 (on the top, right), L2 (on the bottom, left) and D2 (on the bottom, right), considering a proportional importance of the demand nodes.

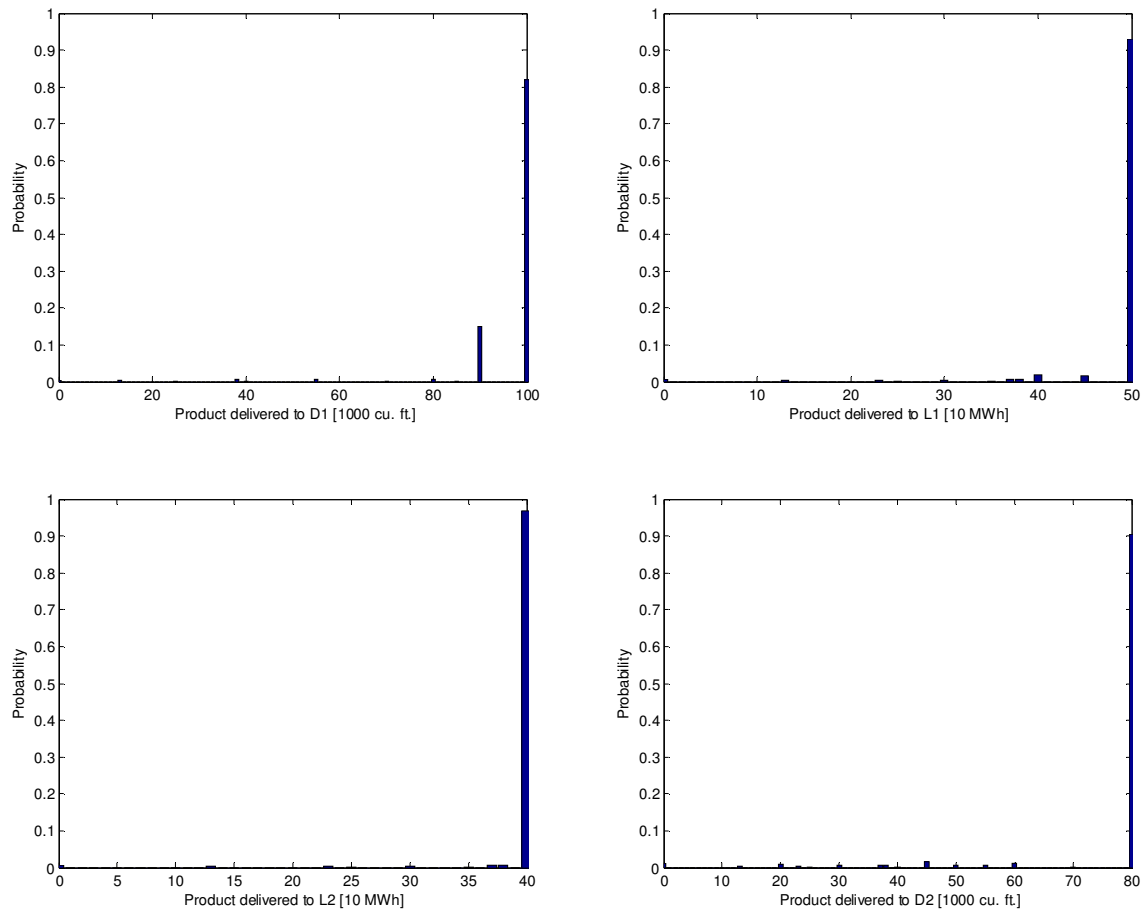


Figure 14: Steady state probability of the product delivered to the demand nodes D1 (on the top, left), L1 (on the top, right), L2 (on the bottom, left) and D2 (on the bottom, right), considering an equal importance of the demand nodes.

Table 1: Steady state probabilities of (i) delivering the (maximum, optimal) required product to the demand nodes (top) and (ii) delivering a quantity of product exceeding the 90% of the corresponding demands (bottom).

Importance criterion	P(D1 = 100 [1000 cu. ft.])	P(L1 = 50 [10 MWh])	P(L2 = 40 [10 MWh])	P(D2 = 80 [1000 cu. ft.])
Sequential	0.9927	0.9867	0.9723	0.7526
Proportional	0.8195	0.7563	0.7563	0.7568
Equal	0.8205	0.9306	0.9678	0.9063

Importance criterion	P(D1 > 90 [1000 cu. ft.])	P(L1 > 45 [10 MWh])	P(L2 > 36 [10 MWh])	P(D2 > 72 [1000 cu. ft.])
Sequential	0.9927	0.9867	0.9723	0.7526
Proportional	0.9778	0.9155	0.9507	0.9160
Equal	0.9717	0.9482	0.9816	0.9063

As expected, in the case of *sequential importance* of the demand nodes, *D1* is the demand node most satisfied, whereas *D2* is the least served: the corresponding probabilities of being completely satisfied are 0.9927 and 0.7526, respectively (Table 1, top). Differently, in the ranking produced by the *proportional importance* criterion, *D2* is more important than *L1* and *L2*. Thus, the probability of delivering the required product to *D2* should increase, whereas the probability of satisfying the last two demand nodes should decrease. In facts, the steady state probabilities of delivering the required maximum product to *L1* and *L2* decrease (Table 1, top). On the contrary, the probability $P(D2 = 80 [1000 \text{ cu. ft.}])$ of delivering the maximum product to *D2* remains almost the same, since (i) the path needed to reach *D2* is affected by the uncertainty on the capacity of many arcs and (ii) *D2* is not the more important demand node. However, an effect of the proportional importance criterion on *D2* can be seen by analyzing the steady state probability $P(D2 > 72 [1000 \text{ cu. ft.}])$ that the product given to *D2* exceeds the 90% of its demand: actually, this value increases considerably from 0.7526 of the previous case (sequential importance, Table 1, top) to 0.9160 (proportional importance, Table 1, bottom). Finally, the criterion of *equal importance*, turns out to give preference to the nodes that have the lowest demands (this is expected since a lower demand has higher probability to be satisfied by an equal partition of the product). Actually, in this case the steady state probabilities of supplying *L1*, *L2* and *D2* at the optimum level increase with respect to the case of proportional importance (from 0.7563 to 0.9306 for *L1*, from 0.7563 to 0.9678 for *L2* and from 0.7568 to 0.9063 for *D2*, see Table 1 top); the probabilities to serve *D1* remain almost the same (around 0.82, see Table 1 top).

4.2. Case study B

Figure 15 shows the electric power distribution network here considered: it is adapted from the IEEE 123 nodes test feeder [IEEE, 2000] in the sense that regulators, capacitors, switches and feeders with length equals to zero are neglected. With these simplifications, the network is composed of 114 nodes: 1 generation point (node 115) and 113 load/transmission nodes. Node 61 of the original IEEE 123 node test feeders is missing here, since after the removal of switches and transformers it turns out to be an end node with load equal to zero. The arcs (i.e.,

the feeders) connect different nodes and distribute the power through the network. In the analysis by Hierarchical Graph, we focus on the arcs (and not on the nodes), hereafter also called “components”; thus, the total number N of components is 113.

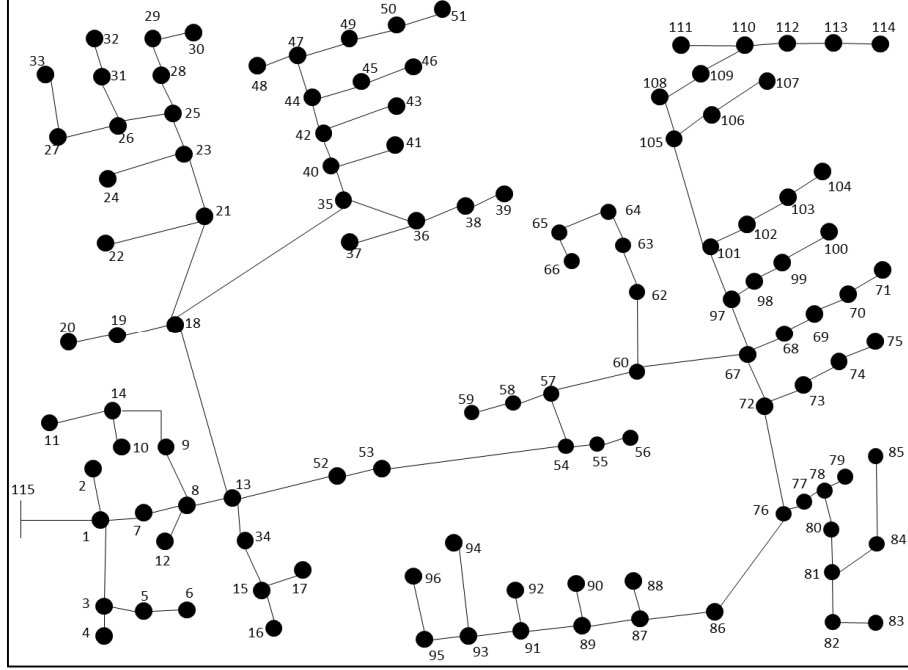


Figure 15: IEEE 123 node test feeders adapted to the purposes of the present analysis

Four states, $i = 1, \dots, 4$, characterize the capacity ($\zeta_{comp,i}$, $comp = 1, \dots, N$) of the arcs: the first one ($i = 1$) represents the worst state and corresponds to the complete failure of the link, i.e., no product can flow through it ($\zeta_{comp,1} = 0$, $comp = 1, \dots, N$); the fourth one ($i = 4$) corresponds to the best state, i.e., to its nominal designed capacity. To obtain a reasonable value for the nominal capacity of the arcs useful for the purposes of the present analysis, we have solved the DC power flow equations (DCPF) that provide the physical flows of electricity through the lines given i) the total power injected in the network (assumed equal to the sum of all the power required by the demand nodes), ii) the network topology (i.e., the adjacency matrix) and iii) the reactance of all the arcs [McCalley, 2012]; the nominal capacity of a link is set equal to the optimal power flow through that link.

Table 2 reports the demands of the network nodes provided by the IEEE database and the maximal capacity $\zeta_{comp,4}$ of the arcs ($comp = 1, \dots, N$) obtained by the DC power flow equations.

Table 2: Left: loads [KW] for all the nodes of the network (except the generation node whose load is zero).
Right: maximal capacity $\zeta_{comp,4}$ of the arcs.

NODES	LOADS [kW]	NODES	LOADS [kW]	ARCS (comp)		$\zeta_{comp,4}$ [kW]	ARCS (comp)		$\zeta_{comp,4}$ [kW]
1	40	58	20	1	2	20	57	60	1815
2	20	59	20	1	3	100	58	59	20
3	0	60	20	1	7	3330	60	62	370
4	40	62	40	3	4	40	60	67	1425
5	20	63	40	3	5	60	62	63	330
6	40	64	75	5	6	40	63	64	290
7	20	65	140	7	8	3310	64	65	215
8	0	66	75	8	12	20	65	66	75
9	40	67	0	8	9	100	67	68	120
10	20	68	20	8	13	3190	67	72	865
11	40	69	40	9	14	60	67	97	440
12	20	70	20	13	34	100	68	69	100
13	0	71	40	13	18	1115	69	70	60
14	0	72	0	13	52	1975	70	71	40
15	0	73	40	14	11	40	72	73	120
16	40	74	40	14	10	20	72	76	745
17	20	75	40	15	16	40	73	74	80
18	0	76	245	15	17	20	74	75	40
19	40	77	40	18	19	80	76	77	240
20	40	78	0	18	21	280	76	86	260
21	0	79	40	18	35	755	77	78	200
22	40	80	40	19	20	40	78	79	40
23	0	81	0	21	22	40	78	80	160
24	40	82	40	21	23	240	80	81	120
25	0	83	20	23	24	40	81	82	60
26	0	84	20	23	25	200	81	84	60
27	0	85	40	25	26	80	82	83	20
28	40	86	20	25	28	120	84	85	40
29	40	87	40	26	27	40	86	87	240
30	40	88	40	26	31	40	87	88	40
31	20	89	0	27	33	40	87	89	160
32	20	90	40	28	29	80	89	90	40
33	40	91	0	29	30	40	89	91	120
34	40	92	40	31	32	20	91	92	40
35	40	93	0	34	15	60	91	93	80
36	0	94	40	35	36	80	93	94	40
37	40	95	20	35	40	635	93	95	40
38	20	96	20	36	37	40	95	96	20
39	20	97	0	36	38	40	97	98	120
40	0	98	40	38	39	20	97	101	320
41	20	99	40	40	41	20	98	99	80
42	20	100	40	40	42	615	99	100	40
43	40	101	0	42	43	40	101	102	100
44	0	102	20	42	44	555	101	105	220
45	20	103	40	44	45	40	102	103	80
46	20	104	40	44	47	515	103	104	40
47	105	105	0	45	46	20	105	106	80
48	210	106	40	47	48	210	105	108	140
49	140	107	40	47	49	200	106	107	40
50	40	108	0	49	50	60	108	109	140
51	20	109	40	50	51	20	109	110	100
52	40	110	0	52	53	1935	110	111	20
53	40	111	20	53	54	1895	110	112	80
54	0	112	20	54	55	40	112	113	60
55	20	113	40	54	57	1855	113	114	20
56	20	114	20	55	56	20	115	1	3490
57	0	115	0	57	58	40			

Changes in the arc capacities are due to random failures (as in the previous case study A): the state transitions over time are modeled as Markov processes.

4.2.1. Hierarchical Graph representation and hierarchical clustering for case study B

An unsupervised spectral clustering algorithm (Section 3.1 and the Appendix) is applied to the IEEE electric power distribution network of interest and identifies five hierarchical clustering levels: level 1 is composed by the generation node and one cluster representative of the whole network, whereas level 5 coincides with the original network. In Figures 16 – 18 the clusters generated at levels 2, 3 and 4 are reported, respectively: the corresponding Hierarchical Graph representation to levels 2 and 3 is also given for illustration purposes. Notice that level 2 is characterized by two clusters, level 3 by 8 clusters and level 4 by 29 clusters (beside the generation point): these clustered representations correspond to different possible “levels of detail” that the analyst may choose to study the network.

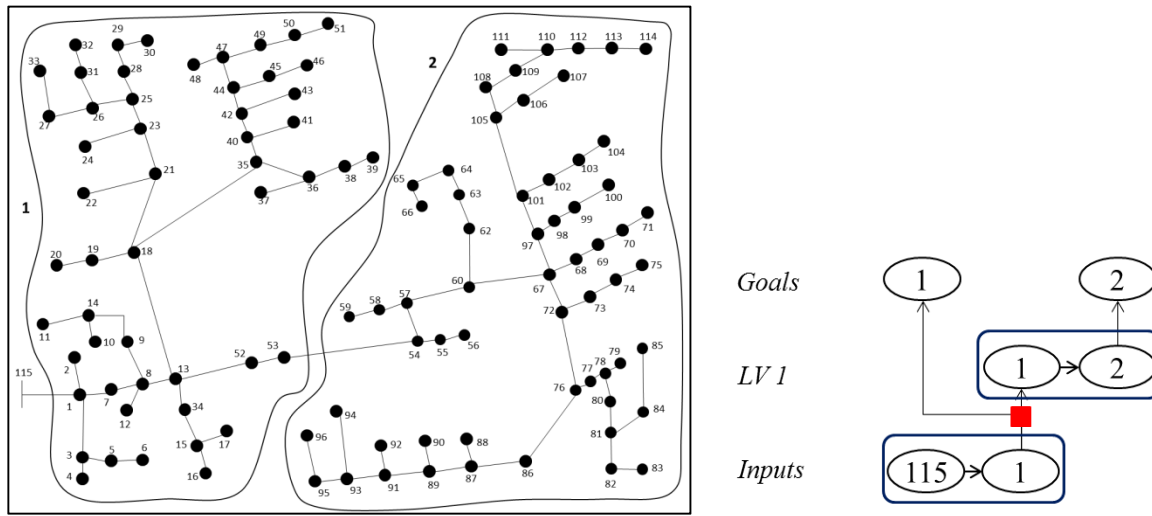


Figure 16: IEEE network clustered at hierarchical level 2 (left) together with the corresponding Hierarchical Graph representation (right); LV: Level.

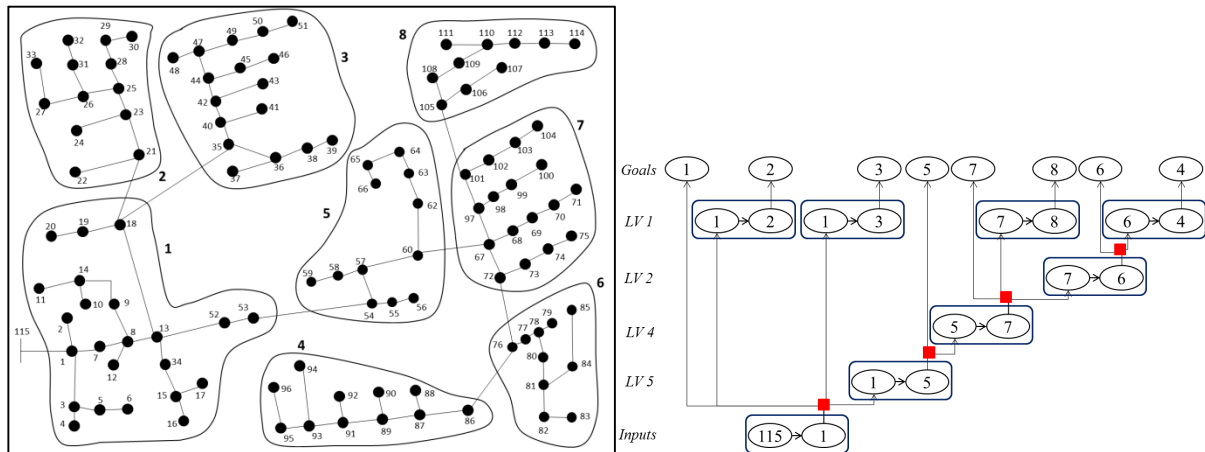


Figure 17: IEEE network clustered at hierarchical level 3 (left) together with the corresponding Hierarchical Graph representation (right); LV: Level.

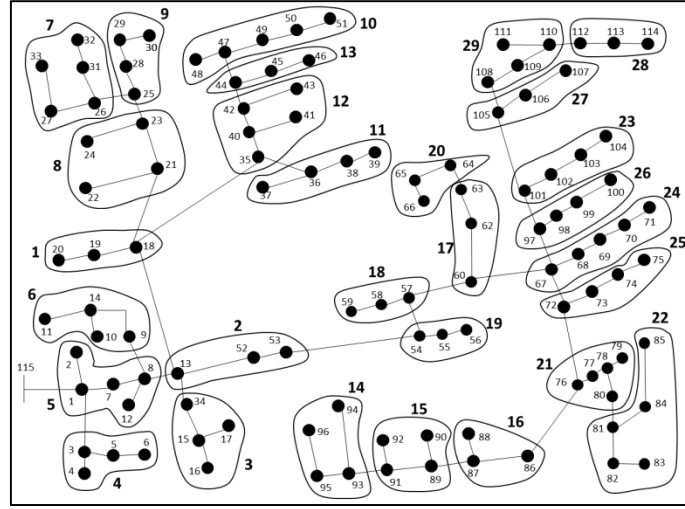


Figure 18: IEEE network clustered at hierarchical level 4.

4.2.2. Results of case study B

In order to evaluate the robustness of the electric power distribution network of Section 4.2, an unsupervised spectral clustering algorithm (Appendix) is first applied to reduce the size of the system and then Monte Carlo simulation is performed at a given clustering level in combination with the Hierarchical Graph representation techniques of Section 3.1. The outcome of the procedure is represented by the steady state probability distribution of the electricity delivered at the demand nodes (or clusters). In this case we consider a proportional importance of the clusters/demand nodes (see Section 2).

As illustrated in Section 4.2.1, five hierarchical clustering levels are identified by the unsupervised spectral clustering algorithm. In the following, for the sake of completeness the results obtained from the analysis of the “fictitious” clustered networks are given for all the levels 2, 3 and 4 of the clustering hierarchy and are compared to those obtained from the analysis of the original (i.e., not clustered) system.

In order to perform a fair comparison, the values of the product delivered to the nodes obtained analyzing the real network are “grouped” on the basis of the clustered structure of the corresponding “fictitious” network: for example, at hierarchical clustering level 4 the products delivered to nodes 9, 10, 11 and 14 have to be summed to allow the comparison with the corresponding cluster 6 of the “fictitious” network (Figure 18). The comparison is then made by means of two quantitative indicators: 1) the mean value μ_k of the product delivered to the cluster k at steady state and 2) the expected product not supplied $idNS_k$ to cluster k , computed as:

$$idNS_k = \frac{D_k - \mu_k}{D_k}, \quad (2)$$

where D_k is the demand of cluster k (equal to the sum of the demands of the nodes of the original network contained in cluster k).

We have performed the comparison between the performances of the clustered and real networks in three cases:

- *Case 1*: all the arcs are characterized by the steady state probability vector $\underline{\Pi}^1$ of Figure 19 left; in this case the system is expected to have a satisfactory performance as all the arcs remain in their best state (i.e., state 4) with very high probability (i.e., 0.9982).
- *Case 2*: all the arcs are characterized by the steady state probability vector $\underline{\Pi}^2$ of Figure 19, middle; in this case the arcs turn into state 3 (i.e., a state of partially reduced functionality) with high probability (i.e., around 0.5) so that the global performance of the system is not expected to be very high;
- *Case 3*: same as case 2, except for the fact that the arcs contained in clusters 8 and 12 at hierarchical clustering level 4 are assumed to be critical and characterized by the steady state probability vector $\underline{\Pi}^3$ of Figure 19, right; in particular, they degrade to their worse states (e.g., state 2) with high probability.

		Steady state probability			Steady state probability			Steady state probability
$\underline{\Pi}^1 =$	1	0.0006	$\underline{\Pi}^2 =$	1	0.0005	$\underline{\Pi}^3 =$	1	0.0291
	2	0.0003		2	0.0010		2	0.0971
	3	0.0009		3	0.4995		3	0.8651
	4	0.9982		4	0.4990		4	0.0087

Figure 19: Steady state probabilities of the four levels of capacity of the arcs of the network in Figure 15 for cases 1 (left), 2 (middle), 3 (right).

In the following, comments on the results obtained are given with reference to Tables 4 – 9. In particular, in Tables 4 – 6 the clusters at hierarchical levels 2, 3 and 4, respectively, are ranked in ascending order based on the mean values (μ_k) of the probability distributions of the product received at steady state (i.e., the clusters at the top of the ranking receive less product than the clusters at the bottom). Tables 4 – 6 also show the ranking of the clusters based on the corresponding demands D_k . In Tables 7 – 9 the clusters at hierarchical levels 2, 3 and 4, respectively, are ranked in ascending order based on the product not supplied ($idNS_k$) (i.e., the mean value of the product supplied to clusters at the top of the ranking is closer to their demands than that of the clusters at the bottom). Tables 7 – 9, right, illustrate also the ranking of the clusters based on their distance (measured in terms of number of arcs) from the source point: for example, at level 3 cluster 8 is connected to the input source by four arcs (i.e., 115_1, 53_54, 60_67 and 101_105) (see Figure 17). The reference to the distance of a demand node from the source is motivated by the fact that in general the higher the distance from the source, the higher the probability of reduced supply of product.

Case 1

The results obtained from simulations on the clustered and real networks are very similar at all the hierarchical clustering levels. In particular, the higher is the hierarchical level (i.e., the finer is the clustering), the closer are the results of the artificial network to those of the real network.

The probability distribution functions (pdfs) of the product delivered to the clusters of the fictitious and real networks at steady state present the same modal values (i.e., the same peaks) at hierarchical levels 2, 3 and 4 and mean values very close to each other at all hierarchical clustering levels (the maximum percentage difference is around 1.1 % at level 2). For illustration purposes, the pdfs of the product delivered at steady state to the cluster 1 at hierarchical levels 2, 3 and 4 obtained from the original (left) and clustered (right) networks are given in Figures 20 – 22, respectively.

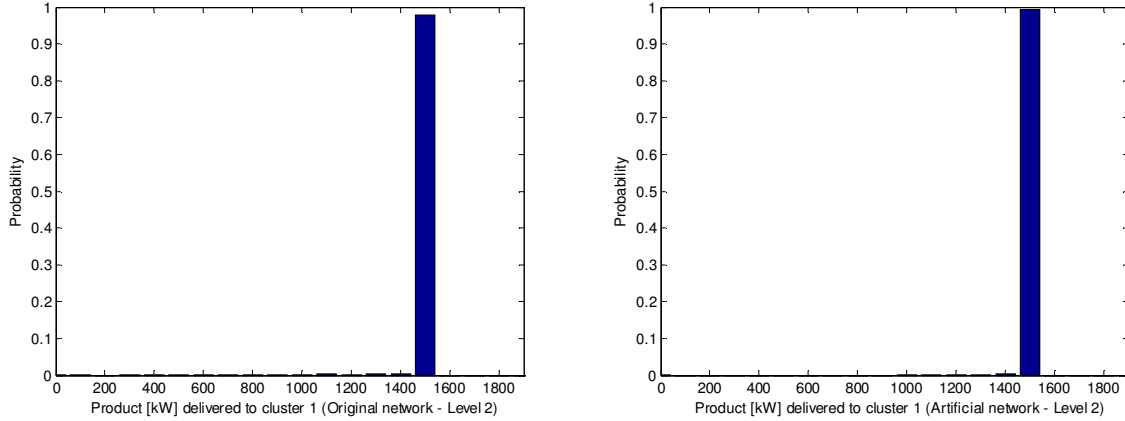


Figure 20: steady state probability distributions of the product delivered to cluster 1, considering the original network (left) and the artificial network clustered at hierarchical level 2 (right).

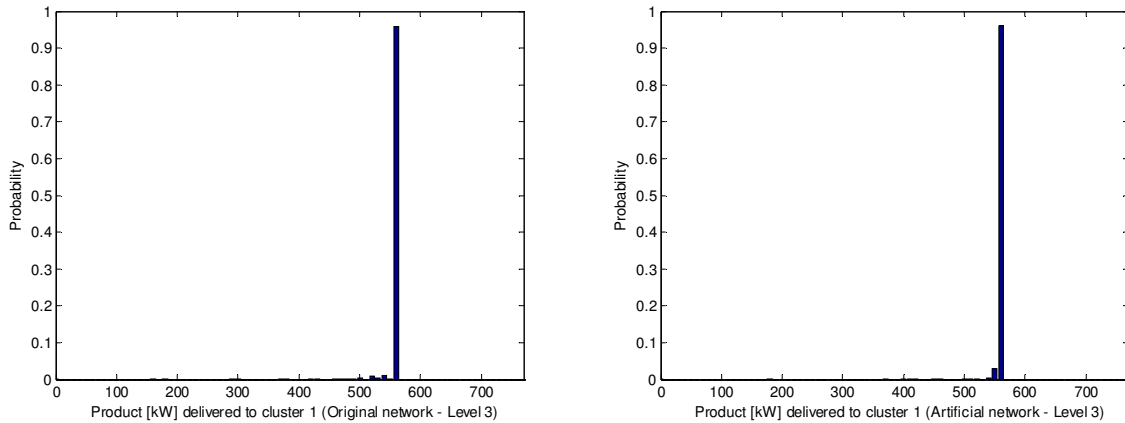


Figure 21: steady state probability distributions of the product delivered to cluster 1, considering the original network (left) and the artificial network clustered at hierarchical level 3 (right).

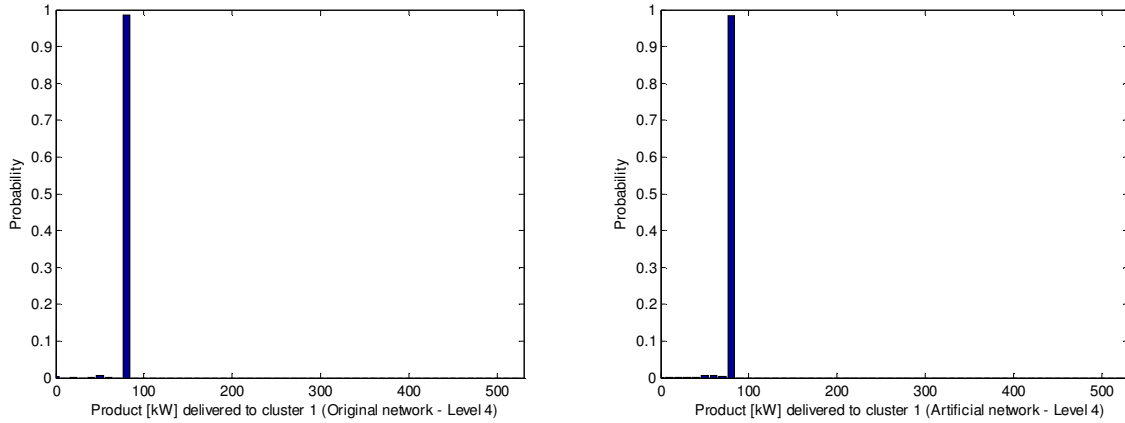


Figure 22: steady state probability distributions of the product delivered to cluster 1, considering the original network (left) and the artificial network clustered at hierarchical level 4 (right).

As a consequence of the strong similarity in the distributions of the product delivered, the ranking of the clusters based on the mean values μ_k is almost the same at all hierarchical clustering levels (see Tables 4 – 6): for example, at hierarchical level 3 the ranking is {8, 4, 2, 7, 5, 6, 1, 3}, which is exactly the same as the one based on the demands D_1, \dots, D_8 (Table 5). At hierarchical level 4, although the rankings are not identical, there is a strong correspondence between the rankings of “groups” of clusters: for example, clusters 13, 18, 19 and 29 are always placed at the top of the ranking, clusters 10, 20, 21 are always found at the bottom, whereas clusters 3, 4, 5, 6, 16, 17, 23 occupy the central part of the ranking (Table 6). This result is expected given the assumption of proportional importance of the demand nodes and the configurations of the network mainly in the best conditions.

The values of the product not supplied $idNS_k$ are in general very low (i.e., lower than 0.025), so that the presence of critical (i.e., less supplied) cluster(s) is not evident. In general, the ranking produced by the clustered and real networks are similar (Tables 7 and 9). Differences in the ranking can be found at clustering level 4, but they can be neglected given the very low values of the corresponding product not supplied index. Also a correspondence between the rankings of groups of clusters can still be found: for example, the clusters closer to the generation source, e.g., 4, 5, 13, 18, 19, are slightly more supplied than those farther, e.g., 14, 15, 16, 22, 27, 28, 29 for both the artificial and real systems, confirming the physical coherence of the approach.

Case 2

In general, the results obtained from simulations on the artificial and real networks are less similar to each other than in case 1, where the arc state capacities present less variability. However, the trends in the network behavior that have been pointed out in the previous case can still be identified.

The pdfs of the product delivered to the clusters of the fictitious and real networks at steady state present (almost) the same peak values at hierarchical levels 2, 3 and 4. Also the mean

values are very close to each other at hierarchical levels 3 and 4 (actually, the maximum percentage difference is 6.9 % for cluster 1 at level 3 and 3.9 % for cluster 5 at level 4); instead, at level 2 the difference is larger, i.e., 10.6 % for cluster 1. Thus, as expected, the higher the hierarchical level (i.e., the finer the clustering), the closer the results of the artificial network to those of the real network. For illustration purposes, the pdfs of the product delivered at steady state to cluster 1 at hierarchical levels 2, 3 and 4 obtained from the original (left) and artificial (right) networks are given in Figures 23 – 25.

With respect to the previous case 1, it can be noticed that the distributions present higher variability; in particular, this effect is more evident in the results obtained on the artificial networks. For example, referring to cluster 1, the variances of the distributions of the product delivered are 3514, 725 and 25 at levels 2, 3, 4, respectively, for the original network, whereas they are 25283, 3199 and 35, respectively, for the clustered network. This can be due to the presence of the clusters and to the definition of their performance (Section 3.2). Actually, the state of a cluster is represented by indicator id_k (eq. 1) that tries to capture and synthetize the main features of the nodes inside the cluster itself. By so doing, the (detrimental) effect of the degradation of an arc is “spread” through the entire cluster instead of having a contained local impact only on the physically connected nodes. This leads to an increase in the variability of the performance of the network: as expected, such variability decreases with the level of detail of the analysis, i.e., with the reduction of the cluster size.

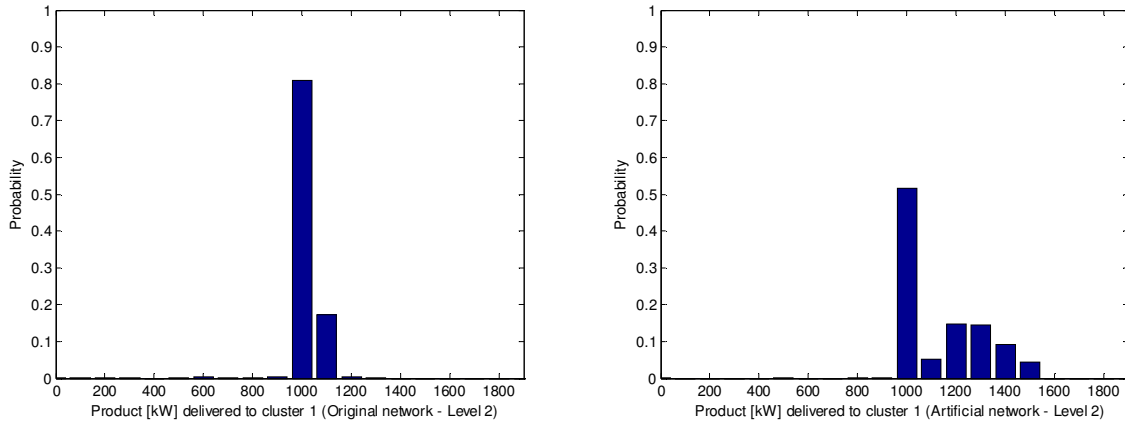


Figure 23: steady state probability distribution of the product delivered to cluster 1 considering the original network (left) and the artificial network at hierarchical clustering level 2 (right).

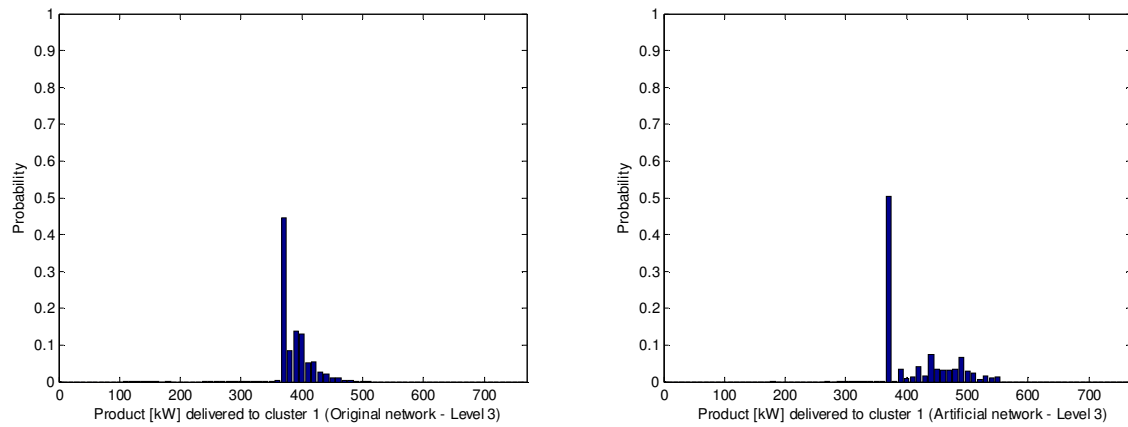


Figure 24: steady state probability distribution of the product delivered to cluster 1 considering the original network (left) and the artificial network at hierarchical clustering level 3 (right).

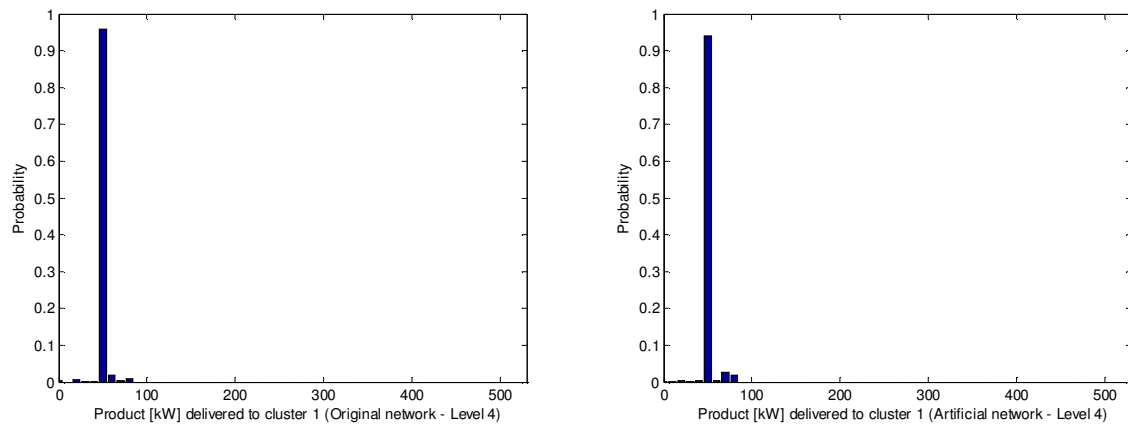


Figure 25: steady state probability distribution of the product delivered to cluster 1 considering the original network (left) and the artificial network at hierarchical clustering level 4 (right).

Given this difference between the distributions obtained from the fictitious and original networks with respect to case 1, the ranking of the clusters based on these mean values is the same only at level 2 (where there are just two clusters), whereas it is slightly different at the other clustering levels. For example, at level 3 the rankings of clusters 5 and 6 are switched.

At hierarchical level 4, although the rankings are not identical, there is a strong correspondence between the rankings of “groups” of clusters: for example, clusters 13, 18, 19 and 29 are always placed at the top of the ranking, clusters 10, 12, 20, 21, are always found at the bottom, whereas clusters 3, 4, 5, 6, 16, 17, 23 occupy the central part of the ranking (Table 6).

The ranking based on the product not supplied is similar for the groups of clusters at the top and at the bottom (i.e., at the tails) of the ranking, so that it is possible to highlight those that are more supplied (e.g., clusters 1 and 5 at level 3 and clusters 1, 2, 3, 4, 5, 6 at level 4) with respect to those who are less supplied (e.g., clusters 4 and 8 at level 3 and 14, 28, 29 at level 4, see Table 9). As for case 1, the tails of the ranking reflect the distance of the clusters from

the generation point: at the top we find the elements closer to the input source, whereas at the bottom those that are farther (see Tables 8 – 9, right).

Case 3

The pdfs of the product delivered to the clusters of the fictitious and real networks at steady state present almost the same modal (peak) values at hierarchical levels 2, 3 and 4. Instead, the mean values are farther from each other than in the previous case 2: for example, at hierarchical level 3 the maximum percentage difference is 14.5 % for cluster 3, whereas at level 4 it is 7.9 % for cluster 10. At level 2 the maximal percentage difference is larger, i.e., 15.9 % for cluster 1. Again, the higher the hierarchical level, the higher the similarity of the results produced by the artificial network and by the real network. For illustration purposes, the pdfs of the product delivered at steady state to cluster 1 at hierarchical levels 2, 3 and 4 obtained from the original (left) and artificial (right) networks are given in Figures 26 – 28.

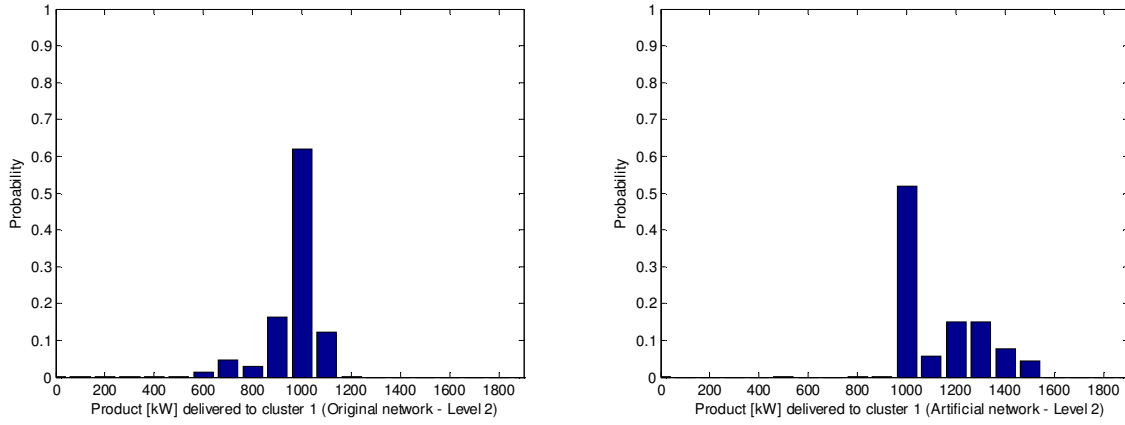


Figure 26: steady state probability distribution of the product delivered to cluster 1 considering the original network (left) and the artificial network at hierarchical clustering level 2 (right).

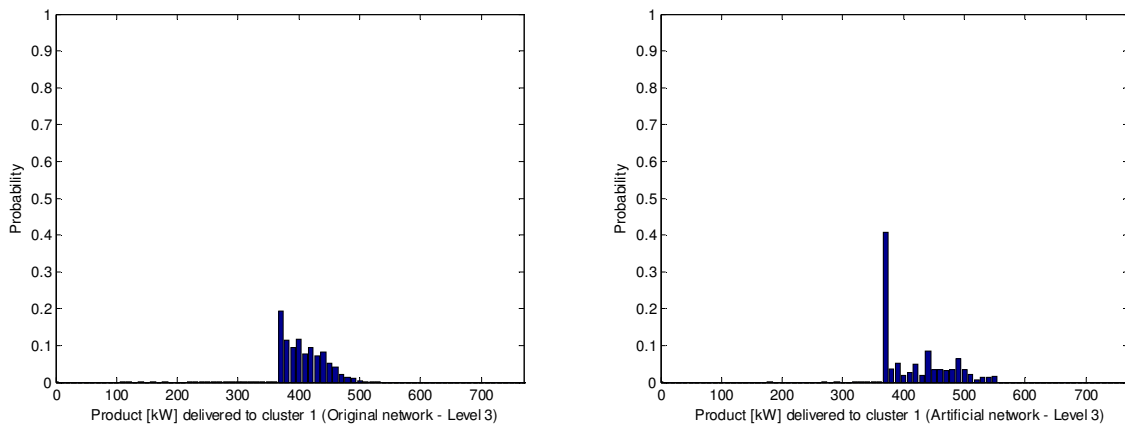


Figure 27: steady state probability distribution of the product delivered to cluster 1 considering the original network (left) and the artificial network at hierarchical clustering level 3 (right).

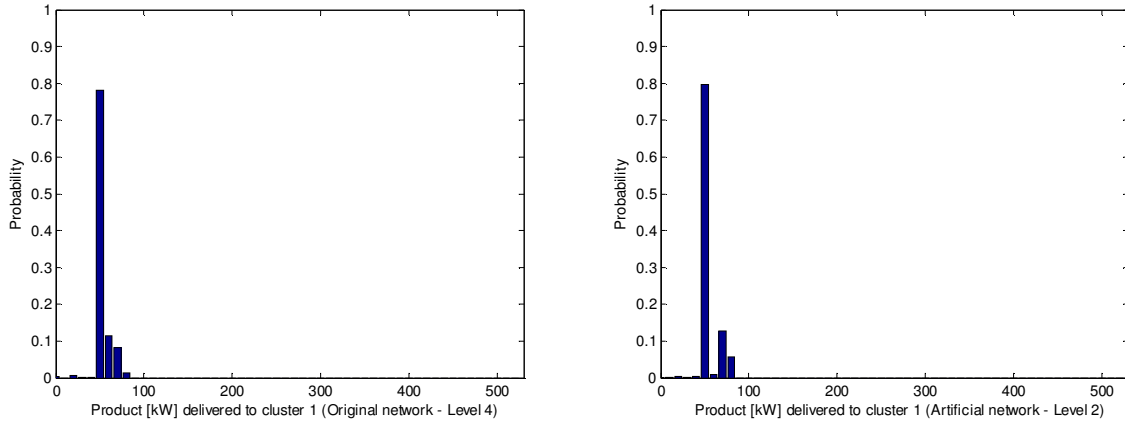


Figure 28: steady state probability distribution of the product delivered to cluster 1 considering the original network (left) and the artificial network at hierarchical clustering level 4 (right).

In this case, the criticality of clusters 8 and 12 at level 4 strongly influences the rankings performed according to the mean (μ_k) and the product not supplied ($idNS_k$) indicators.

At hierarchical level 4, although the rankings are not identical, there is a strong correspondence between the rankings of “groups” of clusters produced according to the mean values for the artificial and real networks: for example, clusters 13, 18, 19 and 29 are always placed at the top of the ranking, clusters 20, 21, 10 are always found at the bottom, whereas clusters 3, 4, 6, 16, 17, 23 occupy the central part of the ranking (Table 6). An exception in this ranking is represented by cluster 5 that even if it requires a demand of 100 kW, it is ranked after some clusters (e.g., clusters 9 and 12) that require higher demands (e.g., 120 kW). This is due to the position of cluster 5 at level 4, which is directly connected to the generation node: thus, this privileged position allows it to receive the product that cannot be allocated anywhere else in the network due to degradation of the arc capacities (in particular, of the arcs contained in clusters 8 and 12 that lead to a reduction of product supplied to themselves and to the connected clusters, i.e., 7, 9, 10, 11 and 13).

With respect to the product not supplied, the simulations carried out on the artificial networks at hierarchical levels 2 and 3 cannot capture the reduction in the performances of the arcs inside clusters 8 and 12 at level 4. Actually, according to the simulation performed on the original network, at level 2 cluster 1 should be the least supplied, whereas at level 3, clusters 2 and 3 should be the most impacted, since they are those that contain the arcs with the worst capacities; those are followed by clusters 4 and 8 that are the farthest from the input source. On the contrary, the simulation of the fictitious networks at level 2 identifies cluster 2 as the most critical and at level 3, clusters 4 and 8 (Table 7). At level 4, the simulations of the fictitious and original networks produce the same results (Table 9): the clusters most supplied are clusters 1, 2, 3, 4, 5, 6 (i.e., those closer to the input source) and those less served are 7, 8, 9, 10, 12, 13 (i.e., those critical and those connected to the critical ones), followed by clusters 14, 15, 22, 28, 29 (i.e., those that are farther from the generation point). Thus, it can be concluded that in those cases where the performance of some arcs is biased towards very low values, a finer level of analysis is needed to capture the global behavior of the network.

The computational cost of the simulations depends on the configuration of the system: the higher the capacities of the arcs, the faster the process to allocate the product by Hierarchical Graph. However, in all the three cases explored, it is evident a considerable reduction in the computational time when the level of detail of the analysis is reduced (i.e., when the network is clustered). In Table 3, the computational times needed to perform the simulations on the clustered networks at hierarchical levels 2, 3 and 4 and on the original network are given for cases 1, 2 and 3 detailed above.

Table 3: Computational time [s] of 10000 Monte Carlo simulations for the analysis of the real and clustered networks at hierarchical levels 2, 3, 4 for cases 1, 2 and 3 considered.

Level	Case 1		Case 2		Case 3	
	REAL	CLUSTERED	REAL	CLUSTERED	REAL	CLUSTERED
2	454	11	1304	11	1392	50
3	454	33	1304	40	1392	78
4	454	110	1304	201	1392	256

Thus, the analyses carried out on clustered networks provide approximate results that are useful in a decision making process for a preliminary design step, reducing considerably the computational cost of the simulation.

Table 4: Mean values, μ_k , of the steady state probability distributions of the product delivered to cluster k for the real and fictitious (clustered) networks at hierarchical level 2, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their demand, D_k .

Level 2

Case 1				Case 2				Case 3				Optimum required - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED			
μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	D_k [kW]	k
1582	1	1592	1	1078	1	1193	1	1026	1	1190	1	1595	1
1867	2	1888	2	1250	2	1316	2	1248	2	1315	2	1895	2

Table 5: Mean values, μ_k , of the steady state probability distributions of the product delivered to cluster k for the real and fictitious (clustered) networks at hierarchical level 3, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their demand, D_k .

Level 3

Case 1				Case 2				Case 3				Optimum required - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED			
μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	D_k [kW]	k
216	8	218	8	144	8	147	8	144	8	147	8	220	8
255	4	258	4	171	4	172	4	169	2	173	4	260	4
277	2	279	2	185	2	195	2	170	4	189	2	280	2
454	7	458	7	303	7	312	7	303	7	311	7	460	7
464	5	468	5	311	5	324	6	311	5	324	6	470	5
478	6	481	6	321	6	330	5	320	6	331	5	485	6
557	1	559	1	393	1	420	1	412	1	423	1	560	1
747	3	752	3	499	3	525	3	444	3	504	3	755	3

Table 6: Mean values, μ_k , of the steady state probability distributions of the product delivered to cluster k for the real and fictitious (clustered) networks at hierarchical level 4, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their demand, D_k .

Level 4

Case 1				Case 2				Case 3				Optimum required - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED		- ranking	
μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	μ_k [kW]	k	D_k [kW]	k
40	13	40	13	26	13	27	13	23	13	25	13	40	13
40	18	40	18	27	18	27	18	27	18	27	18	40	18
40	19	40	19	27	19	27	19	27	19	28	19	40	19
59	29	59	29	39	29	39	29	39	29	39	29	60	29
78	14	79	7	52	14	52	14	47	8	49	7	80	1
78	28	79	8	52	28	52	28	49	7	49	8	80	2
79	1	79	11	53	7	53	7	52	14	49	11	80	7
79	2	79	14	53	11	53	11	52	15	52	14	80	8
79	7	79	15	53	15	53	15	52	28	52	28	80	11
79	8	79	27	53	27	53	27	53	27	53	15	80	14
79	11	79	28	54	1	54	1	55	11	53	27	80	15
79	15	80	1	54	8	54	8	57	1	58	1	80	27
79	27	80	2	55	2	56	2	59	2	60	2	80	28
98	16	99	3	66	16	66	16	66	16	66	16	100	3
99	3	99	16	66	23	66	23	66	23	66	23	100	4
99	6	99	17	67	17	67	17	67	17	67	17	100	5
99	17	99	23	68	3	68	3	71	3	69	3	100	6
99	23	100	4	68	6	69	4	71	4	70	4	100	16
100	4	100	5	70	4	69	6	71	6	70	6	100	17
100	5	100	6	78	5	75	5	73	9	73	9	100	23
118	22	118	22	79	9	79	22	75	12	74	12	120	9
118	24	119	9	79	22	79	26	78	22	78	5	120	12
118	25	119	12	79	24	80	9	79	24	79	22	120	22
118	26	119	24	79	25	80	24	79	25	80	24	120	24
119	9	119	25	79	26	80	25	79	26	80	25	120	25
119	12	119	26	80	12	81	12	84	5	80	26	120	26
286	20	287	20	191	20	192	20	191	20	192	20	290	20
360	21	361	21	242	21	241	21	242	21	242	21	365	21
510	10	510	10	340	10	341	10	292	10	315	10	515	10

Table 7: Product not supplied, $idNS_k$, to the cluster k for the real and fictitious (clustered) networks at hierarchical level 2, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their distance from the input source.

Level 2

Case 1				Case 2				Case 3				Distance - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED		Distance - ranking	
$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$dist_k$	k
0.008	1	0.002	1	0.324	1	0.252	1	0.341	2	0.254	1	1	1
0.015	2	0.004	2	0.340	2	0.306	2	0.357	1	0.306	2	2	2

Table 8: Product not supplied, $idNS_k$, to the cluster k for the real and fictitious (clustered) networks at hierarchical level 3, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their distance from the input source.

Level 3

Case 1				Case 2				Case 3				Distance - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED		Distance - ranking	
$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$dist_k$	k
0.005	1	0.002	1	0.298	1	0.250	1	0.264	1	0.245	1	1	1
0.011	2	0.004	2	0.338	5	0.298	5	0.338	5	0.296	5	2	2
0.011	3	0.004	3	0.338	6	0.304	2	0.340	6	0.324	7	3	2
0.013	5	0.004	5	0.339	2	0.305	3	0.341	7	0.325	2	5	2
0.013	7	0.004	7	0.339	3	0.322	7	0.345	8	0.332	3	7	3
0.014	6	0.008	4	0.341	7	0.332	6	0.346	4	0.332	6	6	4
0.018	8	0.008	6	0.342	4	0.332	8	0.396	2	0.332	8	8	4
0.019	4	0.009	8	0.345	8	0.338	4	0.412	3	0.335	4	4	5

Table 9: Product not supplied, $idNS_k$, to the cluster k for the real and fictitious (clustered) networks at hierarchical level 4, with respect to cases 1, 2 and 3. Right: ranking (in ascending order) of the clusters based on their distance from the input source.

Level 4

Case 1				Case 2				Case 3				Distance - ranking	
REAL		CLUSTERED		REAL		CLUSTERED		REAL		CLUSTERED		$dist_k$	k
$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k	$idNS_k$	k		
0	4	0	1	0.220	5	0.250	5	0.160	5	0.220	5	1	5
0	5	0	2	0.300	4	0.300	2	0.263	2	0.250	2	2	2
0	13	0	4	0.313	2	0.310	4	0.288	1	0.275	1	2	4
0	18	0	5	0.320	3	0.310	6	0.290	3	0.300	4	2	6
0	19	0	6	0.320	6	0.320	3	0.290	4	0.300	6	3	1
0.008	9	0	13	0.325	1	0.325	1	0.290	6	0.300	19	3	3
0.008	12	0	18	0.325	8	0.325	8	0.313	11	0.310	3	3	19
0.010	3	0	19	0.325	18	0.325	12	0.325	18	0.325	18	4	8
0.010	6	0.008	9	0.325	19	0.325	13	0.325	19	0.330	17	4	12
0.010	10	0.008	12	0.330	17	0.325	18	0.330	17	0.333	24	4	18
0.010	17	0.008	24	0.333	12	0.325	19	0.337	21	0.333	25	5	9
0.010	23	0.008	25	0.337	21	0.330	17	0.338	27	0.333	26	5	11
0.013	1	0.008	26	0.338	7	0.333	9	0.340	16	0.337	21	5	13
0.013	2	0.010	3	0.338	11	0.333	24	0.340	23	0.338	15	5	17
0.013	7	0.010	10	0.338	15	0.333	25	0.341	20	0.338	20	6	7
0.013	8	0.010	16	0.338	27	0.338	7	0.342	24	0.338	27	6	10
0.013	11	0.010	17	0.340	10	0.338	10	0.342	25	0.340	16	6	20
0.013	15	0.010	20	0.340	16	0.338	11	0.342	26	0.340	23	6	24
0.013	27	0.010	23	0.340	23	0.338	15	0.350	14	0.342	22	7	25
0.014	20	0.011	21	0.341	20	0.338	20	0.350	15	0.350	14	7	26
0.014	21	0.013	7	0.342	9	0.338	27	0.350	22	0.350	28	8	21
0.017	22	0.013	8	0.342	22	0.340	16	0.350	28	0.350	29	8	23
0.017	24	0.013	11	0.342	24	0.340	21	0.350	29	0.375	13	9	16
0.017	25	0.013	14	0.342	25	0.340	23	0.375	12	0.383	12	9	22
0.017	26	0.013	15	0.342	26	0.342	22	0.388	7	0.388	7	9	27
0.017	29	0.013	27	0.350	13	0.342	26	0.392	9	0.388	8	10	15
0.020	16	0.013	28	0.350	14	0.350	14	0.413	8	0.388	10	10	29
0.025	14	0.017	22	0.350	28	0.350	28	0.425	13	0.388	11	11	14
0.025	28	0.017	29	0.350	29	0.350	29	0.433	10	0.392	9	11	28

5. CONCLUSIONS

In this paper, we have proposed a new representation technique, i.e., the Hierarchical Graph, to analyze the performance of interconnected critical infrastructures (CIs) under a multi-state system-of-systems (SoS) framework. In particular, the robustness of the SoS has been evaluated in terms of the product delivered at steady state to the demand nodes.

First, we have analyzed a small-sized SoS composed by two interconnected CIs (gas and electricity networks) and a supervisory control and data acquisition (SCADA) system and we have evaluated its robustness by Monte Carlo (MC) simulation considering different importance criteria (sequential, proportional, and equal) for the demand nodes. We have

shown that the Hierarchical Graph representation can support this kind of analyses that are useful for decision makers to understand margins of improvement of the SoS to optimize the delivery of product to the demand nodes by changing their importance.

Then, we have evaluated a moderately large-sized power distribution network by adopting a hierarchical clustering algorithm to analyze the SoS at different levels of detail and simplify the Hierarchical Graph representation. In this case, only a proportional importance of the demand nodes has been considered. The results have shown that the Hierarchical Graph can be adopted together with hierarchical clustering algorithms to provide approximate results by analyzing “fictitious” clustered networks instead of the entire large-sized, real network. This can be useful in a first preliminary phase of design of SoS, in order to have satisfactory, physically coherent results with relatively low computational cost.

REFERENCES

- Alata, M., Molhim, M., and Ramini, A. (2008). "Optimizing of Fuzzy C-Means Clustering Algorithm Using GA." *Proceedings of World Academy of Science, Engineering and Technology*, Vol 29, 29, 224-229.
- Barry, L. N. (1995). *Stochastic modeling: analysis and simulation*, McGraw-Hill, New York.
- Clauset, A., Moore, C., and Newman, M. E. J. (2008). "Hierarchical structure and the prediction of missing links in networks." *Nature*, 453(7191), 98-101.
- Fang, Y. P., and Zio, E. (2013). "Unsupervised spectral clustering for hierarchical modelling and criticality analysis of complex networks." *Reliability Engineering & System Safety*, 116, 64-74.
- Ferrario, E., and Zio, E. (2014). "Goal Tree Success Tree-Dynamic Master Logic Diagram and Monte Carlo simulation for the safety and resilience assessment of a multistate system of systems." *Engineering Structures*, 59, 411-433.
- Filippone, M., Camastra, F., Masulli, F., and Rovetta, S. (2008). "A survey of kernel and spectral methods for clustering." *Pattern Recognition*, 41(1), 176-190.
- Gheorghe, A. V., and Schlapfer, M. (2006). "Ubiquity of digitalization and risks of interdependent critical infrastructures." *2006 IEEE International Conference on Systems, Man, and Cybernetics*, Vols 1-6, *Proceedings*, 580-584.
- Gómez, C., Sanchez-Silva, M., Dueñas-Osorio, L., and Rosowsky, D. (2013). "Hierarchical infrastructure network representation methods for risk-based decision-making." *Structure and Infrastructure Engineering*, 9(3), 260-274.
- Haimes, Y. Y. (2012). "Modeling complex systems of systems with Phantom System Models." *Systems Engineering*, 15(3), 333-346.
- Hu, Y. S., and Modarres, M. (2000). "Logic-based hierarchies for modeling behavior of complex dynamic systems with applications." *Fuzzy systems and soft computing in nuclear engineering*, D. Ruan, ed., Springer-Verlag, Berlin Heidelberg.
- IEEE. (2000). "IEEE power and energy society. Distribution test feeders." <http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html>.
- Kröger, W., and Zio, E. (2011). *Vulnerable Systems*, Springer, London.
- Leguizamón, S., Pelgrum, H., and Azzali, S. (1996). "Unsupervised Fuzzy C-means classification for the determination of dynamically homogeneous areas." *Revista SELPER*, 12(12), 20-24.
- Lind, M. (2011a). "An introduction to multilevel flow modeling." *Nuclear safety and simulation*, 2(1), 22-32.
- Lind, M. (2011b). "Reasoning about causes and consequences in Multilevel Flow Models." *Advances in Safety, Reliability and Risk Management*, C. Guedes Soares, ed., CRC Press, 2359-2367.
- McCalley, J. D. (2012). "The DC Power Flow Equations 1.0 Introduction." Department of Electrical and Computer Engineering, Iowa State University.
- Nozick, L. K., Turnquist, M. A., Jones, D. A., Davis, J. R., and Lawton, C. R. (2005). "Assessing the performance of interdependent infrastructures and optimising investments " *International Journal of Critical Infrastructures*, 1(2-3), 144-154.

- Ouyang, M. (2014). "Review on modeling and simulation of interdependent critical infrastructure systems." *Reliability Engineering & System Safety*, 121, 43-60.
- Pepyne, D. L., Panayiotou, C. G., Cassandras, C. G., and Ho, Y. C. (2001). "Vulnerability assessment and allocation of protection resources in power systems." *Proceedings of the 2001 American Control Conference, Vols 1-6*, 4705-4710.
- Porter, M. A., Onnela, J. P., and Mucha, P. J. (2009). "Communities in Networks." *Notices of the American Mathematical Society*, 56(9), 1082-1097.
- Ravasz, E., and Barabasi, A. L. (2003). "Hierarchical organization in complex networks." *Physical Review E*, 67(2).
- Schaeffer, S. E. (2007). "Graph clustering." *Computer Science Review*, 1(1), 27-64.
- von Luxburg, U. (2007). "A tutorial on spectral clustering." *Statistics and Computing*, 17(4), 395-416.
- Zio, E. (2007). *An Introduction to the Basics of Reliability and Risk Analysis*, World Scientific Publishing Co. Pte. Ltd.
- Zio, E. (2009). *Computational methods for reliability and risk analysis*, World Scientific Publishing Co. Pte. Ltd., Singapore.

APPENDIX: UNSUPERVISED SPECTRAL CLUSTERING FOR HIERARCHICAL MODELING

Complex systems are characterized by modularity that allows identifying groups of elements highly interconnected within them and sparsely linked to other dense groups in the network [Porter et al., 2009]. In addition, several studies show that networks often exhibit hierarchical organization [Ravasz and Barabasi, 2003; Clauset et al., 2008]. These features lead to combine hierarchical modeling and clustering analysis to represent complex networks.

Cluster analysis aims at identifying groups of “similar behavior” in their data. Several clustering techniques exist; they can be divided into two categories [Filippone et al., 2008]: *hierarchical*, e.g., dendrogram, that are able to identify structures that can be further decomposed in substructures and so on, and *partitioning*, e.g., K-means, fuzzy c-means, self-organizing maps, neural gas, that obtain a single partition of data without any other decomposition and they are often based on the optimization of an appropriate objective function. Recently, the partitioning methods have been further developed and two big families of algorithms can be identified: kernel and spectral [Filippone et al., 2008].

In this work, we have applied the unsupervised spectral clustering algorithm (USCA) and the Fuzzy c-means (FCM) clustering as in [Fang and Zio, 2013], for its simplicity to implement. By recursively operating the USCA and the FCM a hierarchical structure of the system can be obtained.

The *spectrum* of a graph is defined as the list of eigenvalues of its adjacency matrix $\{ADJ\}$ [Schaeffer, 2007], which rows and columns are the nodes and the cells assume value 1, if the nodes on the rows are connected to the nodes on the columns, otherwise, they assume value 0. It is often more convenient to study the eigenvalues of the Laplacian matrix $\{\mathcal{L}\}$ than those of the adjacency matrix itself [Schaeffer, 2007]. The spectral decomposition of the Laplacian matrix can give useful information about the properties of the graph [Filippone et al., 2008]. The Laplacian matrix is computed as $\mathcal{L} = D - ADJ$, where D is the degree matrix, i.e., the diagonal matrix with the degrees of the nodes on the diagonal; the degrees are obtained by the

sum of the columns of the adjacency matrix). In particular, the normalized Laplacian matrix is adopted in the USCA and it is defined as: $\mathcal{L}_N = \mathbf{D}^{-1/2} \mathcal{L} \mathbf{D}^{-1/2}$.

The main operative steps of the algorithm are reported in the following [Fang and Zio, 2013]:

- 1) Compute the normalized graph Laplacian matrix \mathcal{L}_N .
- 2) Compute the first k eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, and the corresponding eigenvectors $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_k$, of matrix \mathcal{L}_N . The first k eigenvalues are such that they are very small whereas λ_{k+1} is relatively large. All eigenvalues are ordered increasingly.
- 3) The number of clusters is set equal to k , according to the eigengap heuristic theory [von Luxburg, 2007].
- 4) Let $\{\mathbf{U}\}$ be the matrix containing the vectors $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_k$ as columns. Normalize the rows of the matrix $\{\mathbf{U}\}$ to norm 1 obtaining the matrix $\{\mathbf{T}\}$. Denote as [Pepyne et al. 2001], $i = 1, \dots, n$, the vector corresponding to the i -th row of $\{\mathbf{T}\}$, where n is the number of row.
- 5) Resort to the Fuzzy c-means algorithm [Leguizamón et al., 1996; Alata et al., 2008] to partition the points [Pepyne et al. 2001], $i = 1, \dots, n$, into $c = k$ clusters, Cl_1, \dots, Cl_k .

The outputs of the procedure are the clusters C_1, \dots, C_k of the original data points.

Successive USCA and FCM lead to a hierarchical structure of the system where, at the top of the hierarchy, the system is represented by just one node and, at the bottom, by the whole original network. In the middle, each hierarchical level represents a different degree of resolution of the systems (from the top to the bottom it can be seen an increasing quantity of information about the local connectivity) and corresponds to artificial networks that include artificial nodes and links (these last ones are composed by those original network links connecting – in parallel – the original nodes in the clusters forming the artificial nodes).

