Contribution à l'évaluation des risques liés au TMD (transport de matières dangereuses) en prenant en compte les incertitudes

El Abed El Safadi

UNIVERSITÉ GRENOBLE ALPES

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Automatique Productique**

Arrêté ministériel : 7 août 2006

Présentée par

El Abed EL SAFADI

Thèse dirigée par **Jean-Marie FLAUS** et
codirigée par **Olivier ADROT**

préparée au sein du **Laboratoire G-SCOP**
dans **l'École Doctorale EEATS**

Contribution à l'évaluation des risques liés au TMD (Transport de Matières Dangereuses) en prenant en compte les incertitudes

Thèse soutenue publiquement le **09 juillet 2015**,
devant le jury composé de :

M. Gilles DUSSERRE
Directeur de recherche, Ecole des Mines d'Alès, Rapporteur
M. Tarek RAÏSSI
Professeur des universités, CNAM, Rapporteur
Mme. Maria DI MASCOLO
Directrice de Recherche, CNRS, Examinateur
M. Stéphane PAGNON
Inspecteur, DREAL Rhône Alpes, Examinateur
M. Jean-Marie FLAUS
Professeur des universités, UJF, Directeur de thèse
M. Olivier ADROT
Maître de conférences, UJF, Co-directeur de thèse

Résumé

Le processus d'évaluation des risques technologiques, notamment liés au Transport de Matières Dangereuses (TMD), consiste, quand un événement accidentel se produit, à évaluer le niveau de risque potentiel des zones impactées afin de pouvoir dimensionner et prendre rapidement des mesures de prévention et de protection (confinement, évacuation...) dans le but de réduire et maitriser les effets sur les personnes et l'environnement.

La première problématique de ce travail consiste donc à évaluer le niveau de risque des zones soumises au transport des matières dangereuses. Pour ce faire, un certain nombre d'informations sont utilisées, comme la quantification de l'intensité des phénomènes qui se produisent à l'aide de modèles d'effets (analytique ou code informatique). Pour ce qui concerne le problème de dispersion de produits toxiques, ces modèles contiennent principalement des variables d'entrée liées à la base de données d'exposition, de données météorologiques,... La deuxième problématique réside dans les incertitudes affectant certaines entrées de ces modèles. Pour correctement réaliser une cartographie en déterminant la zone de danger où le niveau de risque est jugé trop élevé, il est nécessaire d'identifier et de prendre en compte les incertitudes sur les entrées afin de les propager dans le modèle d'effets et ainsi d'avoir une évaluation fiable du niveau de risque.

Une première phase de ce travail a consisté à évaluer et propager l'incertitude sur la concentration qui est induite par les grandeurs d'entrée incertaines lors de son évaluation par les modèles de dispersion. Deux approches sont utilisées pour modéliser et propager les incertitudes : l'approche ensembliste pour les modèles analytiques et l'approche probabiliste (Monte-Carlo) qui est plus classique et utilisable que le modèle de dispersion soit analytique ou défini par du code informatique. L'objectif consiste à comparer les deux approches pour connaitre leurs avantages et inconvénients en termes de précision et temps de calcul afin de résoudre le problème proposé. Pour réaliser les cartographies, deux modèles de dispersion (Gaussien et SLAB) sont utilisés pour évaluer l'intensité des risques dans la zone contaminée. La réalisation des cartographies a été abordée avec une méthode probabiliste (Monte Carlo) qui consiste à inverser le modèle d'effets et avec une méthode ensembliste générique qui consiste à formuler ce problème sous la forme d'un ensemble de contraintes à satisfaire (CSP) et le résoudre ensuite par inversion ensembliste.

La deuxième phase a eu pour but d'établir une méthodologie générale pour réaliser les cartographies et améliorer les performances en termes de temps du calcul et de précision. Cette méthodologie s'appuie sur 3 étapes : l'analyse préalable des modèles d'effets utilisés, la proposition d'une nouvelle approche pour la propagation des incertitudes mixant les approches probabiliste et ensembliste en tirant notamment partie des avantages des deux

approches précitées, et utilisable pour n'importe quel type de modèle d'effets spatialisé et statique, puis finalement la réalisation des cartographies en inversant les modèles d'effets. L'analyse de sensibilité présente dans la première étape s'adresse classiquement à des modèles probabilistes. Nous discutons de la validité d'utiliser des indices de type Sobol dans le cas de modèles intervalles et nous proposerons un nouvel indice de sensibilité purement intervalle cette fois-ci.

Mots clés : Évaluation des risques, Modélisation incertaine, Méthodes probabiliste et ensembliste, Analyse de sensibilité, Problème de Satisfaction de Contraintes.

Abstract

When an accidental event is occurring, the process of technological risk assessment, in particular the one related to Dangerous Goods Transportation (DGT), allows assessing the level of potential risk of impacted areas in order to provide and quickly take prevention and protection actions (containment, evacuation ...). The objective is to reduce and control its effects on people and environment.

The first issue of this work is to evaluate the risk level for areas subjected to dangerous goods transportation. The quantification of the intensity of the occurring events needed to do this evaluation is based on effect models (analytical or computer code). Regarding the problem of dispersion of toxic products, these models mainly contain inputs linked to different databases, like the exposure data and meteorological data. The second problematic is related to the uncertainties affecting some model inputs. To determine the geographical danger zone where the estimated risk level is not acceptable, it is necessary to identify and take in consideration the uncertainties on the inputs in aim to propagate them in the effect model and thus to have a reliable evaluation of the risk level.

The first phase of this work is to evaluate and propagate the uncertainty on the gas concentration induced by uncertain model inputs during its evaluation by dispersion models. Two approaches are used to model and propagate the uncertainties. The first one is the set-membership approach based on interval calculus for analytical models. The second one is the probabilistic approach (Monte Carlo), which is more classical and used more frequently when the dispersion model is described by an analytic expression or is defined by a computer code. The objective is to compare the two approaches to define their advantages and disadvantages in terms of precision and computation time to solve the proposed problem. To determine the danger zones, two dispersion models (Gaussian and SLAB) are used to evaluate the risk intensity in the contaminated area. The risk mapping is achieved by using two methods : a probabilistic method (Monte Carlo) which consists in solving an inverse problem on the effect model and a set-membership generic method that defines the problem as a constraint satisfaction problem (CSP) and to resolve it with an set-membership inversion method.

The second phase consists in establishing a general methodology to realize the risk mapping and to improve performance in terms of computation time and precision. This methodology is based on three steps :

- Firstly the analysis of the used effect model.
- Secondly the proposal of a new method for the uncertainty propagation based on a mix between the probabilistic and set-membership approaches that takes advantage of both approaches and that is suited to any type of spatial and static effect model.

– Finally the realization of risk mapping by inversing the effect models. The sensitivity analysis present in the first step is typically addressed to probabilistic models. The validity of using Sobol indices for interval models is discussed and a new interval sensitivity indice is proposed.

Key-words : Risk assessment, uncertain modeling, probabilistic and set-member ship method, sensitivity analysis, constraint satisfaction problem.

Remerciements

Tout travail réussi dans la vie nécessite en premier lieu, la bénédiction de **Dieu** et ensuite l'aide et le support de plusieurs personnes. Je tiens donc à remercier et à adresser ma reconnaissance à toute personne qui m'a aidé de loin ou de près afin de réaliser ce travail. On m'a appris que les remerciements les plus longs sont parfois les moins sincères. Je serai donc bref.

Je remercie chaleureusement mon directeur de thèse, Jean-Marie Flaus, pour m'avoir encadré pendant tout le déroulement de celle-ci. Je le remercie de ses conseils avisés, son excellence scientifique ainsi que pour ses remarques intéressantes qui ont contribué à faire avancer et améliorer ce travail. Je tiens à remercier de tout mon cœur Olivier Adrot d'avoir accepté de co-diriger cette thèse. Je le remercie pour son soutien moral et technique ainsi que pour sa patience qui m'ont été très précieux.

J'exprime toute ma reconnaissance à Mr Gilles Dusserre et Mr Tarek Raissi, pour l'honneur qu'ils m'ont fait en acceptant de rapporter cette thèse. Je voudrais également remercier Mme Maria Di Mascolo et Mr Stéphane Pagnon pour l'intérêt qu'ils ont porté à ces travaux en acceptant de les examiner.

Enfin je tiens à remercier du fond du coeur tous mes proches, mes amis et ma famille pour tous les merveilleux moments hors thèse et pour le soutien qu'ils m'ont toujours apporté.

A mes chers parents...

Table des matières

Liste des figures

Liste des tableaux

Introduction

Historiquement, depuis le début du XXème siècle, les progrès scientifiques et technologiques, qui ont tant concouru à améliorer la qualité de vie, ont dans le même temps intensifié les risques qu'ils pouvaient faire peser. L'ère de la révolution industrielle a attiré l'attention sur l'intérêt de maîtriser ces risques, comme le risque d'explosion ou d'émission de substances dangereuses, et sur la nécessité de prévenir les accidents technologiques afin de réduire le plus possible leurs effets et leurs conséquences sur la population et l'environnement. Aujourd'hui, notre monde est donc globalement plus risqué, que ce soit à cause de la nature et de la diversité des évènements dangereux qui peuvent survenir, que d'une meilleure connaissance des phénomènes qui, entrainant une plus grande sensibilisation, induit aussi un niveau d'acceptation du risque plus faible, tant au niveau du tout à chacun que de la réglementation. C'est dans le but d'améliorer la sécurité des personnes et de protéger l'environnement que les risques liés aux Transports de Matières Dangereuses (TMD) sont devenus des enjeux importants pour la population et l'état, notamment en zone fortement urbanisée. Une matière dangereuse est une substance qui peut représenter un danger pour l'homme, les biens ou l'environnement, en raison de son caractère inflammable, de sa toxicité, de son caractère corrosif, ou de sa radioactivité. Les risques associés aux TMD sont donc consécutifs à un accident se produisant lors du transport d'une telle matière.

Dans le processus d'évaluation des risques technologiques, l'étape essentielle consiste dans l'évaluation de l'intensité des risques quand un événement accidentel se produit, c'est-à-dire quantifier les effets ou impacts associés, afin de répondre rapidement et de prioriser les actions de secours pour la protection de la population et de l'environnement. Soumise à des critères parfois subjectifs et à des contraintes à la fois opérationnelles et économiques, l'évaluation des risques se doit d'être le plus possible transparente, répétable, systématique. De ce fait, l'évaluation de l'intensité des risques doit être à la fois précise et fiable, tout en étant réalisée en un temps raisonnable.

L'évaluation de l'intensité d'un risque technologique peut se faire en utilisant un modèle d'effets capable d'estimer d'un point de vue quantitatif les effets induits par le phénomène

dangereux considéré, afin de déterminer la zone géographique de danger où l'intensité du risque est jugée trop élevée. Ce modèle, plus ou moins complexe, est qualifié d'analytique si une expression formelle est manipulée. Sinon, il s'agit d'un modèle de type boite noire et on parle alors de code informatique (ou de logiciel). De manière générale, l'effet peut être toxique, thermique ou explosif et dans cette étude, la dispersion d'un produit chimique toxique est plus spécifiquement considérée. Ainsi, des modèles de dispersion atmosphérique sont utilisés pour estimer la concentration en gaz à une position géographique donnée émise à partir de sources telles que les installations industrielles, les véhicules TMD ou les rejets chimiques accidentels. Les modèles de dispersion peuvent être classés en trois catégories, que sont les modèles de dispersion gaussiens, les modèles de type intégral et les modèles 3D reposant sur la mécanique des fluides (CFD). Tous ces modèles de dispersion ont en commun de faire intervenir de nombreuses entrées, comme celles liées aux données d'exposition, aux données météorologiques, aux caractéristiques du produit émis et à la géolocalisation de la source d'émission. Ces entrées peuvent être mesurées, estimées ou déduites d'une connaissance a priori, mais la plupart d'entre elles ne sont connues qu'avec imprécision, d'où la notion d'incertitude de modèle qui a pour objectif de décrire ce que l'on connait mal.

Les incertitudes de modèles influent naturellement sur les résultats de l'évaluation des risques lors de la détermination de la zone géographique de danger où la concentration en gaz est plus grande qu'un seuil réglementaire donné, afin d'éviter par exemple les effets irréversibles ou létaux sur la santé. Par conséquent, pour avoir une évaluation fiable des effets et correctement réaliser cette cartographie, il est nécessaire d'identifier et de prendre en compte les incertitudes sur les entrées des modèles de dispersion. Ce travail conduit à générer une seconde zone, dite d'incertitudes, où la concentration en gaz peut être plus élevée ou plus faible que le seuil donné selon les valeurs (inconnues) réellement prises par les entrées incertaines du modèle.

Dans ce contexte, la première problématique abordée dans cette thèse consiste à évaluer le niveau de risque des zones soumises au transport des matières dangereuses, tandis que la deuxième problématique réside dans la prise en compte des incertitudes affectant certaines entrées dans les modèles de dispersion utilisées dans cette évaluation. Cette seconde problématique est développée en deux temps. Cette thèse vise tout d'abord à évaluer et propager les incertitudes sur la concentration en gaz induites par les grandeurs d'entrée incertaines lors de l'évaluation des modèles de dispersion. L'objectif est ensuite de traiter le problème de cartographie en évaluant les zones de danger et d'incertitudes, qui sont utilisées pour évaluer l'intensité des risques dans la zone impactée. Cela nous a conduit finalement à proposer une méthodologie générale pour réaliser les cartographies dans le but d'améliorer les performances en termes de temps de calculs et de précision par rapport à d'autres méthodes plus classiques.

Organisation de la thèse

Ce document est structuré en quatre chapitres principaux et s'achève par une conclusion générale.

Le premier chapitre énonce le contexte général et la problématique de notre travail. Il est organisé en cinq parties. La première partie présente les concepts généraux de la gestion des risques, notamment les risques de TMD, avec un rappel sur les statistiques et les événements marquants associés. La seconde partie présente le projet GEOFENCING-MD, qui soutient ces travaux de recherche et définit le cadre applicatif de ce travail. La troisième partie pose la problématique associée à l'évaluation du niveau de risque des zones soumises au TMD en tenant compte des différentes sources d'incertitudes correspondantes. La quatrième partie introduit les différents types de modèles de dispersion, les façons de représenter les sources d'incertitudes et les différentes approches utilisées dans la propagation des incertitudes, ainsi que l'analyse de sensibilité. La dernière partie détaille les objectifs des travaux proposés pour traiter les problématiques posées.
Dans le chapitre 2, nous présentons les différentes techniques retenues dans ce travail dont l'utilisation est devenue relativement classique dans le cadre de la propagation des incertitudes dans un modèle. Dans la première partie, nous présentons deux approches : l'approche probabiliste reposant sur la méthode de Monte Carlo et l'approche ensembliste basée sur les techniques d'analyse par intervalles. Concernant la méthode de Monte Carlo, nous décrivons le processus utilisé pour la propagation des incertitudes et pour la méthode d'analyse par intervalles nous présentons un rappel du calcul par intervalles ainsi que des principaux problèmes rencontrés lors de la manipulation des intervalles. La deuxième partie de ce chapitre est consacrée à une application à la dispersion atmosphérique où l'objectif est d'estimer l'intervalle de confiance sur la concentration en gaz émis, évalué à l'aide d'un modèle gaussien compte tenu des différentes entrées incertaines prises en compte.
Dans le chapitre 3, nous abordons la problématique de la génération des cartographies, c'est-à-dire la détermination des zones de danger et de sécurité lorsqu'un accident TMD survient. Pour réaliser les cartographies, deux modèles de dispersion, le premier plus simple et analytique (modèle gaussien), le second plus sophistiqué et de type code informatique (SLAB), sont utilisés pour évaluer l'intensité des risques dans la zone impactée. Dans la première partie de ce chapitre, nous abordons la réalisation des cartographies avec la méthode probabiliste (Monte Carlo) qui consiste à inverser le modèle de dispersion tandis qu'avec la méthode ensembliste la réalisation des cartographies consiste à formuler ce problème sous la forme d'un ensemble de contraintes à satisfaire (CSP) et à le résoudre ensuite par inversion ensembliste en tenant compte des incertitudes. Nous expliquons comment il est possible d'utiliser l'approche générique CSP couplée à

une méthode d'inversion ensembliste pour résoudre d'autres types de problèmes comme l'identification des défauts dans les capteurs qui mesurent la concentration en gaz ou la localisation de la source de fuite. La deuxième partie est consacrée à une application à la dispersion atmosphérique qui vise à montrer les résultats obtenus dans la réalisation des cartographies avec les deux modèles étudiés (gaussien et SLAB) en utilisant les deux approches (probabiliste et ensembliste) lorsque le modèle s'y prête.

Dans le chapitre 4, nous présentons dans la première partie une méthodologie générale et originale pour réaliser les cartographies et pour améliorer les performances en termes de temps du calculs et de précision. Cette méthodologie est divisée en 3 étapes :

- La première étape consiste à analyser préalablement les modèles de dispersion en fonction de leur nature (analytique, par code) en réalisant une analyse de sensibilité globale pour déterminer quelles sont les entrées incertaines les moins influentes qu'il est raisonnable de fixer à une valeur nominale pour par exemple réduire le nombre de simulations du modèle de dispersion, en étudiant la monotonie des modèles par rapport aux entrées incertaines pour être capable d'évaluer l'intervalle de sortie même pour un modèle de type code informatique et avoir des informations sur la multi-occurrence de ces mêmes entrées dans ces modèles pour mieux maitriser le pessimisme en cas de calcul par intervalles. Nous proposons notamment dans cette optique la construction d'un nouvel indice de sensibilité, intervalle cette fois-ci, plus spécifiquement adapté aux modèles par intervalles.
- La deuxième étape, qui s'appuie sur les résultats de l'étape précédente, propose une nouvelle approche pour propager les incertitudes mixant les approches probabiliste et ensembliste. Cette approche dite mixte, reposant initialement sur la simulation de Monte-Carlo, est utilisable pour tout type de modèles de dispersion spatialisé et statique et consiste à tirer des ensembles de valeurs (intervalles) plutôt que de simples valeurs pour appréhender plus de situations possibles, tout en utilisant un nombre de tirages réduit.
- La dernière étape repose sur un algorithme d'inversion utilisant les résultats de l'étape précédente pour réaliser les cartographies en inversant les modèles de dispersion.

  La deuxième partie de ce chapitre est consacrée à une application à la dispersion atmosphérique en utilisant les deux modèles (gaussien et SLAB). Nous appliquons la méthodologie proposée pour en montrer les avantages et comparons les résultats obtenus en termes de cartographies avec ceux obtenus pour les approches probabiliste et ensembliste. Nous nous intéressons plus spécifiquement au problème de précision et de temps de calculs.

Pour terminer, la partie conclusion propose un bilan général de ce travail et évoque, bien évidemment, quelques perspectives de recherche.

Chapitre 1

Risques liés au Transport de Matières Dangereuses

1.1 Contexte

1.1.1 La gestion des risques : principes généraux

Le risque peut être défini par la confrontation d'un aléa (phénomène naturel ou technologique dangereux) et d'une zone géographique où existent des enjeux qui peuvent être humains, économiques ou environnementaux. La gestion des risques peut être définie comme un ensemble d'activités coordonnées dans le but de diriger et piloter un état, une entreprise ou d'autres formes d'organisations en vue de réduire le risque à un niveau jugé tolérable ou acceptable. La gestion des risques correspond donc à un domaine scientifique transverse, plongeant ses racines dans toutes les disciplines existantes, la nature de l'organisation et l'origine des risques pouvant être diverses et variées (risque technologique, naturel, professionnel, domestique, routier,...).
Mettre en œuvre cette gestion des risques selon la norme ISO 31000, consiste en l'application systématique de principes, politiques, procédures et pratiques pour les tâches d'identification, d'analyse, d'évaluation et de traitement du risque ainsi que pour les activités de communication, de concertation, d'établissement du contexte, de surveillance et de revue des risques [7–11].
La première étape dans cette procédure est l'analyse des risques qui vise à identifier les sources de danger étudiées et les situations associées qui peuvent conduire à des dommages sur les personnes, l'environnement ou les biens. Lors des étapes suivantes, les conséquences potentielles des phénomènes dangereux sont estimées classiquement à l'aide de grilles de cotations ou de logiciels de simulation plus sophistiqués reposant sur

des modèles mathématiques. De même, les fréquences d'exposition ou les probabilités d'occurrence d'accidents sont estimées à partir de bases de données d'accidents ou de jugements d'experts. D'autres critères tels que la vulnérabilité représentant les enjeux et le niveau de maîtrise des risques (moyens organisationnels, humains ou techniques individuels ou collectifs en place) doivent aussi être pris en compte. Finalement, l'analyse de risque fournit un résultat caractérisant le niveau du risque (et si possible le degré de confiance dans cette évaluation). L'évaluation des risques consiste ensuite à réaliser une comparaison du niveau de risque à des seuils issus de critères de décision déjà définis dans l'étape d'établissement du contexte, afin d'étudier la nécessité de mettre en place des actions correctives. La réduction du risque (ou maîtrise du risque) désigne l'ensemble des actions qui visent à diminuer la probabilité (prévention) ou la gravité (protection) des dommages associés à un risque particulier.
La méthodologie d'évaluation et de réduction des risques est présentée sur la figure (1.1).

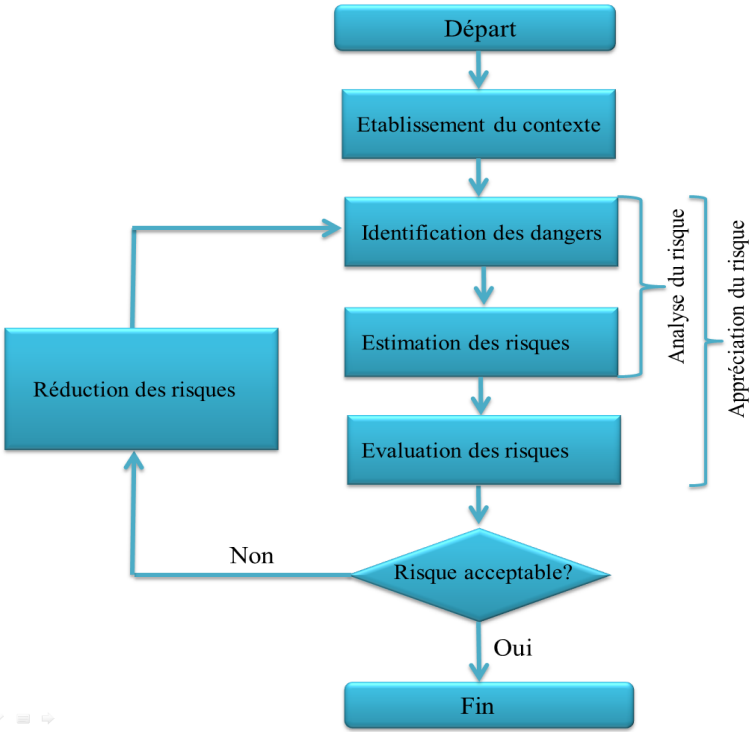

FIGURE 1.1: Processus itératif d'évaluation et de réduction des risques

Un risque acceptable est un risque qui permet de préserver à la fois l'activité de l'organisation (chômage technique, perte d'une partie de l'installation,...), la sécurité des personnes (salariés, habitants du voisinage,...), ainsi que l'environnement au sens large (infrastructure, utilités, pollution,...), de satisfaire aux lois, réglementations et normes en vigueur (conséquences pénales), de se prémunir des coûts financiers ou de perte d'image (opinion public) liés à l'accident,...
Le but de l'évaluation est d'aider les décideurs à déterminer les risques nécessitant un

plan d'actions et la priorité dans la mise en œuvre de ces actions. Cette hiérarchisation sera déterminée en fonction [12] :

- du résultat de la comparaison permettant de déterminer si le risque est négligeable, acceptable, rédhibitoire,
- de la tolérance au risque des différentes parties prenantes et des obligations légales,
- des coûts financiers et des plans d'actions existants.

1.1.2 Risque lié au Transport de Matières Dangereuses (TMD)

Une marchandise dangereuse est une matière ou un objet qui peut présenter des risques pour l'homme, les biens et/ou l'environnement à cause de ses caractéristiques physico-chimiques (toxicité, réactivité...) et physiologiques. Ces marchandises peuvent être transportées sous forme liquide (hydrocarbures, chlore, propane,...), solide (explosifs, nitrate d'ammonium,...) et gaz (gaz de houille comprimé,...). Ces substances ont souvent une concentration et une agressivité supérieures à celles des usages domestiques.
Chaque jour, des marchandises dangereuses (MD) sont transportées selon des modalités différentes partout dans le monde. Elles sont utilisées pour des activités liées à l'énergie, la fabrication ou lors d'activités commerciales. Le risque de transport de matières dangereuses (TMD) désigne les conséquences d'un accident se produisant lors du transport de ces marchandises par voie routière, ferroviaire, voie d'eau ou canalisations. Quand un événement dangereux se produit, causé par une erreur humaine, une dégradation de l'état du moyen de transport..., il est nécessaire d'évaluer le niveau de risque potentiel des zones impactées afin de pouvoir prendre des mesures de prévention et de protection (confinement ou évacuation de la population avoisinante, détournement des moyens de TMD proches pour éviter le sur-accident,...) dans le but de réduire et maitriser les effets sur les personnes et l'environnement.

1.1.3 Principaux risques liés aux matières dangereuses

On distingue cinq principales catégories de risques liés au transport de matières dangereuses [13, 14] :

- **le risque explosion**, qui est issu d'une combustion rapide engendrant une quantité importante de gaz à une température, une pression et une vitesse d'expansion tellement élevées qu'il en résulte des dommages aux alentours. Un périmètre de sécurité sera mis en place à proximité du sinistre dans un rayon de plusieurs centaines de mètres ;

- **le risque d'incendie**, qui correspond à une réaction résultant de la présence de plusieurs facteurs (source de chaleur, comburant, combustible) et qui provoque un dégagement important de chaleur, ayant pour conséquences des brûlures ou des blessures souvent très graves ;
- **le risque toxique**, qui peut provoquer l'empoisonnement, voire la mort, par inhalation, contact ou ingestion d'une substance chimique toxique suite à une fuite de produits toxiques. La dispersion de la matière dangereuse peut se faire dans l'air, l'eau et/ou le sol. Dans le cas aérien, le nuage toxique va s'éloigner du lieu de l'accident au gré des vents actifs à ce moment là ;
- **le risque de radioactivité** dans le cas de matières émettant des rayonnements dangereux qui peuvent atteindre tous les êtres vivants ;
- **le risque infectieux**, qui peut provoquer des maladies graves chez les êtres vivants. Ce risque est spécifique aux matières contenant des micro-organismes infectieux tels que les virus, les bactéries,...

1.1.4 Modes de transport des matières dangereuses

Les principaux produits dangereux transportés en France sont les produits pétroliers (environ 60 millions de tonnes par an) et les produits chimiques (environ 25 à 30 millions de tonnes par an) ce qui représente un tonnage conséquent. Globalement il existe principalement 5 modes de transport [14] :

- **le transport routier**. Le transport en camion est le plus fréquemment utilisé puisqu'il représente environ 76 % du tonnage transporté sur l'ensemble de la France. Le mode routier est plus rapide, plus flexible et plus rentable économiquement.
- **le transport ferroviaire**. Il permet de transporter les marchandises dangereuses par le biais de wagons et représente environ 16 % du tonnage transporté au niveau national. le transport ferroviaire peut être utilisé comme moyen combiné avec le transport routier.
- **le transport par canalisations**. Ce type de transport se compose d'un ensemble de conduites sous pression, de diamètres variables, qui sert à déplacer de façon continue ou séquentielle des fluides ou des gaz liquéfiés. Ce mode de transport est utilisé pour transporter du gaz naturel (gazoducs), des hydrocarbures liquides ou liquéfiés (oléoducs, pipelines), certains produits chimiques (éthylène, propylène,...).
- **le transport maritime**. Le transport maritime est en véritable progression grâce au le commerce international. L'avantage de ce mode est la possibilité de recouvrir les zones de livraison les plus étendues du globe et la grande capacité de transport.
- **le transport fluvial**. En France, le tonnage de marchandises transportées par voie fluviale représente environ 3% du tonnage total.

1.1.5 Causes possibles d'un accident de TMD

Les principales causes d'accident en TMD sont les suivantes [15] :

- **une explosion** peut être due à un choc avec production d'étincelles notamment pour les citernes qui transportent des produits inflammable sous forme gazeuse, liquide ou solide, à la mise en contact de plusieurs produits incompatibles qui vont générer une réaction incontrôlée ou encore à la mise à feu inopinée de munitions ou d'explosifs.
- **un incendie** est lié à la présence de produits inflammables et peut être provoqué par l'échauffement anormal d'un organe du véhicule, un choc mécanique ou électrostatique avec production d'étincelles ou encore une inflammation accidentelle.
- **un nuage toxique** peut provenir d'une fuite de produit chimique toxique ou d'un dégagement de fumées toxiques dues à un incendie.

1.1.6 Statistiques et événements marquants

En se référant aux données de la base ARIA du Bureau Analyses des Risques et Pollutions Industriels (BARPI), 3280 accidents survenus durant le transport de matières dangereuses sont recensés entre 1992 et 2011. Leur répartition par mode de transport est indiquée sur la figure 1.2.

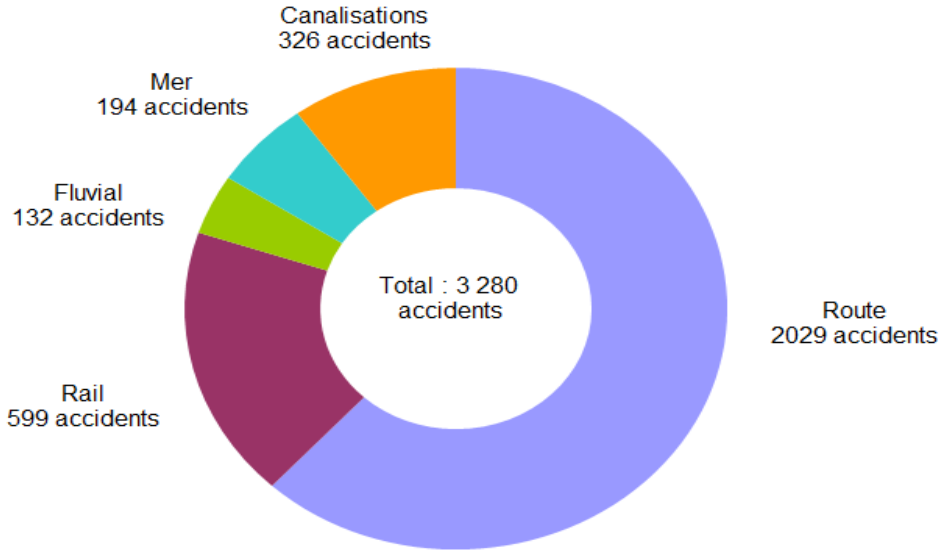

FIGURE 1.2: Nombre d'accidents liés au transport de matières dangereuses recensés par mode de transport entre 1992 et 2011

En proportion, 62% des accidents liés au TMD se produisent sur le réseau routier, entraînant alors la perte des produits dans 33% des cas concernés, des incendies dans 10% des cas ou des explosions dans 3% des cas. Le reste des accidents se produit pour 18% sur voie ferroviaire, 10% sur des canalisations, 6% en mer et 4% sur les voies fluviales.

Le transport par route est le mode le plus utilisé pour transporter les matières dangereuses et le plus exposé aux accidents, ce qui explique la fréquence plus importante des accidents. Les causes sont diverses : faute de conduite du conducteur ou d'un tiers, mauvais état du véhicule, mauvais état des routes, météorologie défavorable. En France, le transport de produits pétroliers (76% du nombre total de trajets) et de gaz (14%), qui sont donc les marchandises dangereuses les plus transportées, est essentiel à l'approvisionnement en fuel, essence, gazole ou GPL des millions de ménages français sur l'ensemble du territoire national.

La base de données Gaspar (Gestion Assistée des Procédures Administratives relatives aux Risques naturels et technologiques) de la Direction Générale de la Prévention des Risques (DGPR) recense **12 000 communes françaises soumises au risque lié au transport de matières dangereuses**.
Les régions les plus exposées sont celles comportant de grands axes routiers et autoroutiers et situées le long des corridors fluviaux : Rhin, Rhône, Seine, Moselle, Escaut. Six régions concentrent plus de la moitié des communes classées à risque lié au transport de matières dangereuses : Nord-Pas-de-Calais, Rhône-Alpes, Lorraine, Poitou-Charentes, Midi-Pyrénées, Haute-Normandie.

La figure 1.3 représente une cartographie des risques TMD en Rhône-Alpes.

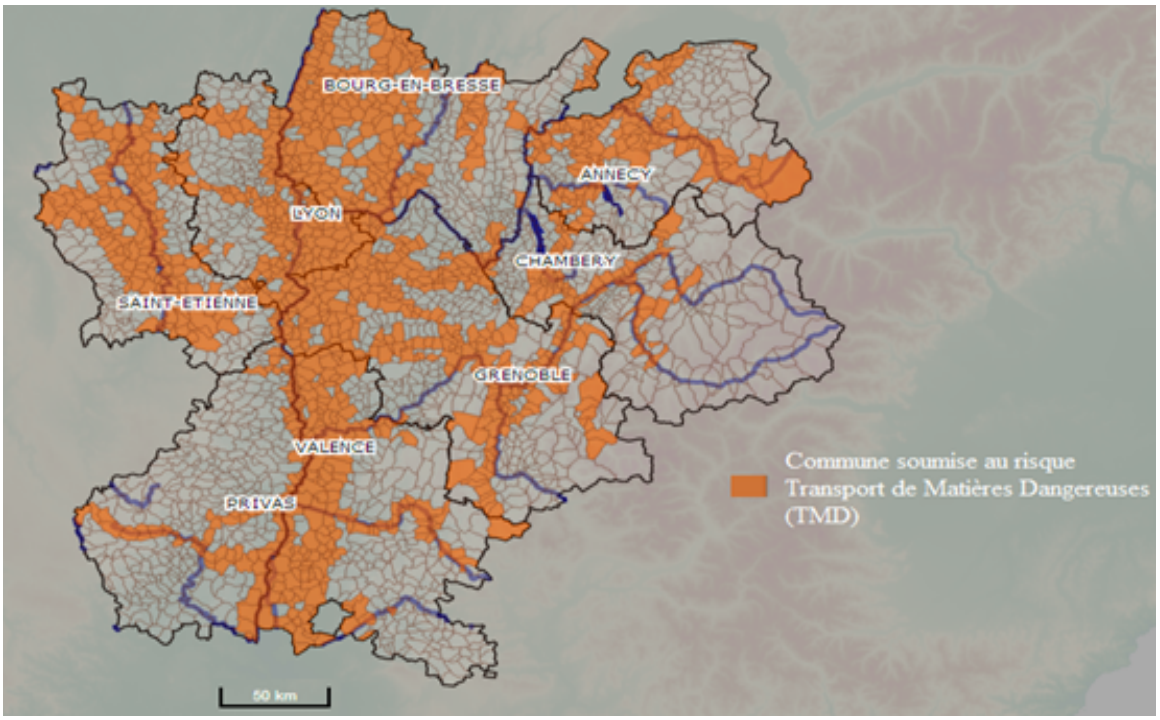

FIGURE 1.3: Carte des risques TMD en Rhône-Alpes, « Source : L'institut des Risques Majeurs (IRMa) »

Date	Localisation	Type d'accident	Conséquences
1997	Port Sainte Foy (Dordogne)	Collision entre un camion-citerne transportant des produits pétroliers et un autorail [15]	12 morts et 43 blessés
1997	Granieu	Renversement d'un camion-citerne transportant du propane au centre d'un village [14]	Un périmètre de sécurité de 250 mètres est mis en place et 15 personnes sont évacuées
2001	Grenoble	Renversement d'un camion transportant 23 tonnes de toluène diisocyanate au niveau du Pont de Catane à Grenoble [14]	Un périmètre de sécurité a été délimité, la circulation interrompue
2008	Longvic	Collision entre un camion et un train de fret transportant des hydrocarbures	Une fuite de carburant du réservoir du camion a produit une pollution dans un secteur sensible du fait de la présence de la nappe phréatique proche de la surface
2012	Rouen	Un chauffeur transportant 9 m^3 d'essence et 22 m^3 de gazole perd le contrôle de son véhicule, l'attelage franchit la glissière centrale et percute un poids-lourd circulant en sens inverse.	Les carburants qui s'écoulent de la citerne éventrée s'enflamment et propagent l'incendie à des chemins de câbles sous le tablier, les secours évacuent les 2 chauffeurs, 4 forains et 1 policier sont blessés.

TABLE 1.1: Exemples d'accidents de TMD en France

En 2012, 131 événements impliquant des véhicules routiers de transport de matières dangereuses (TMD) non radioactives ont été enregistrés dans la base ARIA. Quelques exemples sont donnés dans le tableau 1.1.
Si les accidents de transport de matières dangereuses (TMD) restent heureusement peu fréquents, ils peuvent être très graves : nombreuses victimes, importants dommages matériels et environnementaux, pertes économiques. Tous ces éléments, nous ont donc amenés à considérer dans la suite de ce document plus spécifiquement le risque de transport de matière dangereuses par voie routière.

1.2 Présentation du projet Geofencing

Le projet GEOFENCING-MD de gestion des transports de matières dangereuses en agglomération urbaine est porté depuis juin 2011 par le pôle de compétitivité LUTB (Lyon Urban Truck & Bus).

Nous avons vu que les risques engendrés par le transport de matières dangereuses (TMD) peuvent être graves avec des conséquences parfois désastreuses pour la société en général (citoyens, environnement, infrastructures), particulièrement en milieu urbain. L'objectif de ce projet est de développer un outil télématique de supervision et de gestion des TMD sur une agglomération urbaine, afin d'en améliorer le fonctionnement logistique et

de limiter les risques lors des approvisionnements et du transit des matières dangereuses sur son territoire. Cet outil doit être utilisable par tous les acteurs impliqués dans le TMD (chargeurs, transporteurs, autorités publiques, groupes d'intervention, opérateurs d'infrastructures, etc...)[16].

Ce projet inclut la collecte de données de capteurs embarqués sur les véhicules de transport de matières dangereuses permettant par exemple leur géolocalisation et leur traçabilité en temps réel, ainsi que la transmission de ces données vers un serveur unique gérant l'accès aux informations collectées. Ainsi, la centralisation des données permet d'imaginer le fonctionnement de ce serveur unique comme une « tour de contrôle » (par analogie au contrôle du trafic aérien) en charge du bon transit des matières dangereuses dans l'agglomération urbaine qu'elle supervise. Cette tour de contrôle se compose d'un serveur centralisé et de systèmes de communication, de l'ensemble des applications nécessaires à la gestion du transport ainsi que des interfaces utilisateurs. Elle définit une architecture technique et fonctionnelle d'un système d'information de suivi et de supervision des TMD en temps réel.

Ce travail de thèse s'inscrit dans le projet GEOFENCING-MD. Plus précisément, il consiste à suivre en ligne et en temps réel les différents moyens de transport, essentiellement les camions, transportant des matières dangereuses qui vont donc être géolocalisés à l'aide d'un GPS. Dans le cas où un accident se produit sur l'un des camions, le but consiste à évaluer à l'aide d'une carte les zones de danger et de sécurité où le niveau de risque est jugé rédhibitoire, respectivement acceptable, autour du camion accidenté afin de mettre en place des barrières de sécurité (fence) pour dérouter les véhicules soumis au TMD proches du périmètre défini et réduire la probabilité de survenue d'un sur-accident [17]. Ainsi, durant cette thèse, une contribution dans la réalisation d'un démonstrateur de suivi et de supervision des TMD a été développée, essentiellement concernant l'évaluation du niveau de risque à travers la réalisation de cartographies qui évaluent les zones de danger et de sécurité. Une description plus détaillée de cette contribution est fournie en annexe A.

De nombreux projets/études ont abordé le thème de suivi télématique et de supervision en temps réel du transport de matières dangereuses, que ce soit par route, par rail, par mer ou par voie navigable. Ainsi, plusieurs projets européens (ARTS VISU TMD, GOODROUTE, GRAIL CHEM, MITRA, MENTORE, METIS, M-TRAD), et français (TR@IN MD, DETRACE, SISTEMS,...) ont traité ce sujet, pour certains au-delà du domaine des matières dangereuses. GEOFENCING-MD se distingue par l'utilisation de capteurs embarqués. Les informations disponibles sur ces différents projets et systèmes étudiés sont de nature très diverse et parfois peu détaillées sur certains aspects techniques. En règle générale, des projets de recherche mettent en avant des systèmes très

technologiques, à la pointe de l'innovation quant aux technologies de l'information et de la communication.

Aux États-Unis, plusieurs rapports d'études d'organismes gouvernementaux préconisent une approche globale du suivi et de la surveillance des TMD en temps réel, notamment en développant des applications basées sur le geofencing (concept de géolocalisation qui permet de surveiller à distance la position et le déplacement d'un objet pour prendre des mesures si celui-ci s'écarte ou rentre dans une zone virtuelle prédéfinie). Le "Department of Homeland Security" accorde ainsi des subventions aux projets qui mettent en oeuvre des systèmes de suivi intégrant cette fonctionnalité.

Les résultats obtenus dans ces différents projets montrent l'intérêt de mettre en œuvre un suivi et une supervision du transport de matières dangereuses dans le cadre d'une approche systémique. Cependant, il n'existe pas aujourd'hui, en Europe, d'architecture technique et fonctionnelle d'un système d'information de suivi et de supervision des transports de matières dangereuse en temps réel et multimodale. GEOFENCING-MD vise à proposer une partie de la solution pour un tel système.

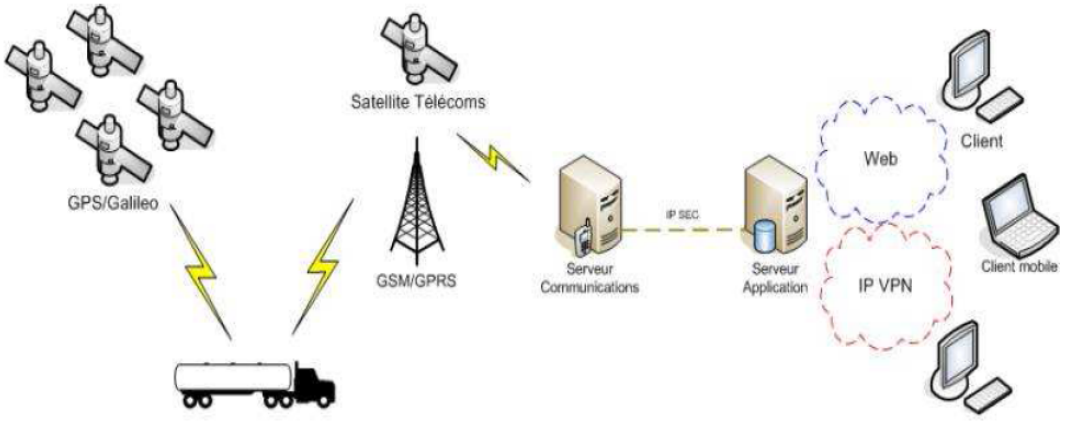

FIGURE 1.4: Architecture technique et fonctionnelle du système d'information

1.3 Problématique

En lien avec les objectifs du projet geofencing précisés dans la section précédente, la première problématique de cette thèse consiste à évaluer le niveau de risque des zones soumises au transport des matières dangereuses [18] en fonction des données reçues du véhicule. L'évaluation des risques TMD est basée sur différents critères qui prennent en compte la gravité des conséquences de l'accident, la vulnérabilité de la zone touchée et la probabilité d'occurrence des événements considérés. Pour faire cette évaluation des risques, on a donc besoin d'un certain nombre d'informations et de modèles d'effets, classiquement utilisés pour ce qui concerne les problèmes de dispersion de produits toxiques,

d'incendie ou encore d'explosion. D'autres informations comme la vulnérabilité de la zone impactée témoignant des enjeux qui avoisinent le lieu de l'accident sont nécessaires. Ces enjeux peuvent être des enjeux humains, à savoir les personnes physiques directement ou indirectement exposées aux conséquences de l'accident, des enjeux économiques comme les entreprises voisines du lieu de l'accident, les routes, les voies de chemin de fer... et des enjeux environnementaux. Parfois les conséquences d'un accident peuvent également avoir un impact sanitaire (pollution des nappes phréatiques par exemple) et, par voie de conséquence, un effet sur l'homme (on parlera alors d'un effet différé). Les informations sur les enjeux sont souvent directement issues de bases de données. L'accident ayant eu lieu, l'information sur la probabilité d'occurrence de cet évènement ne nous intéressera en revanche pas. Dans ces conditions, c'est essentiellement l'évaluation de l'intensité du phénomène dangereux qui retiendra notre attention dans la suite de ce document, et dans ce contexte, le type de phénomène dangereux étudié sera plus spécifiquement la dispersion atmosphérique de gaz toxiques résultant de l'accident TMD. Ce phénomène peut être modélisé et évalué à chaque instant en utilisant des modèles d'effet, ces modèles contiennent des paramètres liés à la base de données d'exposition, de données météorologiques, des caractéristiques du produit émis et de la géolocalisation de la source d'émission.

La deuxième problématique réside dans les incertitudes affectant un certain nombre d'entrées dans ces modèles d'effets [19]. Ces incertitudes peuvent concerner les conditions météorologiques comme la vitesse du vent, grandeur qui peut beaucoup varier d'un instant à l'autre et qui est rarement mesurée directement sur le lieu de l'accident, ou bien les paramètres du modèle d'effets comme la quantité du gaz relâché difficile à estimer en pratique. Ces incertitudes conduisent naturellement à de l'incertitude sur le niveau de risque évalué. Donc pour correctement évaluer les effets et déterminer la zone de danger et de sécurité, il est nécessaire d'identifier et de prendre en compte les incertitudes sur les entrées des modèles d'effets afin de fournir un niveau de risque fiable et précis.

1.4 État de l'art

1.4.1 Modèles d'effets

Pour évaluer l'intensité des risques, un modèle d'effets sous la forme d'un ensemble de relations mathématiques plus ou moins complexes, est donc utilisé pour représenter les relations entre les variables physiques et les paramètres caractéristiques de la dispersion atmosphérique (vitesse du vent, point d'émission, paramètres de dispersion,...) et l'impact (concentration du produit dans l'air en un point donné) à évaluer et à comparer à des

seuils réglementaires (effet létaux ou irréversibles sur la santé,...). L'approche utilisée pour déterminer les zones de danger et de sécurité est basée sur une valeur du seuil de concentration. Si en un point donné, la concentration en gaz toxique est inférieure à cette concentration seuil, on est dans la zone de sécurité. Au contraire, si en un point donné, la concentration en gaz toxique est supérieure à cette concentration seuil, on est dans la zone de danger.

Trois grandes catégories d'entrées influencent la dispersion atmosphérique d'un produit :

- les caractéristiques du rejet (nature du nuage de produit, mode d'émission,...),
- les conditions météorologiques (champ de vent, de température,...),
- les spécificités de l'environnement (nature du sol, obstacles, topographie,...).

Il existe trois familles de modèles de dispersion atmosphérique, qui sont classées par ordre de complexité croissante :

- les modèles de type gaussien qui permettent d'estimer la dispersion des gaz passifs. Pour rappel, le gaz est dit passif lorsqu'il n'apporte aucune perturbation mécanique à l'écoulement atmosphérique et se disperse du fait de la seule action du fluide porteur, à savoir l'air,
- les modèles de type intégral à utiliser dès que le rejet perturbe l'écoulement atmosphérique de l'air, et qui sont utilisés pour les gaz lourds,
- les modèles tridimensionnels qui permettent de prendre en compte la complexité de l'environnement (obstacles, relief,...) en s'appuyant sur la résolution des équations de la mécanique des fluides.

Par ailleurs, les modèles de dispersion peuvent se ranger en deux grands groupes :

- les modèles analytiques,
- les modèles par code.

Globalement, les modèles analytiques s'attachent à modéliser le phénomène de dispersion à partir d'équations paramétrées et simplifiées pour lesquelles des expressions formelles sont accessibles et peuvent donc être manipulées, comme par exemple le modèle gaussien de dispersion. Au contraire, les modèles par code sont plus sophistiqués, exprimés sous forme de code de calcul, pour lesquels on n'a pas de forme analytique du modèle mais un logiciel qui estime la dispersion atmosphérique. Ce sont concrètement des modèles de type boites noires.

1.4.1.1 Modèles de type gaussien

Un des modèles d'effets utilisé dans cette étude est le modèle gaussien de dispersion atmosphérique. Les modèles gaussiens peuvent être utilisés pour modéliser et simuler la dispersion à une échelle locale et sur un périmètre allant de quelques dizaines de mètres à quelques dizaines de kilomètres autour de la source de fuite. À ces échelles et sous certaines hypothèses, une solution analytique du champ de concentration est obtenue sous

la forme d'un panache gaussien, le transport et la diffusion du gaz vont alors dépendre du vent et de la turbulence atmosphérique d'origine mécanique ou thermique (figure 1.5).

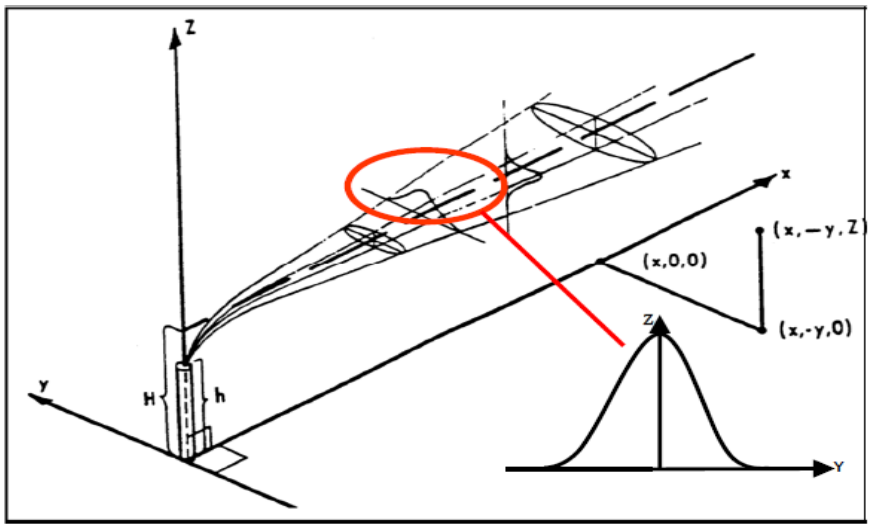

FIGURE 1.5: Répartition gaussienne de la concentration dans un panache de gaz passif [1]

Le modèle gaussien s'applique aux rejets de gaz passifs, le produit rejeté doit donc avoir :
- une densité à peu près égale à celle de l'air (ou bien il est très dilué) ;
- une température identique à celle de l'air ;
- une vitesse initiale relative nulle.

La vitesse du vent doit être d'au moins 1 à 2 m/s et le terrain doit être homogène et plat. En effet, la présence de reliefs, d'obstacles (murs, bâtiments...) introduirait des perturbations importantes de l'écoulement de l'air qui ne sont pas pris en compte par ces modèles. La forme générale du modèle est donnée par :

$$c_k = \frac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} \sigma_y \sigma_z} \times exp\left[-\frac{1}{2}\left(\frac{y_k}{\sigma_y}\right)^2\right] \times \left\{exp\left(-\frac{1}{2}\left(\frac{z_k - h}{\sigma_z}\right)^2\right) + \alpha exp\left(-\frac{1}{2}\left(\frac{z_k + h}{\sigma_z}\right)^2\right)\right\} \tag{1.1}$$

où
- c_k est la concentration en gaz émis (en micro-grammes par mètre cube) en tout point situé à x_k mètres sous le vent de la source, y_k mètres latéralement de l'axe central du panache, et z_k mètres au-dessus du niveau du sol,
- l'indice k désigne des évaluations différentes de la sortie du modèle,
- le terme q (en grammes par seconde) est le débit d'émission,
- u_{ref} est la vitesse du vent (en mètres par seconde) mesurée à une altitude donnée z_{ref},
- h est l'altitude du point d'émission (en mètres),
- α désigne le coefficient de réflexion du sol qui modélise la capacité de réflexion ou d'absorption du produit sur le sol, l'eau ou les végétaux. Ce coefficient est compris entre 0 et 1 ($\alpha = 0$: absorption totale et $\alpha = 1$: réflexion totale).

La concentration est calculée sur l'axe du panache, une loi gaussienne permet ensuite d'en déduire la concentration dans tout le panache. Dans ce modèle existent un certain nombre de paramètres ou de variables d'entrée connus avec imprécision comme le débit de fuite q difficile à connaitre en pratique, la vitesse du vent u_{ref} fluctuante et pas forcément mesurée à proximité de l'accident, ou encore les différents écarts-types de la distribution gaussienne σ_y et σ_z qui sont estimés pour une classe atmosphérique.
Les écarts types de la loi gaussienne dépendent :
– de la distance à la source ou de la durée de transfert,
– des caractéristiques de la structure de l'atmosphère,
– et de la rugosité du site.
L'utilisation des modèles gaussiens impose donc la détermination de ces écarts-types [1, 20–22], qui s'expriment généralement en fonction de la distance x_k à la source (valable pour des distances supérieures à 100 m et inférieures à 10 km).
Les expressions des écarts types σ_y et σ_z proposées par Pasquill correspondent à des durées d'échantillonnage de 10 minutes, et une hauteur de source qui n'excède pas quelques centaines de mètres. Elles se présentent sous la forme suivante :

$$\sigma = a.x_k^b + c$$

Les valeurs des coefficient a, b et c sont reportées dans les tableaux suivants pour respectivement σ_y et σ_z (x_k, σ_y et σ_z sont exprimés en kilomètres).

Stabilité atmosphérique (Pasquill)	a	b	c
A	0,215	0,858	0
B	0,155	0,889	0
C	0,105	0,903	0
D	0,068	0,908	0
E	0,05	0,914	0
F	0,034	0,908	0

TABLE 1.2: Coefficients relatifs à σ_y

Stabilité atmosphérique (Pasquill)	a	b	c
A	0,467	1,89	0,01
B	0,103	1,11	0
C	0,066	0,915	0
D	0,0315	0,822	0
E si $x_k < 1$ km	0,0232	0,745	0
E si $x_k > 1$ km	0,148	0,15	-0,126
F si $x_k < 1$ km	0,0144	0,727	0
F si $x_k > 1$ km	0,0312	0,306	-0,017

TABLE 1.3: Coefficients relatifs à σ_z

La méthode la plus couramment utilisée pour évaluer le niveau de la turbulence atmosphérique actuelle est la méthode développée par Pasquill en 1961. Il a classé la turbulence atmosphérique en six classes de stabilité nommées A, B, C, D, E et F ; la classe A étant la plus instable ou turbulente, et la classe F la plus stable ou moins turbulente. Le tableau 1.4 donne la liste des six classes avec les conditions météorologiques qui définissent chaque classe.

Vitesse du vent à 10m	Jour			Nuit	
	Rayonnement solaire incident			Nébolusité	
[m/s]	Fort	Modéré	Faible	4/8-7/8	<3/8
< 2	A	A-B	B	F	F
2-3	A-B	B	C	E	F
3-5	B	B-C	C	D	E
5-6	C	C-D	D	D	D
> 6	C	D	D	D	D

TABLE 1.4: Classes de stabilité de Pasquill

1.4.1.2 Modèle de type intégral

L'utilisation d'un modèle intégral permet de modéliser les mécanismes physiques qui n'étaient pas pris en compte par les modèles gaussiens comme les effets de gravité pour les rejets de gaz lourds, les effets de flottabilité pour les rejets de gaz légers et l'effet de "jet" pour les rejets à vitesse d'émission élevée. Ce type de modèle est basé sur des équations de la mécanique des fluides simplifiées pour permettre une résolution rapide. Cette simplification se traduit par l'introduction de paramètres représentant globalement les mécanismes non modélisés. A cet effet, les coefficients des modèles intégraux sont déterminés à partir d'expérimentations.

Pour la modélisation des nuages de gaz passifs, l'outil intégral utilise un modèle de type gaussien. Les modèles intégraux comprennent, dans la plupart des cas, un module de calcul permettant de déterminer de façon plus ou moins forfaitaire le terme source de rejet en fonction des conditions de stockage du produit et du type de rejet (rupture guillotine, ruine du réservoir, évaporation de flaque...). Ce type de modèle s'applique aux gaz neutres, aux gaz denses et parfois aux gaz légers (pour les versions les plus récentes des logiciels). La turbulence atmosphérique est prise en compte par l'intermédiaire de classes de stabilité atmosphérique, pour éviter une modélisation lourde de la turbulence. Comme pour les modèles gaussiens, au-delà de la dizaine de kilomètres, les résultats ne sont plus valables car d'autres phénomènes de turbulence et de diffusion doivent être considérés.

Il existe plusieurs logiciels utilisant un modèle de type intégral tels que SLAB, PHAST, HGSYSTEM, SCIPUFF, TRACE, ALOHA,... Le logiciel SLAB semble pouvoir répondre aux exigences du présent travail de recherche. Il s'agit en effet d'un logiciel libre de droit pour lequel les codes sources sont téléchargeables entre autre à partir du site internet de l'US Environmental Protection Agency.

Modèle de type intégral : SLAB

Le modèle SLAB [2] est un modèle de dispersion atmosphérique par code. Ce modèle est basé sur le concept d'entraînement de l'air dans un nuage de gaz lourd et sur l'effet d'affaissement de celui-ci du fait de la gravité (théorie de Zeman [23]).

SLAB est un logiciel exécutable qui a été développé dans le langage de programmation Fortran. SLAB prend en entrée un fichier texte et génère en sortie un fichier texte (figure 1.6) ; il ne possède pas d'interface facilitant l'automatisation des calculs. Ermak et Chan [2, 24] ont réalisé le codage informatique de SLAB et les développements qui ont suivi.

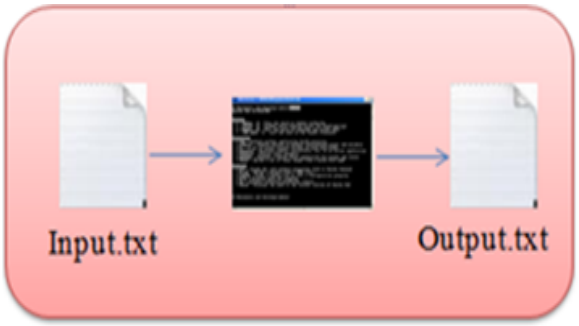

FIGURE 1.6: Fonctionnement du logiciel SLAB

Le modèle complexe par code SLAB estime la dispersion atmosphérique d'un gaz lourd par la résolution des équations simplifiées de conservation de la masse, de la quantité de mouvement, de l'énergie et des espèces. Ces équations sont résolues dans l'espace de manière à ce que le nuage puisse être traité comme un panache stationnaire, une bouffée transitoire ou une combinaison des deux selon la durée du rejet [3].

La représentation géométrique utilisée par le modèle complexe SLAB est présentée sur la figure 1.7 [25] :

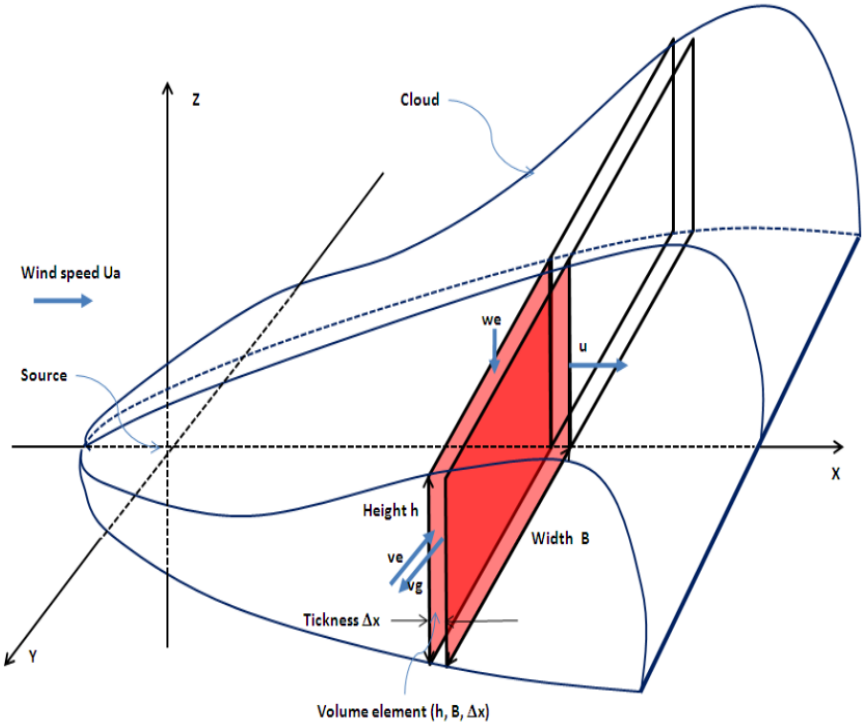

FIGURE 1.7: Calcul de l'évolution du nuage par la méthode intégrale ([2, 3])

Un champ tridimensionnel de concentrations moyennées sur une période précisée par l'utilisateur est calculé par le modèle SLAB. Il permet aussi d'accéder aux dimensions du nuage.

Le logiciel SLAB, qui évalue des concentrations volumiques en gaz moyennées sur le temps, renvoie en sortie trois types d'informations :
- les paramètres de contour de concentration (paramètres internes pour la simulation)
- la concentration en gaz dans le plan horizontal $Z = ZP(I)$
- la concentration maximale sur le ligne centrale

Le deuxième type d'informations, qui est celui qui nous intéressera par la suite, correspond à un tableau de concentrations volumiques moyennées dans le temps et calculées dans le plan horizontal à une hauteur $Z = ZP(I)$ au dessus du sol. Jusqu'à quatre plans peuvent être sélectionnés par l'utilisateur et être spécifiés en entrée. Dans ce tableau, les concentrations calculées en différents points spatialement répartis, sont listées en fonction de la distance sous le vent x. A chaque position x sous le vent, la demi-largeur effective du nuage notée BBC est donnée. Ce type d' informations est suivi par les concentrations volumiques évaluées en six endroits sur l'axe y (horizontal et perpendiculaire à l'axe du vent) définis par :

$$y = N.BBC$$

avec

$$N \in \{0; 0.5; 1.0; 1.5; 2; 2.5\}.$$

Les résultats pour le plan $ZP(1)$ sont donnés en premier, suivis par ceux du plan $ZP(2)$ et ainsi de suite jusqu'à ce que les résultats pour le dernier plan soient renvoyés (figure 1.8).

time averaged (tav = 10. s) volume concentration: concentration in the z = 0.00 plane.

downwind distance x (m)	time of max conc (s)	cloud duration (s)	effective half width bbc (m)	average concentration (volume fraction) at (x,y,z) y/bbc= 0.0	y/bbc= 0.5	y/bbc= 1.0	y/bbc= 1.5	y/bbc= 2.0	y/bbc= 2.5
1.00E+00	1.90E+02	3.81E+02	4.82E-01	4.84E-03	4.57E-03	1.67E-03	4.15E-05	3.00E-08	0.00E+00
1.02E+00	1.91E+02	3.81E+02	4.83E-01	4.96E-03	4.68E-03	1.71E-03	4.28E-05	3.16E-08	0.00E+00
1.04E+00	1.91E+02	3.81E+02	4.84E-01	5.08E-03	4.79E-03	1.75E-03	4.44E-05	3.39E-08	0.00E+00
1.07E+00	1.91E+02	3.81E+02	4.85E-01	5.24E-03	4.93E-03	1.80E-03	4.64E-05	3.69E-08	0.00E+00
1.10E+00	1.91E+02	3.81E+02	4.86E-01	5.43E-03	5.11E-03	1.87E-03	4.88E-05	4.06E-08	0.00E+00
1.13E+00	1.91E+02	3.81E+02	4.88E-01	5.66E-03	5.32E-03	1.94E-03	5.19E-05	4.54E-08	0.00E+00
1.18E+00	1.91E+02	3.81E+02	4.90E-01	5.95E-03	5.59E-03	2.04E-03	5.57E-05	5.18E-08	0.00E+00

time averaged (tav = 10. s) volume concentration: concentration in the z = 1.00 plane.

downwind distance x (m)	time of max conc (s)	cloud duration (s)	effective half width bbc (m)	average concentration (volume fraction) at (x,y,z) y/bbc= 0.0	y/bbc= 0.5	y/bbc= 1.0	y/bbc= 1.5	y/bbc= 2.0	y/bbc= 2.5
1.00E+00	1.90E+02	3.81E+02	4.82E-01	1.00E+00	1.00E+00	5.31E-01	1.31E-02	9.52E-06	0.00E+00
1.02E+00	1.91E+02	3.81E+02	4.83E-01	1.00E+00	1.00E+00	5.30E-01	1.33E-02	9.79E-06	0.00E+00
1.04E+00	1.91E+02	3.81E+02	4.84E-01	1.00E+00	1.00E+00	5.30E-01	1.34E-02	1.03E-05	0.00E+00
1.07E+00	1.91E+02	3.81E+02	4.85E-01	1.00E+00	1.00E+00	5.29E-01	1.36E-02	1.08E-05	0.00E+00
1.10E+00	1.91E+02	3.81E+02	4.86E-01	1.00E+00	1.00E+00	5.28E-01	1.38E-02	1.15E-05	0.00E+00
1.13E+00	1.91E+02	3.81E+02	4.88E-01	1.00E+00	1.00E+00	5.27E-01	1.41E-02	1.23E-05	0.00E+00
1.18E+00	1.91E+02	3.81E+02	4.90E-01	1.00E+00	1.00E+00	5.26E-01	1.44E-02	1.34E-05	0.00E+00

FIGURE 1.8: Tableau de concentrations volumiques moyennées

Une fois le champ de concentrations en gaz estimé dans l'espace et le temps, il reste à en évaluer ses effets. Dans le cas d'un contact avec un nuage toxique par inhalation, les effets sont estimés sur la base de la dose inhalée qui est comparée à des doses estimées au moyen de valeurs seuil de toxicité issues des fiches de toxicité aigüe de référence. Les effets sélectionnés sont les effets irréversibles et les premiers effets létaux. Les effets générés par l'inhalation d'un produit toxique donné sont alors fonction de deux variables : la concentration du produit et la durée d'exposition.

1.4.1.3 Modèles tridimensionnels

Les modèles CFD (Computational Fluid Dynamics) sont des modèles de simulation de la Mécanique des Fluides permettant de simuler des écoulements en configurations complexes. Ce type de modèles s'attache à résoudre le système d'équations de la mécanique des fluides qui gouvernent la dispersion (conservation de la quantité de mouvement, conservation de l'énergie, conservation de la masse).

Les simplifications des équations de la mécanique des fluides sont beaucoup moins poussées que celles effectuées dans les modèles intégraux. Les modèles numériques CFD permettent de simuler la dispersion atmosphérique de gaz émis de manière chronique ou accidentelle en prenant en compte l'ensemble des phénomènes intervenant de façon significative sur la dispersion, qu'ils soient liés à l'atmosphère comme la turbulence thermique, ou au site comme les obstacles ou le relief.

Les logiciels Fluidyn Panair, FLACS, Fluent, Aria risk sont des modèles de type CFD/3D Les outils CFD tridimensionnels, une fois validés, ont un champ d'application étendu et prennent en compte des débits variables et des sources multiples.

1.4.1.4 Comparaison des différents modèles de dispersion

Le tableau 1.5 montre les principaux avantages et inconvénients de ces trois types de modèles.

	Modèle gaussien	**Modèle intégral**	**Modèle CFD**
Principaux avantages	- Facilité de mise en oeuvre - Coût de calcul très faible - Grande variété de paramétrisations disponible, validation sur des cas académiques bien connus	- Facilité de mise en oeuvre - Prise en compte des effets de gaz lourds et gaz légers - Quantification du terme source dans certains logiciels - Modèles calés sur des expériences	- Champ d'application étendu - Débits variables - Prise en compte de la réalité du terrain - Conditions météorologiques extrêmes
Principaux inconvénients	- Simplifications de la modélisation physique - Conditions météorologiques moyennes - Pas d'effet de gravité (gaz de même densité que l'air) -Champ lointain (distance de l'ordre de 100 m à une dizaine de km de la source) et terrain plat - Nuages ne s'éloignant pas trop du sol	- Pas d'obstacle ni de relief - Pas de conditions météorologiques extrêmes - Erreurs dues à la simplification des équations de la mécanique des fluides - Champ lointain (distance de l'ordre de 20 m à une dizaine de km de la source) - Valable à condition que la concentration soit homogène à l'intérieur du nuage	- Temps de calcul important - Difficultés de mise en œuvre - Dépendant de la qualité des données d'entrée liées aux conditions et aux limites du problème à traiter (topographie, profils verticaux de vents et de température,...). - Complexité des algorithmes, moyens informatiques importants

TABLE 1.5: Tableau de comparaison des modèles de dispersion [4–6]

1.4.2 Sources des incertitudes de modèle

D'un point de vue utilisation pratique, la difficulté principale réside dans l'incertitude des données utilisées dans ces modèles. Certaines informations comme la quantité de produit se dispersant sont difficiles à obtenir, d'autres comme la vitesse du vent ou les paramètres de dispersion ne sont connus qu'avec imprécision. Pour avoir une évaluation fiable des effets et évaluer correctement les zones de danger et de sécurité, il faut tenir compte des incertitudes sur les paramètres et sur les variables d'entrée des modèles.
Les incertitudes paramétriques sont celles portant sur les coefficients du modèle (caractéristiques physiques, paramètres estimés pour une classe atmosphérique,...). Ces incertitudes peuvent induire une variabilité significative des résultats recherchés.
Les incertitudes des variables d'entrée sont celles portant sur les grandeurs physiques d'un modèle et provenant des incertitudes sur les données mesurées ou observées utilisées pour estimer ces dernières. Dans les modèles d'effets, les incertitudes sur les variables d'entrée de modèle peuvent être très nombreuses et toucher différentes parties du modèle. Les principales sources d'incertitudes pour les grandeurs physiques peuvent résulter des différentes origines suivantes comme l'insuffisance des données observées, les erreurs (biais) de mesure, la précision technologique des capteurs,...

1.4.3 Modélisation et propagation des incertitudes

Il existe plusieurs approches pour modéliser les incertitudes comme :
- l'approche probabiliste utilisée par Morgan et Henrion (1990)[26] et Gentle (2003)[27] visant à représenter une grandeur par un ensemble de valeurs aléatoires suivant une loi de probabilité donnée (gaussienne, uniforme, triangulaire,...),
- l'approche ensembliste introduite par Moore R.E., (1979)[28] visant à représenter une grandeur non pas par une valeur moyenne ou nominale, mais par l'ensemble des valeurs qu'elle peut prendre, à savoir un support intervalle,
- l'approche possibiliste (ensembles flous) introduite par Zadeh [29] et développée par Dubois et Prade [30] consistant à représenter les paramètres incertains par des distributions de possibilités,
- l'approche hybride proposée par Guyonnet et al. (2002, 2003) [31, 32], Ferson et Ginzburg, (1995) [33] combinant des informations de type probabiliste et possibiliste.

Une fois modélisées les incertitudes sur les entrées, la propagation des incertitudes consiste à déterminer l'incertitude sur la sortie qui est induite par les grandeurs d'entrée incertaines lors de son évaluation par le modèle de dispersion étudié. Les deux approches probabiliste et ensembliste ont retenues toute notre attention. La première, plus classique permettra d'obtenir des points de comparaison. La seconde, plus originale, est

comme nous le verrons au cours de ce document, naturellement adaptée à la modélisation des incertitudes envisagées au niveau du modèle d'effets et propose dans le même temps des outils permettant de propager ces incertitudes lors de la détermination des zones de danger ou sécurité. Nous présenterons donc dans les sections suivantes les principes généraux des deux approches utilisées. Cependant, les méthodes de propagation d'incertitudes doivent nécessairement être adaptées à l'approche utilisée pour modéliser les entrées incertaines. Ainsi la simulation de Monte Carlo pour l'approche probabiliste et l'analyse par intervalles pour l'approche ensembliste seront plus amplement détaillées un peu plus tard dans le chapitre 2 dédié aux méthodes utilisées pour propager des incertitudes.

1.4.3.1 Approche Probabiliste

L'approche probabiliste est une méthode classique et générale pour prendre en compte les incertitudes de modèle qui vise à représenter une grandeur incertaine par une variable aléatoire. L'incertitude est propagée par la technique de Monte-Carlo. Le principe de cette technique consiste à se servir du modèle comme d'une boîte noire. Le principe de base est simple [34, 35]. La première étape consiste à déterminer au hasard une valeur pour chaque entrée incertaine selon sa densité de probabilité, la deuxième étape consiste à calculer la sortie du modèle de dispersion pour chaque tirage des entrées et finalement répéter la procédure afin d'obtenir un échantillon suffisant des valeurs de sortie et ainsi estimer l'incertitude sur celle-ci. La méthode de Monte-Carlo sera plus amplement détaillée dans la section 2.1.1.
La figure 1.9 présente le principe général de la propagation des incertitudes avec l'approche probabiliste.

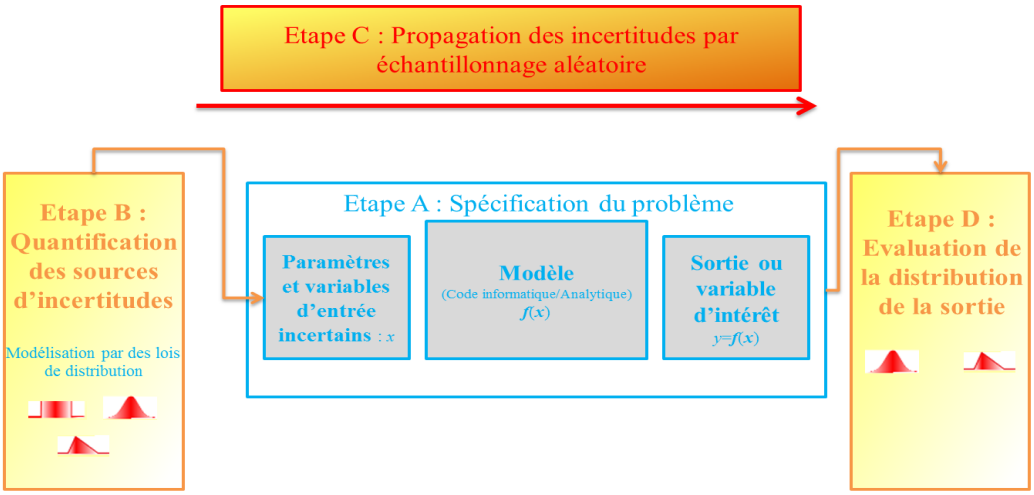

FIGURE 1.9: Principe général de la propagation d'incertitudes par Monte-Carlo

Compte tenu du nombre fini de tirages aléatoires des entrées du modèle, il n'est pas possible de prendre en compte toutes les situations possibles. De plus, les temps de calcul, dépendants de la taille de l'échantillon et de la complexité du modèle, sont généralement importants, mais ce problème peut être en partie résolu à l'aide d'une analyse de sensibilité.

Fonction de distribution

Un événement aléatoire E est un événement qui a une chance de se produire et une probabilité $p(E)$ est une mesure numérique de cette chance. Une probabilité est un nombre compris entre 0 et 1, bornes incluses.
On a :

$$0 \leq p(E) \leq 1$$

Si tous les évènements E sont équiprobables parmi un ensemble de r évènements possibles, alors : $p(E) = \frac{1}{r}$. On parle alors de loi uniforme sur l'ensemble de taille r.

Les lois de distribution sont un bon moyen pour exprimer l'incertitude sur la plupart des entrées d'un modèle, mais deux lois sont plus couramment employées : la distribution uniforme et la distribution normale. Selon MacDonald [36] les distributions normales sont adaptées aux entrées dont l'incertitude provient du bruit de mesure, alors que l'autre distribution convient pour décrire les incertitudes provenant d'un manque d'information sur des entrées physiques qui doivent logiquement rester bornées.

Les principales lois de distribution [26, 37] sont illustrées sur la figure 1.10.

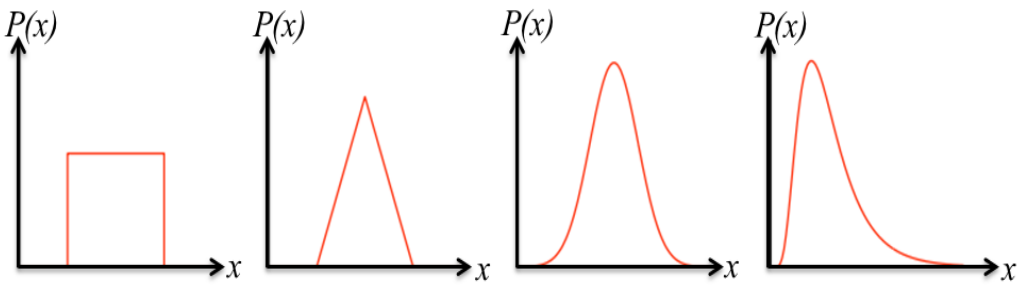

FIGURE 1.10: Distributions statistiques

Distribution uniforme : La distribution uniforme (figure 1.10a) est la distribution statistique la plus simple, cette loi modélise un phénomène uniforme sur un intervalle donné : toutes les valeurs prises par la variable aléatoire sont équiprobables.

Distribution triangulaire : Cette distribution (figure 1.10b) peut être définie par trois paramètres : le minimum, le maximum et le mode. Elle permet d'exprimer une probabilité plus grande pour les valeurs centrales sans toutefois connaître la vraie distribution, ainsi les valeurs proches de la valeur mode sont plus susceptibles de se réaliser.

Distribution normale : C'est la loi de distribution la plus connue, parfois sous le vocable de loi de Laplace-Gauss, cette distribution (figure 1.10c) est symétrique. Elle est caractérisée par le fait que les valeurs les plus tirées se situent au centre de la distribution, ce nombre s'amenuisant progressivement de part et d'autre de la valeur moyenne. La moyenne et la médiane sont identiques. L'inconvénient est que le support d'une loi normale n'est pas borné, ce qui peut être problématique pour des entrées incertaines ayant un sens physique.

Distribution log-normale : Les valeurs sont positivement asymétriques, et non symétriques comme dans une distribution normale, c'est-à-dire que la médiane, la moyenne et le mode ne se trouvent pas au même endroit. Elle est très utile lorsque les valeurs négatives ne sont pas possibles. Cette distribution (figure 1.10d) sert à représenter les valeurs qui ne tombent jamais sous zéro mais qui ont un potentiel positif illimité.

Exemple 1. Nous considérons le modèle linéaire à entrées indépendantes suivantes :

$$f(x_1, x_2) = 3x_1 + 2x_2$$

Tel que x_1 et x_2 suivent respectivement une loi de distribution normale sur $N(20, 16)$ et $N(60, 64)$.

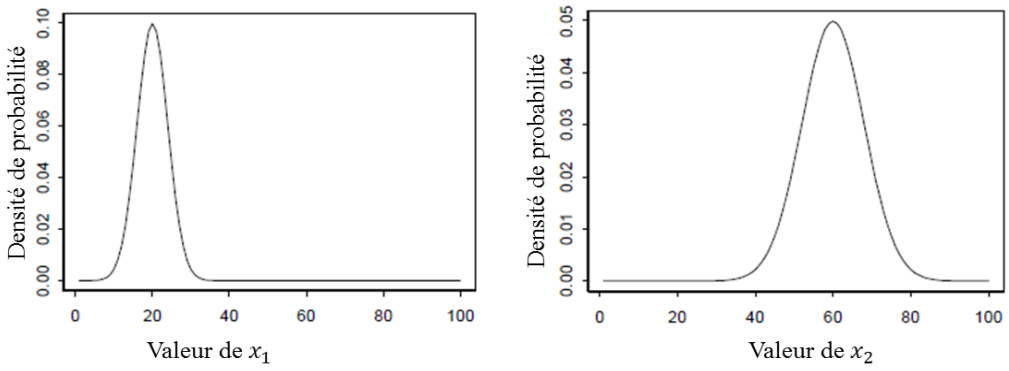

FIGURE 1.11: Incertitude sur x_1 et x_2

L'objectif est de déterminer la distribution de probabilité de $f(x_1, x_2)$ à partir des distributions de x_1 et x_2. En notant par $E(.)$ et $V(.)$ l'espérance et la variance de la variable aléatoire passée en argument, comme x_1 et x_2 sont deux variables indépendantes de distribution normale alors :

1. $3x_1 + 2x_2$ suit une distribution normale.
2. $E(3x_1 + 2x_2) = 3E(x_1) + 2E(x_2)$.
3. $V(3x_1 + 2x_2) = 3^2V(x_1) + 2^2V(x_2)$

La densité de probabilité de $f(x_1, x_2)$ sera : $N(180, 400)$.

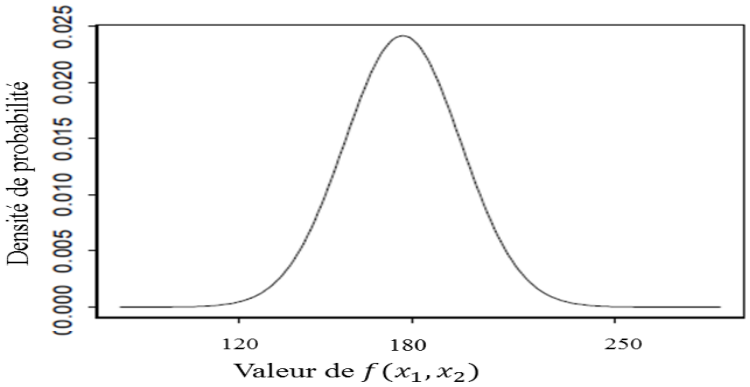

FIGURE 1.12: Incertitude sur $f(x_1, x_2)$

Pour ce modèle simple, on peut déterminer expérimentalement la densité de probabilité de $f(x_1, x_2)$ en utilisant la technique de Monte-Carlo en quatre étapes :

1. Définir les distributions de x_1 et x_2,
2. Générer aléatoirement $N = 1000$ échantillons à partir des lois de distribution définies à l'étape 1,
3. Calculer $f(x_1, x_2)$ pour chaque échantillon de x_1 et x_2 généré,
4. Estimer la distribution de $f(x_1, x_2)$.

Différentes approches sont possibles pour estimer la distribution, nous avons opté ici pour une évaluation graphique à base d'histogramme comme illustré sur la figure 1.13

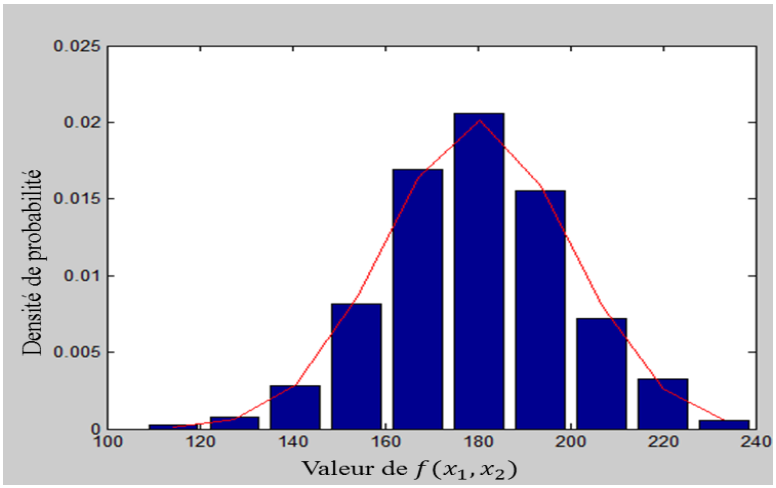

FIGURE 1.13: Densité de probabilité estimée de $f(x_1, x_2)$

Exemple 2. Nous considérons le modèle linéaire suivant :

$$y = f(x_1, x_2) = x_1 + x_2$$

Tel que x_1 et x_2 sont 2 variables aléatoires indépendantes équidistribuées respectivement sur les supports bornés $[x_1^-, x_1^+]$ et $[x_2^-, x_2^+]$.

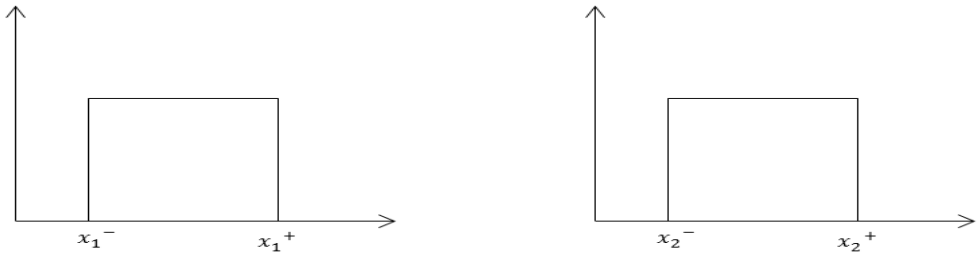

FIGURE 1.14: Incertitude sur x_1 et x_2

L'objectif est de déterminer la distribution de probabilité de $f(x_1, x_2)$ à partir des distributions de x_1 et x_2. La somme de deux variables stochastiques à distribution uniforme conduit à une distribution trapézoïdale :

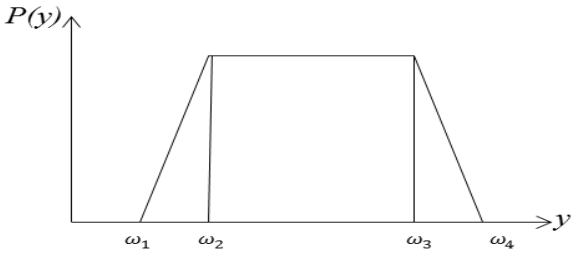

FIGURE 1.15: Incertitude sur $f(x_1, x_2)$

avec : $\omega_1 = x_1^- + x_2^-$, $\omega_2 = min(x_1^- + x_2^+, x_1^+ + x_2^-)$, $\omega_3 = max(x_1^- + x_2^+, x_1^+ + x_2^-)$ et $\omega_4 = x_1^+ + x_2^+$. On observe ainsi que même pour des modèles simples, les lois de distribution ne sont pas stables et la loi de distribution de la sortie peut devenir très rapidement complexe (le produit de deux distributions normales n'est par exemple plus une distribution normale).

Les points forts de la méthode probabiliste résident premièrement dans une modélisation riche des grandeurs incertaines en utilisant à la fois des informations sur la distribution et le support, deuxièmement dans la capacité à propager les incertitudes de manière simple en simulant un grand nombre de fois le modèle, et ce avec n'importe quel type de modèle, qu'il soit sous forme analytique ou code informatique.
Les points faibles de cette méthode, notamment concernant notre travail, proviennent de la difficulté à savoir si les entrées incertaines étudiées (vitesse du vent,..) suivent vraiment les lois de distribution choisies. La deuxième difficulté réside dans la complexité à déterminer d'un point de vue pratique la nature de la distribution de la sortie de modèle

parce qu' il n'est plus possible de déterminer la distribution de la sortie, autrement que par simulation numérique et que la loi de distribution obtenue est de toute façon généralement quelconque, si bien qu'au final seule l'information relative au support de la sortie est utilisée.

1.4.3.2 Approche Ensembliste

De manière générale, les méthodes ensemblistes désignent toutes les méthodes permettant de manipuler des sous-ensembles de R^n et de résoudre des contraintes liant les éléments de ces ensembles afin de caractériser l'ensemble des solutions d'un problème à traiter. On parle alors de méthode garantie dans le sens où toutes les solutions sont appréhendées sans en perdre lors du processus de traitement. Les méthodes ensemblistes permettent ainsi d'élaborer des preuves formelles de propriétés, à partir d'une évaluation algorithmique d'un résultat. Elles permettent d'établir un lien entre les méthodes numériques, limitées quant à leur garantie de résultat, et les méthodes analytiques, limitées quant à la complexité des systèmes qu'elles permettent de traiter.

Les sous-ensembles élémentaires (aussi appelés récipients) utilisés peuvent prendre des formes diverses et variées comme des zonotopes, des polytopes, des ellipsoïdes, des boîtes (ou vecteur d'intervalles)... Ces ensembles sont dans tous les cas décrits à l'aide de variables intervalles et leur manipulation est réalisée à l'aide de l'analyse par intervalles définissant les règles de calculs sur les intervalles (Moore, 1979 [28]), (Neumaier, 1990 [38]). Lorsque le domaine de solution est trop complexe et ne peut être exactement représenté par un simple sous-ensemble élémentaire, des approximations intérieures (dont tous les points sont solution) ou extérieure (aucune solution n'existe en dehors de cette approximation) sont déterminées en travaillant sur des unions de sous-ensembles (généralement des unions de boîtes).

Historiquement, les outils ensemblistes sont tout d'abord développés par les informaticiens dans le cadre de l'analyse des intervalles, afin de tenir compte des imprécisions sur la valeur des nombres. Ces imprécisions peuvent provenir des données issues d'une chaîne d'instrumentation, problème alors lié à la qualité du capteur (auquel cas le fabriquant indique une précision technologique), aux conditions opératoires, à la présence de signaux parasites ou tout simplement d'un défaut (biais systématique, dérive,..., à cause d'un mauvais étalonnage ou d'un encrassement). Ces imprécisions peuvent aussi provenir de l'outil de calcul informatique lui-même. En effet, même si un nombre est parfaitement connu, son codage informatique sous forme d'une série plus ou moins longue de bits est nécessairement fini. Ainsi, un nombre rationnel tel que 1/3 ne peut être exactement représenté. La valeur réellement utilisée dans les calculs est un nombre tronqué 0.3333...

où la quantité de chiffres après la virgule dépend du nombre de bits assurant le codage. De plus, de nombreuses fonctions usuelles sont en réalité calculées à l'aide d'approximations ce qui nécessite une gestion des erreurs d'arrondis. Parallèlement à la montée en puissance des calculateurs dans les années 60, se pose donc la question de savoir comment ces imprécisions évoluent au cours des calculs et comment quantifier l'erreur sur le résultat final. C'est donc en s'attachant à résoudre ces difficultés que Moore R.E. ouvre la voie de l'analyse par intervalles, en publiant notamment l'ouvrage de référence (Moore, 1966 [39]) mettant en place les fondements de cet outil, suivi de (Moore, 1979 [28]). Plus tard, cet outil permet l'étude de problèmes mathématiques complexes comme la résolution de systèmes d'équations linéaires ou non-linéaires (Neumaier, 1990 [38]), la résolution de problème inverse ou encore l'optimisation globale d'un critère non convexe permettant la recherche de manière garantie de tous ses minimiseurs (Hansen, 1992 [40]), (Didrit, 1997 [41]) alors que les méthodes plus classiques d'optimisation non-linéaire (Gradient, Newton-Raphson,...) ne renvoient qu'un seul optimum pouvant n'être que local.

La communauté automaticienne n'est pas en reste et participe de manière active au développement de ces méthodes. En effet, le principal attrait de l'outil ensembliste réside dans sa capacité à appréhender les incertitudes sur les paramètres ou les variables d'un modèle. C'est donc tout naturellement que l'outil ensembliste fait son apparition dans le domaine de l'estimation paramétrique dans les années 80. Appelée estimation paramétrique ensembliste, l'objectif consiste alors à déterminer l'ensemble des valeurs acceptables des paramètres, c'est-à-dire cohérentes avec les mesures, le modèle et les bornes de l'erreur d'équation. L'essentiel des travaux portant sur l'approche ensembliste en estimation paramétrique sont regroupés dans l'ouvrage collectif (Milanese et al, 1996 [42]) et de nombreux articles de synthèse ont été publiés (Walter et al, 1990 [43]). Plus récemment, des méthodes dites de caractérisation ont été développées pour identifier les supports intervalles des différentes grandeurs incertaines dans le cas de modèles incertains aussi bien linéaires que non-linéaires (Adrot et al, 2006 [44]), (Blesa, 2011 [45]).

En définitive, les méthodes ensemblistes permettent naturellement de représenter l'imprécision d'un modèle. Au lieu de représenter des paramètres ou des mesures connus avec imprécision par des valeurs nominales, ces grandeurs sont décrites par leurs supports bornés (le plus souvent des intervalles), c'est-à-dire les ensembles de valeurs qu'ils sont susceptibles de prendre.

$$[x] = [x^-, x^+]$$

avec x^- et x^+ désignant les bornes inférieure et supérieure de la variable x.
Contrairement à l'approche probabiliste, aucune hypothèse relative à une loi de distribution n'est requise. Une fois l'étape d'estimation paramétrique réalisée, l'idée consiste à étendre les relations mathématiques des modèles de la physique à ces ensembles, puis

à propager ces supports au cours du processus d'évaluation du modèle grâce à l'analyse par intervalles pour appréhender toutes les situations possibles en construisant des tubes de trajectoires.

Cette propriété est intéressante pour de nombreux problèmes à traiter dans le domaine des sciences de l'ingénieur (Jaulin et al, 2001 [46]) comme l'estimation d'état (Hadj-Sadok et al., 2001 [47]), (Jaulin et al., 1993,1996 [48, 49]), (Flaus et al., 2003 [50]), (Adrot et al., 2009 [51]), la surveillance (détection et localisation de défauts ou diagnostic) (Armengol et al, 1999 [52]), (Puig et al, 2002 [53]), (Ploix et al, 2000, 2006 [54, 55]), (Adrot et al, 2008 [56]), (Combastel et al, 2009 [57]), (Puig et al,2009 [58]), (Adrot et al, 2010 [59]), l'analyse de sûreté (Shahriari, 2007 [60]), (Raka, 2011 [61]),...

La figure 1.16 présente le principe général de la propagation des incertitudes avec l'approche ensembliste.

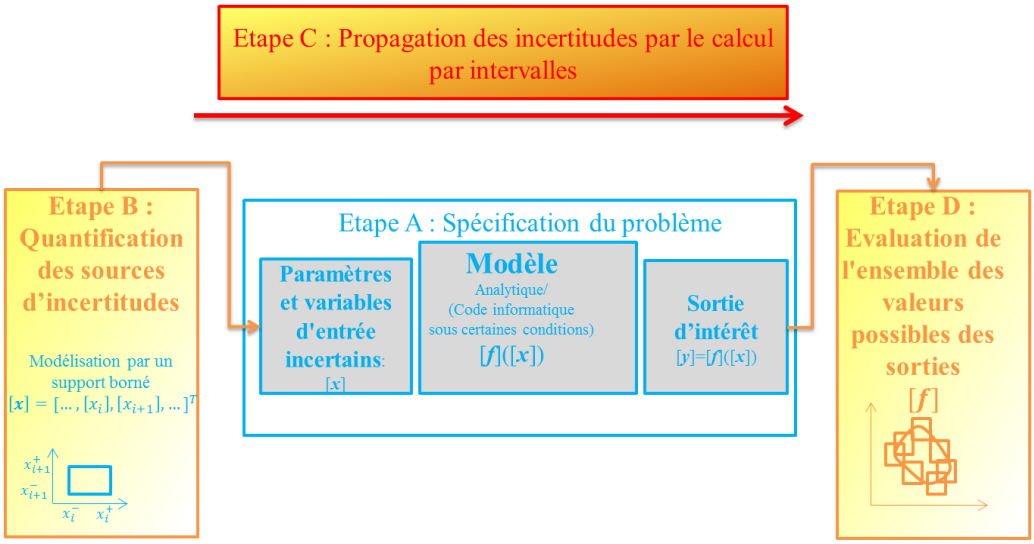

FIGURE 1.16: Principe général de la propagation d'incertitudes par l'analyse par intervalles

L'approche ensembliste consiste grossièrement à transformer le modèle analytique étudié en un modèle intervalle en remplaçant chaque grandeur incertaine x par son support intervalle $[x]$ et chaque opérateur sur les réels par sa version intervalle [28, 38]. Pour un modèle réel noté $y = f(x)$, sa version intervalle sera notée $[y] = [f]([x])$.

Exemple 3. Nous considérons le modèle linéaire suivant :

$$f(x_1, x_2) = x_1 - x_2$$

Tel que x_1 et x_2 définissent 2 variables pouvant prendre des valeurs entre -1 et 1, en d'autre terme $x_1 \in [-1, 1]$ et $x_2 \in [-1, 1]$.

L'objectif est de déterminer l'ensemble image $\{f(x_1, x_2)/x_1 \in [-1, 1] \wedge x_2 \in [-1, 1]\}$ de la fonction f connaissant les supports des variables d'entrée x_1 et x_2. On construit donc la fonction intervalle associée $[f]([x_1], [x_2]) = [x_1] - [x_2]$. Le résultat de la soustraction entre 2 intervalles bornés est un intervalle dont les bornes se déduisent de celles de $[x_1]$ et $[x_2]$ comme suit : $[x_1] - [x_2] = [x_1^- - x_2^+, x_1^+ - x_2^-]$.

Cette règle de calcul par intervalles conduit à :

$[f]([x_1], [x_2]) = [-1, 1] - [-1, 1] = [-2, 2]$.

On retrouve bien que la soustraction de deux valeurs prises de manière quelconque entre -1 et 1 conduit bien à une valeur comprise en -2 et 2. Ainsi si x_1 et x_2 désignent deux entrées incertaines, la sortie $y = f(x1, x2)$ est elle aussi incertaine et $[-2, 2]$ correspond à l'ensemble des valeurs qu'elle peut prendre.

En conclusion, les points forts des méthodes ensemblistes résident dans leur capacité à être à la fois :

- un outil de calcul numérique permettant de traiter une grande variété de problèmes mathématiques (optimisation globale de critères non convexes, inversion de modèles lors de la phase d'identification des paramètres ou du recalage sur des observations,...) tout en garantissant l'obtention de toutes les solutions au problème posé,
- un outil de modélisation en contexte incertain permettant de prendre en compte l'imprécision sur la valeur des nombres (que ce soit pour les erreurs d'arrondis liées aux calculs ou aux incertitudes sur la valeur de paramètres ou de grandeurs observées) et la manipulation des modèles obtenus pour propager ces incertitudes. Grâce au calcul par intervalles, évaluer l'intervalle de sortie d'un modèle peut se faire très rapidement en un coup. La notion de calcul garanti assure que toutes les valeurs de la sortie seront appréhendées, en résulte en contre partie comme point faible le fait que l'intervalle de sortie calculé pourra être surestimé.

Le point faible de cette approche réside dans la nécessité de travailler sur une forme analytique du modèle pour en concevoir une version intervalle, mais elle est cependant applicable sur des modèles par code informatique sous certaines conditions (en étudiant la monotonie du modèle par rapport aux différentes entrées, il est possible d'évaluer les bornes supérieures et inférieures de la sortie sans faire à proprement parlé de calcul par intervalles). L'approche ensembliste, et notamment les règles de calcul par intervalles, seront plus amplement détaillées dans la section 2.1.2.

1.4.4 Analyse de sensibilité

Une analyse de sensibilité est une étape préalable à la propagation des incertitudes, utile afin d'identifier les principales sources d'incertitude parmi les différentes entrées d'un modèle d'effets. L'intérêt de cette analyse consiste à connaitre les entrées dont les variations ont une forte influence sur la variation de la sortie du modèle. En effet, les autres entrées, qui ont une influence moindre, nécessitent moins d'attention lors de l'étape de modélisation où il n'est pas nécessaire de faire des efforts supplémentaires pour mieux connaitre ces grandeurs incertaines et ainsi améliorer la précision de leurs supports (et par la même occasion avoir une meilleure connaissance de la sortie). Il peut même être intéressant de fixer ces entrées à des valeurs nominales pour réduire le nombre de grandeurs incertaines dans des modèles coûteux en temps de calcul, et ce sans trop porter atteinte à la complétude du modèle, c'est-à-dire sans perdre trop de solutions sur le support de la sortie.

Dans les applications dédiées à la gestion du TMD, l'analyse de sensibilité prend tout son sens. En effet, la nécessité de connaître l'influence des incertitudes affectant les entrées sur la zone de danger évaluée (c'est-à-dire l'incertitude sur le niveau des risque), ne doit pas en contre partie se traduire par une réactivité trop faible à cause de temps de calculs importants lors des simulations du modèle d'effets, sans quoi la mise en place de barrières pour empêcher d'autres véhicules de TMD de s'approcher du lieu de l'accident risque d'être trop tardive. De plus, ces modèles utilisant de nombreuses entrées parfois difficiles à déterminer immédiatement après l'accident, l'analyse de sensibilité apporte des informations sur les entrées les plus influentes dont la recherche est à privilégier pour correctement renseigner le modèle d'effets utilisé.

L'analyse de sensibilité (AS) étudie comment la variation de la sortie d'un modèle peut être attribuée aux variations des différentes entrées. Nous souhaitons ainsi quantifier l'impact de l'incertitude attachée aux entrées du modèle sur la sortie prédite par le modèle. Plus particulièrement, nous souhaitons identifier les entrées par rapports auxquelles la sortie est "**sensible**". **L'indice de sensibilité** d'une entrée quantifie donc l'importance de l'influence de son incertitude sur la sortie. Il s'agit de la part de la variabilité de la sortie expliquée par la variabilité de l'entrée. L'analyse de sensibilité consiste à calculer les indices de sensibilité de chacune des entrées, ce qui permet de classer ces dernières en fonction de leur influence sur la sortie.
Il existe plusieurs méthodes d'analyse de sensibilité selon le domaine d'application et le niveau de complexité du système, qui peuvent être regroupées en trois catégories (Saltelli et al., 2004)[62] :
– les méthodes de criblage ou « screening » ,
– les méthodes d'analyse de sensibilité locale ,

– les méthodes d'analyse de sensibilité globale.

1.4.4.1 Méthodes de criblage ou « screening »

Les méthodes de criblage sont qualitatives et permettent d'identifier les entrées qui sont très influentes sur la variation de la sortie afin de faire un premier tri et de réduire le nombre d'entrées à analyser par la suite avec des méthodes plus sophistiquées et/ou plus coûteuses [63–66]. Elles sont fréquemment utilisées dans le cas où le modèle contient un nombre considérable d'entrées incertaines et nécessite un temps de calcul très élevé.
Plusieurs techniques ont été développées pour la méthode de criblage, parmi lesquelles existe la technique du criblage par groupe [65], qui consiste à créer un certain nombre de groupes d'entrées et à identifier les plus influents. En répétant progressivement l'opération en conservant les groupes influents, on extrait au final les entrées influentes. Cette technique nécessite la connaissance du sens de variation de la sortie en fonction du sens de variation de chaque entrée, connaissance qui n'est pas toujours disponible.
La méthode de criblage la plus utilisée pour l'analyse de sensibilité est la méthode de Morris [67, 68]. Cette méthode consiste à répéter r fois ($r = 5$ à 10) un plan OAT (One at A Time) aléatoirement dans l'espace des entrées, c'est-à-dire en ne faisant varier qu'une seule entrée à la fois, les autres restants fixes, en discrétisant chaque entrée en un nombre convenable de niveaux. La méthode de Morris permet ainsi de classer les entrées influentes selon trois catégories :
– les entrées ayant des effets négligeables,
– les entrées ayant des effets linéaires et sans interaction,
– les entrées ayant des effets non linéaires et/ou avec interactions.

1.4.4.2 Méthodes d'analyse de sensibilité locale

L'analyse de sensibilité locale est une méthode quantitative, qui vise à déterminer l'impact local des entrées sur la sortie, en calculant un indice de sensibilité représentant les variations de la sortie du modèle suite à de petites perturbations autour d'une valeur choisie de l'entrée étudiée. A chaque entrée est attribuée une valeur nominale de référence. L'analyse de sensibilité s'effectue en faisant varier la valeur de référence de l'entrée étudiée tout en fixant toutes les autres entrées à leurs valeurs nominales. Ces méthodes, bien que simples et rapides, peuvent apparaitre insuffisantes pour caractériser la sensibilité de modèles complexes car elles ne prennent pas en compte les interactions entre les entrées [3, 69].
Considérons un modèle mathématique qui, à un ensemble d'entrées aléatoires $\boldsymbol{x} = (x_1, ..., x_p)$, fait correspondre, via une fonction f déterministe, une variable de sortie y (ou réponse)

aléatoire :

$$y = f(\boldsymbol{x}) \tag{1.2}$$

La sensibilité S_i de la réponse par rapport à une variation de faible amplitude de l'entrée x_i, s'obtient en évaluant les dérivées partielles,

$$S_i = \frac{\partial y}{\partial x_i} \tag{1.3}$$

où ∂x_i est l'écart entre la valeur perturbée x_i^* de x_i et sa valeur nominale x_{i0}, et ∂y est l'écart entre la réponse du modèle pour la valeur de x_i^*, $y = f(\dots, x_i^*, \dots)$, et la réponse « nominale » du modèle pour x_{i0}, $y_0 = f(\dots, x_{i0}, \dots)$.

La figure 1.17 présente le résultat de la variation d'un modèle étudié. Elle montre que la variation de l'entrée autour du point A a un impact très fort sur la variation de la sortie par rapport a l'impact faible au point B. C'est avec l'analyse de sensibilité locale qu'on pourra identifier les zones de l'espace d'entrée où la variation de la sortie du modèle est importante ou faible [5].

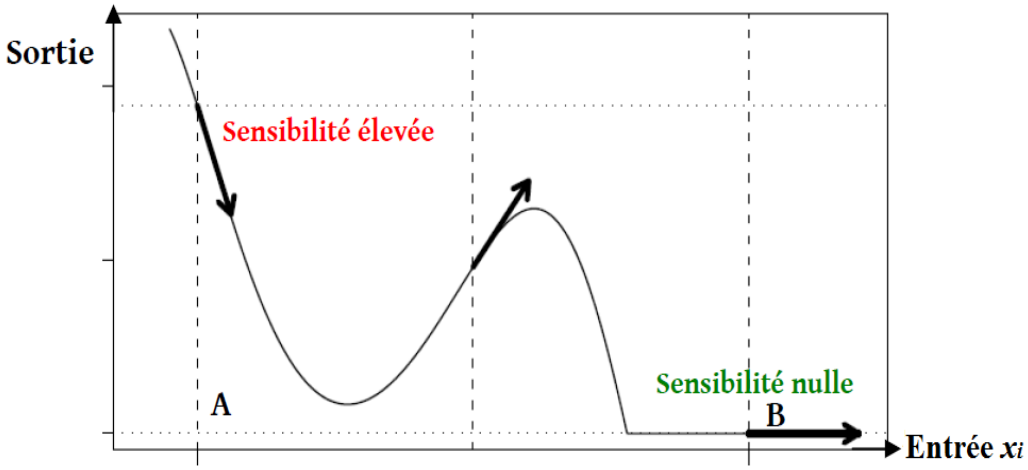

FIGURE 1.17: L'analyse de sensibilité locale

Généralement, les variations des entrées du modèle lors de l'analyse sensibilité locales sont de l'ordre de ±5%, ±8%, ±10%... autour de la valeur nominale.

Exemple 4. [3] Si une variation de 8% de la valeur de la grandeur d'entrée induit une variation de 12% de la valeur de la grandeur de sortie, alors S = 12/8 = 1,5.

- Si $|S|$ est égal à 1, cela signifie que la variation de la valeur d'entrée induit en proportion la même variation en sortie.
- Si $|S|$ est inférieur à 1, cela signifie que la variation de la valeur d'entrée induit en proportion une plus faible variation en sortie.
- Si $|S|$ est supérieur à 1, cela signifie que la variation de la valeur d'entrée induit en proportion une plus forte variation en sortie.

– Si S est négatif, cela signifie que la grandeur de sortie varie dans le sens inverse de la grandeur d'entrée.

On considère en général que la sensibilité devient significative dès lors que l'indicateur S dépasse la valeur 0,5 en valeur absolue.

1.4.4.3 Méthodes d'analyse de sensibilité globale

Les méthodes d'analyse de sensibilité globale permettent d'analyser un modèle numérique en étudiant l'impact de la variabilité des facteurs d'entrée du modèle sur la variable de sortie. Elles permettent de prendre en compte la densité de probabilité de chaque entrée et de faire varier toutes les entrées simultanément[62].
Nous nous intéressons plus précisément aux méthodes basées sur l'étude de la variance qui calculent des indices de sensibilité globale pour quantifier l'influence des entrées basés sur l'hypothèse d'indépendance de celles-ci. Pour le calcul de ces indices, des méthodes spécifiques pour les modèles linéaires et/ou monotones existent, ainsi que des méthodes plus générales, qui ne font aucune hypothèse sur le modèle. Ce dernier type de techniques comprend la méthode de Sobol, que nous utilisons dans ce travail et que nous allons détailler plus spécifiquement par la suite. Auparavant, nous allons définir les indices de sensibilité globale.

Indice SRC (Standardized Regression Coefficient) et coefficient de corrélation

L'indice **SRC** exprime la part de variance de la réponse y due à la variance de la variable x_i.
Supposons que le modèle étudié soit linéaire, et qu'il s'écrive sous la forme suivante :

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$$

Il est possible de quantifier la sensibilité de y à x_i par le rapport de la part de variance due à x_i sur la variance totale.
L'indicateur ainsi construit est l'indice de sensibilité SRC, défini par :

$$SRC_i = \frac{\beta_i^2 V(x_i)}{V(y)}$$

où $V(.)$ représente la variance de la grandeur placée en argument.

Indice PCC (Partial Correlation Coefficient)

Néanmoins, il est parfois difficile d'apprécier la sensibilité de y due à une variable d'entrée x_i, si les simulations successives du modèle sont faites pour des valeurs différentes de toutes les variables d'entrée. En effet, la corrélation entre y et x_i peut être due à une tierce variable. On rencontre parfois en pratique des cas où une corrélation entre deux variables est observée, alors qu'elle n'est en fait due qu'à une corrélation avec une troisième variable [70].

Pour contrer cet effet, l'indice de corrélation partielle PCC a été proposé. Il permet d'évaluer la sensibilité de y à x_i en éliminant l'effet des autres variables, toujours donc sous l'hypothèse de linéarité du modèle.

L'indice de corrélation partielle de y par rapport à x_i, exprimant la sensibilité de y à x_i, est donné par :

$$PCC_i = \rho_{Y,x_i|x_{\sim i}} = \frac{Cov(y, x_i|x_{\sim i})}{\sqrt{V(y|x_{\sim i})V(x_i|x_{\sim i})}}$$

où le terme $Cov(y, x_i|x_{\sim i})$ est la covariance partielle entre y et x_i, $V(y|x_{\sim i})$ est la variance conditionnelle de y et $x_{\sim i}$ est le vecteur $\boldsymbol{x}$ privé de sa i-ème composante.

Méthode de Sobol

Dans le cas de la méthode de Sobol, la sensibilité de la sortie par rapport aux entrées incertaines est donnée par des indices de sensibilité qui s'expriment sous la forme du rapport entre la variance de la sortie calculée lorsque certaines entrées incertaines sont fixées sur la variance de la sortie calculée lorsque toutes les entrées évoluent librement. Le principe de la méthode repose sur une décomposition fonctionnelle de la variance de type ANOVA (acronyme de ANalysis Of VAriance) qui permet de construire des indices de différents ordres.

- Un indice de sensibilité du 1^{er} ordre qui permet d'étudier l'effet sur la sortie d'une entrée seule, évoluant sur tout son support de variation,
- Un indice de sensibilité du second ordre qui permet d'étudier l'effet des interactions de 2 entrées sur la sortie
- Un indice de sensibilité d'ordre total qui permet d'étudier l'effet de l'entrée seule et les effets de son interaction avec toutes les autres entrées sur la variation de la sortie.

Des indices d'ordre supérieur à 2 peuvent être également calculés. L'estimation de ces indices qui nécessite le calcul de variances de la sortie conditionnelles relativement à certaines entrées fixées repose sur des stratégies d'échantillonnage comme :

1. l'échantillonnage de type Monte Carlo,
2. l'échantillonnage de type Quasi-Monte Carlo (QMC)[71, 72],
3. l'échantillonnage de type Quasi-Monte Carlo Randomisé [73],
4. l'échantillonnage de type hyper cube latin(LHS)[74].

L'objectif est de tirer plus efficacement les valeurs d'entrée pour travailler sur des échantillons plus restreints. L'estimation de ces indices utilisés dans notre travail sera détaillée dans la section 4.1.1.1.

Méthode de McKay

La méthode de McKay (1995)[75] consiste à estimer les indices de sensibilité du premier ordre en se basant sur l'échantillonnage conditionnel par hypercube latin répliqué. À partir d'un N-échantillon créé selon le plan d'échantillonnage par hypercube latin, on crée r réplications (paquet de N lignes) en permutant indépendamment et aléatoirement les N valeurs de la variable dont on recherche l'indice (i.e. colonne). La réunion de ces r réplications donnera Nr échantillons pour chaque variable.

Ce schéma d'échantillonnage par hypercube latin répliqué peut être représenté par la figure 1.18 [70].

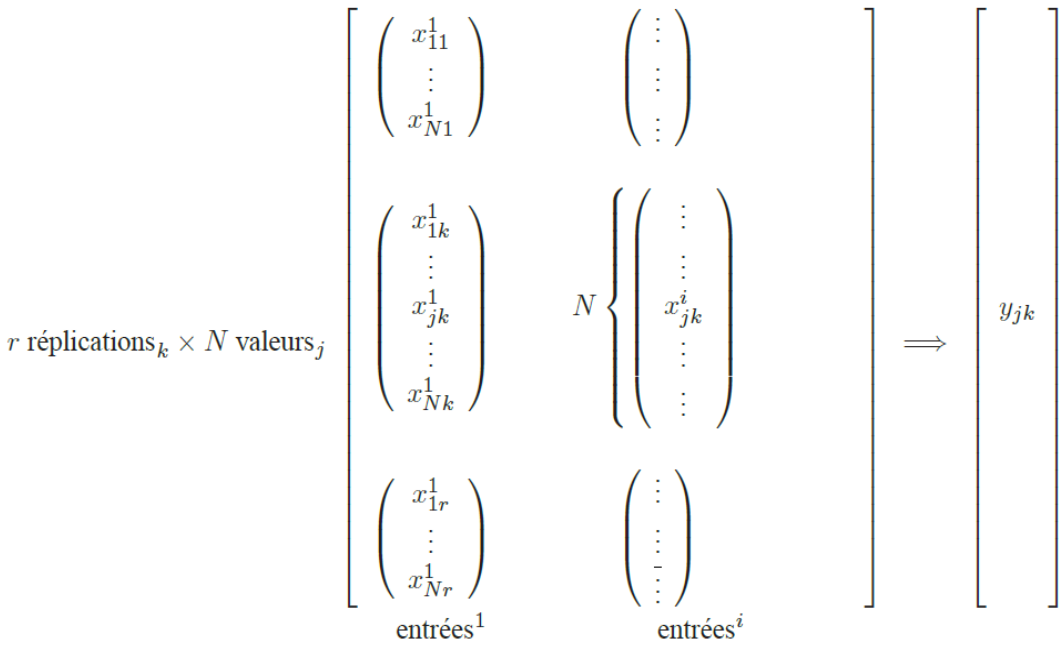

FIGURE 1.18: Échantillonnage par hypercube latin répliqué.

avec :

$1 \leq j \leq N$: N valeurs de chaque entrée (prises dans des intervalles équiprobables),

$1 \leq k \leq r$: r permutations des N-échantillons des entrées,
$1 \leq i \leq p$: p entrées.

Les moyennes suivantes des valeurs de la sortie sont alors définies :

$$\bar{y}_{j.} = \frac{1}{r} \sum_{k=1}^{r} y_{jk} \ , \ \bar{y} = \frac{1}{N} \sum_{j=1}^{N} \bar{y}_{j.}$$

où $\bar{y}_{j.}$ est la moyenne inter réplications et $\bar{y}$ est la moyenne sur toutes les valeurs de y. L'indice de sensibilité du premier ordre de la variable x_i est alors estimé par :

$$S_i = \frac{r \sum_{j=1}^{N} (\bar{y}_{j.} - \bar{y})^2 - \frac{1}{r} \sum_{j=1}^{N} \sum_{k=1}^{r} (y_{jk}^{(i)} - \bar{y}_{j.})^2}{\sum_{j=1}^{N} \sum_{k=1}^{r} (y_{jk} - \bar{y})^2}$$

.

Par rapport à la méthode de Sobol, cette dernière l'avantage de permettre le calcul des indices de sensibilité relevant des interactions entre entrées et est celle qui est la plus rapide à implémenter.

Méthode FAST (Fourier Amplitude Sensitivity Analysis)

La méthode FAST (Fourier Amplitude Sensitivity Test) a été développée par Cukier et al. [76–78], ainsi que Schaibly et Shuler [79].
Considérons une fonction :

$$y = f(\boldsymbol{x}) = f(x_1, ..., x_p)$$

Il s'agit d'une méthode qui utilise la transformée de Fourier multi-dimensionnelle de f pour obtenir une décomposition de la variance de y. Le principe de la méthode FAST est de remplacer les décompositions multi-dimensionnelles par des décompositions uni-dimensionnelles le long d'une courbe parcourant l'espace $[0, 1]^p$. Cette courbe est définie par les entrées qui sont représentées par des fonctions sinus de fréquences différentes. Les échantillonnages sont générés ainsi :

$$x_i(s) = g_i(\sin(\omega_i s))_{i=1,..,p}$$

où les g_i sont des fonctions de transformation à déterminer, permettant un recouvrement uniforme de $[0, 1]^p$, les ω_i sont les fréquences associées à chaque facteur d'entrée, et qui sont linéairement indépendantes (aucune n'est combinaison linéaire des autres) et s est une discrétisation du domaine $[-\pi, \pi]$ respectant une distribution uniforme des

échantillons [70]. L'indice de sensibilité est donné par l'expression suivante :

$$S_i = \frac{\sum_{p=1}^{\infty} (A_{p\omega_i}^2 + B_{p\omega_i}^2)}{\sum_{p=1}^{\infty} (A_p^2 + B_p^2)}$$

où A_p et B_p sont les coefficients de la transformée de Fourrier de y en cosinus et sinus respectivement.

1.5 Objectifs de ce travail

Pour évaluer l'intensité d'un phénomène dangereux généré lors d'un accident industriel et ensuite le niveau de risque des zones soumises au transport des matières dangereuses, différents modèles de dispersion atmosphérique peuvent être utilisés en fonction du type de matières dangereuses concernées et des conditions dans lesquelles se déroulent l'accident. Un autre facteur important dans le choix du modèle de dispersion utilisé dépend des performances attendues en termes de temps de calculs et de précision. Des modèles simples et analytiques conduiront à une réponse plus rapide, mais aussi moins représentative de la réalité du terrain par rapport à des modèles plus sophistiqués et généralement définis sous la forme de code informatique beaucoup plus lourd à traiter. Dans ces modèles existent un certain nombre de paramètres ou de variables d'entrée connus avec imprécision. Pour avoir une évaluation pertinente des effets et correctement réaliser la cartographie en déterminant la zone de danger pour laquelle le niveau de concentration en gaz toxique est supérieur à un seuil préconisé, il est nécessaire d'identifier et de prendre en compte les incertitudes sur les entrées des modèles d'effets afin de fournir un niveau de risque fiable. En effet, il existe alors une zone d'incertitude autour de la zone de danger où on ne peut plus dire avec certitude si la concentration en gaz est systématiquement plus élevée que le seuil préconisé. Ce peut être le cas, ou non, en fonction des valeurs réellement prises par les entrées incertaines, parmi toutes celles qu'elle sont supposées pouvoir prendre. Il est aussi important de garder à l'esprit que la nature du modèle (analytique, code informatique) a une incidence sur les méthodes que l'on pourra utiliser pour propager les incertitudes.

Le **premier objectif** consiste donc à évaluer et propager les incertitudes sur la concentration induites par les grandeurs d'entrée incertaines lors de l'évaluation des modèles de dispersion. Deux approches sont utilisées pour modéliser et propager les incertitudes : l'approche ensembliste pour les modèles analytiques et l'approche probabiliste (Monte-Carlo) qui est plus classique et utilisable que le modèle de dispersion soit analytique ou défini par du code informatique. L'objectif consiste à comparer les deux approches pour

connaitre leur avantages et inconvénients en termes de précision et temps de calcul afin de résoudre le problème proposé.

Le **deuxième objectif** consiste à traiter le problème de cartographie en évaluant la zone de danger et la zone d'incertitudes, qui sont utilisées pour évaluer l'intensité des risques dans la zone impactée. La réalisation des cartographies nécessite de résoudre un problème d'inversion, les modèles d'effets évaluant des concentrations en des positions données alors que l'on recherche les positions où la concentration serait plus grande qu'un seuil donné (aux incertitudes près). En fonction de la nature des modèles utilisés (analytique ou par code), nous mettrons en oeuvre une méthode probabiliste (Monte Carlo) spécifique au modèle d'effets à inverser et une méthode ensembliste. Cette dernière, en reformulant ce problème sous la forme d'un ensemble de contraintes à satisfaire (CSP) et en le résolvant ensuite par inversion ensembliste reste une démarche générique, indépendante du modèle utilisé à partir du moment où on peut en obtenir une version intervalle, et capable de traiter d'autres types de problèmes comme de la surveillance de capteur ou la localisation du camion accidenté.

Le **troisième objectif** consiste à proposer une méthodologie générale pour réaliser les cartographies dans le but d'améliorer les performances en termes de temps du calcul et de précision. Cette méthodologie s'appuie sur 3 étapes : l'analyse préalable des modèles d'effets utilisés, la proposition d'une nouvelle approche pour la propagation des incertitudes mixant l'approche probabiliste et ensembliste tirant notamment partie des avantages des deux approches précitées, et utilisable pour n'importe quel type de modèle d'effets spatialisé et statique puis finalement la réalisation des cartographies en inversant les modèles d'effets. L'analyse de sensibilité présente dans la première étape s'adresse classiquement à des modèles probabilistes. Nous discuterons de la validité d'utiliser des indices de type Sobol dans le cas de modèles intervalles et nous proposerons un nouvel indice de sensibilité purement intervalle cette fois-ci.

1.6 Conclusion

Dans ce chapitre, nous avons précisé dans un premier temps le contexte applicatif de ce travail en rappelant les principes généraux de la gestion des risques, notamment liés au Transport de Matières Dangereuses (TMD). Ensuite nous avons présenté le projet GEOFENCING-MD qui finance ce travail et qui vise à suivre en temps réel les camions transportant des matières dangereuses. A partir de ces éléments, nous avons posé la problématique de cette thèse qui consiste à évaluer le niveau de risque de zones soumises à un relâchement accidentel d'un gaz toxique, et ce en utilisant des modèles d'effets

(analytique ou code informatique) tout en prenant en compte les incertitudes affectant certaines de leurs entrées.

Ensuite nous avons introduit un certain nombre d'éléments nécessaires à la bonne compréhension des chapitres suivants. Nous avons ainsi rappelé les différents types de modèles d'effets, dont le modèle gaussien de dispersion atmosphérique pour les gaz passifs ou les modèles de type intégral pour les gaz lourds, qui seront par la suite utilisés dans les applications. Nous avons présenté les sources d'incertitudes sur les données utilisées dans ces modèles et les différentes méthodes pour modéliser et propager les incertitudes, notamment les approches ensembliste et probabiliste qui ont retenu notre attention. Nous en avons profité pour en préciser les principes généraux et nous avons brièvement évoqué leurs avantages respectifs. Les avantages de l'approche ensembliste résident dans la possibilité de traiter une grande variété de problèmes mathématiques et d'être un outil de modélisation permettant de prendre en compte et de propager à faible cout et de manière garantie les incertitudes de modèle. Le point faible réside dans la nécessité de travailler sur une forme analytique du modèle. Concernant la méthode probabiliste, les points forts résident premièrement dans une modélisation plus riche des grandeurs incertaines en utilisant à la fois des informations sur la distribution et le support, deuxièmement dans la capacité à propager les incertitudes de manière simple en simulant un grand nombre de fois le modèle, et ce avec n'importe quel type de modèle (analytique ou code informatique). Les points faibles proviennent de la difficulté à savoir si les entrées incertaines étudiées suivent vraiment les lois de distributions choisies. La deuxième difficulté réside dans la complexité à déterminer d'un point de vue pratique la nature de la distribution de la sortie de modèle. Pour terminer, la dernière partie se focalise sur les différentes méthodes d'analyse de sensibilité, notamment globale, qui sont des outils classiquement utilisés en contexte incertain. Le but est d'identifier les principales sources d'incertitudes parmi les différentes entrées d'un modèle de dispersion dans le but de connaitre les entrées dont les variations ont une forte influence sur la variation de la sortie du modèle.

Finalement, nous avons posé les objectifs de ce travail qui consistent premièrement à propager les incertitudes sur la sortie de modèles de dispersion, deuxièmement à traiter des problèmes de cartographie en évaluant la zone de danger et sa zone d'incertitude et troisièmement proposer une méthodologie générale pour réalise ces cartographies dans le but d'améliorer les performances en termes de temps du calcul et de précision. Le contexte de la thèse et les objectifs ayant été précisés, le chapitre 2 va s'attarder à détailler les méthodes de propagation d'incertitudes utilisées.

Chapitre 2

Propagation des incertitudes : méthodes et application

Dans ce chapitre, l'objectif est de propager dans les modèles de dispersion utilisés les incertitudes provenant des grandeurs d'entrée et ainsi d'évaluer l'incertitude qui en résulte sur la concentration en gaz. Deux approches sont utilisées pour modéliser et propager les incertitudes : l'approche ensembliste pour les modèles analytiques et l'approche probabiliste (Monte-Carlo) qui est plus classique et reste utilisable que le modèle de dispersion soit analytique ou défini par du code informatique. Ces approches sont ensuite comparées pour en faire ressortir les avantages et inconvénients en termes de temps de calcul, mais aussi de précision. Le terme précision signifie ici que les méthodes utilisées ne seront capables en un temps raisonnable que de fournir une approximation du support de la concentration. On cherche donc à faire en sorte que le support calculé soit le plus proche possible du support exact recherché.

2.1 Méthodes de propagation des incertitudes

La propagation des incertitudes dans un modèle est devenu un enjeu important en calcul scientifique puisqu'il s'agit d'une étape essentielle dès lors où l'on souhaite prendre en compte les incertitudes de modèle. Elle consiste à estimer la dispersion de la sortie du modèle due aux incertitudes associées aux entrées du modèle. Cela revient à minima à estimer, suivant la méthode utilisée, le support de la sortie, c'est-à-dire toutes les valeurs qu'elle est susceptible de prendre, ou bien un intervalle de confiance (par exemple pour un degré de confiance de 95%) si le support n'est pas borné, permettant de délimiter sa variation entre 2 limites inférieure et supérieure compte tenu des entrées incertaines. Il

est aussi possible dans le cas d'une approche probabiliste d'estimer la distribution de la sortie. Cette information ne sera pas recherchée dans notre cas car :

- l'information sur le support est suffisante pour atteindre nos objectifs,
- l'approche ensembliste utilisée ne le permet pas,
- la complexité des modèles utilisés rend cette information à la fois difficile à extraire puis à manipuler.

Il s'agit ici d'analyser la propagation des incertitudes des entrées sur la variable d'intérêt y via la relation :

$$y = f(x_1, x_2, ..., x_p) \tag{2.1}$$

où $x_1, x_2, \ldots, x_p$ sont les entrées considérées comme des variables incertaines et f désigne le modèle utilisé.

Nous présentons ici deux approches distinctes. La première est une approche probabiliste (Monte-Carlo) et la deuxième une approche ensembliste qui visent à représenter une grandeur incertaine respectivement par une variable aléatoire suivant une loi de probabilité donnée ou par un support borné.

2.1.1 Approche Probabiliste : Monte Carlo (MC)

Les méthodes de Monte Carlo sont des outils de simulation utilisés dans de nombreux domaines scientifiques tels que la finance (mesure de risques et fluctuations boursières), l'environnement (gestion du trafic routier,...), les mathématiques (calculs d'intégrales,...) mais aussi la chimie, la biologie et plus sporadiquement la médecine et surtout la physique. Les méthodes de Monte-Carlo sont des techniques appartenant à la fois aux domaines des mathématiques et de l'informatique, qui visent à évaluer une quantité déterministe en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. Le but est d'obtenir numériquement l'ensemble des réponses d'un modèle obtenues à partir de multiples réalisations des variables aléatoires d'entrées.

La méthode de Monte Carlo a connu un essor considérable à partir de la fin de la seconde guerre mondiale, où elle a été utilisée pour résoudre des problèmes complexes essentiellement dans le cadre du projet secret de la défense américaine "Manhattan" concernant le développement de l'arme nucléaire. Cette époque correspond à l'apparition des premiers ordinateurs au milieu des années 1940. Le terme Monte Carlo, qui fait allusion aux jeux de hasard pratiqués à Monte-Carlo, a été développé par Metropolis et Ulam(1949)[80]. Par la suite d'autres auteurs ont contribué à son développement, comme Hammersley et Morton (1956)[81], Hammersley et Handscomb (1964)[82], Haber (1966)[83], Kuipers et Niederreiter (1974)[84], Boyle (1977)[85] et Niederreiter (1978)[72].

Les méthodes de Monte Carlo peuvent être vues comme des méthodes d'approximation, au sens statistique du terme. La description la plus utilisée consiste à dire que les méthodes de ce type se caractérisent par l'utilisation de tirages aléatoires pour résoudre des problèmes centrés sur le calcul d'une valeur numérique. C'est un outil standard pour l'analyse des systèmes complexes multidimensionnels, qui peut servir à propager l'incertitude provenant des entrées d'un modèle mais, elle requiert beaucoup de ressources en calcul informatique.
Dans notre étude, nous ne nous intéresserons pour les entrées incertaines qu'à des lois de distributions de variables aléatoires uniformément répartis entre 2 bornes a et b. La raison est double. Premièrement, c'est une façon simple et pratique de travailler sur des supports bornés et de pouvoir par la suite comparer les résultats obtenus avec l'approche ensembliste qui utilisera des supports intervalles identiques. La seconde raison est que travaillant sur des modèles décrivant des phénomènes physiques, les entrées incertaines concernées ont des domaines de définitions bornés. Utiliser une loi de distribution sur un support non borné pourrait générer lors de l'échantillonnage des valeurs n'ayant pas de sens physique (vitesse du vent négative par exemple) même si la probabilité reste faible, ou pour lesquels le modèle n'est pas défini. Une alternative non utilisée ici consiste à travailler sur des lois tronquées.

2.1.1.1 Principe de la méthode de Monte Carlo

Morgan et Henrion (1990)[26], Gentle (2003)[27], Glasserman (2004) [34]et Ayyub et Klir (2006)[35] en font la description.
Le principe de base est plutôt simple :

1. Définir la sortie, les facteurs d'entrée et expliciter le modèle mathématique ou numérique,
2. Associer une loi de distribution à chaque entrée du modèle sur lequel on effectue une simulation de Monte Carlo,
3. Préciser le nombre d'échantillons N qui peut être choisi arbitrairement en fonction du nombre d'entrées incertaines considérées et du temps de calcul associé à une simulation du modèle (dans notre étude, ce nombre sera souvent de l'ordre de 10^p, p restant faible),
4. Générer un ensemble de valeurs aléatoires (aussi appelées des échantillons aléatoires) de ces distributions,
5. Générer une matrice contenant la combinaison des échantillons effectués pour chaque entrée,
6. Évaluer la sortie pour toutes les combinaisons des entrées en calculant le minimum et le maximum de la sortie.

2.1.1.2 Implémentation de la méthode de Monte Carlo

La figure 2.1 illustre les étapes de calcul de la propagation d'incertitudes en utilisant la méthode Monte Carlo.

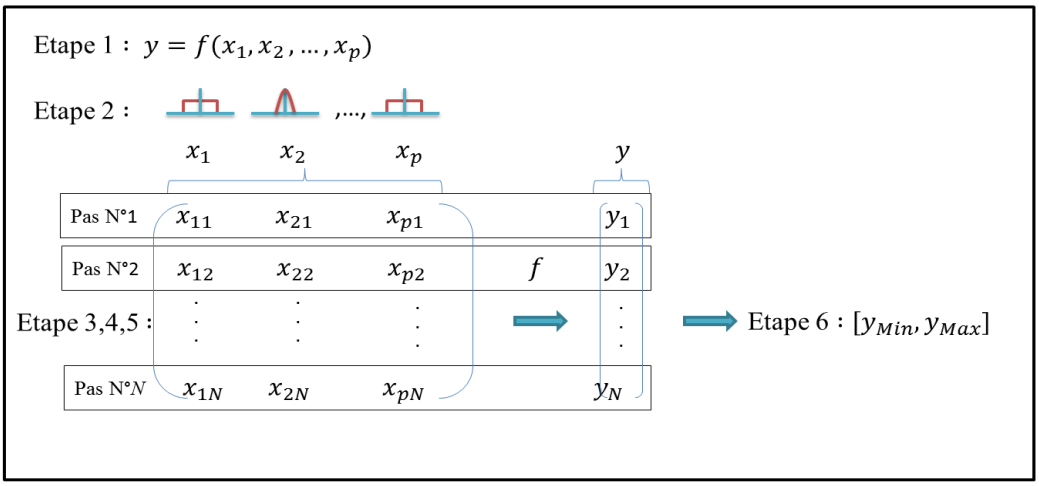

FIGURE 2.1: Les étapes de la propagation d'incertitudes

Trois étapes principales sont exécutées au cours de la mise en œuvre de la simulation de Monte Carlo : la génération de N échantillons de p entrées selon les lois de distribution associées, l'évaluation de la sortie du modèle pour chaque échantillon et enfin l'estimation du support de la sortie du modèle et de l'incertitude associée. Le résultat final de la propagation d'incertitude est une estimation de l'intervalle de confiance $[y_{Min}, y_{Max}]$ (avec un degré de confiance de 100%) de la sortie du modèle, en d'autres termes une estimation du plus petit support borné contenant toutes les valeurs possibles de la sortie. Reposant sur les valeurs maximale y_{Max} et minimale y_{Min} obtenues pour les N valeurs calculées $y_1, y_2, ..., y_N$, l'indicateur u_{-MC} caractérisant l'incertitude sur la sortie est défini par :

$$u_{-MC} = \frac{y_{Max} - y_{Min}}{2} \times \frac{100}{y_{ValeurNominale}}. \tag{2.2}$$

où $y_{ValeurNominale}$ désigne la valeur de la sortie calculée à partir des valeurs nominales des entrées du modèles, c'est-à-dire sans les incertitudes sur ces entrées.

Cet indicateur exprime en pourcentage la largeur de l'intervalle de confiance calculé rapporté à la valeur nominale de la sortie. L'idée est qu'un intervalle de confiance de largeur 1 par exemple n'a pas la même signification si la grandeur de sortie prend ses valeurs autour de 0,1 ou de 10. Cet indicateur est d'autant plus élevé que l'incertitude sur la sortie est importante et permettra par la suite de plus facilement comparer les résultats entre les différentes approches.

La méthode de Monte Carlo est celle la plus utilisée pour la propagation d'incertitudes. Selon la complexité du modèle, une simulation peut prendre un temps substantiel (quelques heures) afin d'obtenir un degré de précision satisfaisant.

Comme il est impossible d'effectuer une infinité d'itérations, il y a toujours une erreur sur les résultats calculés à l'aide de MC. Cette erreur est estimée par l'erreur type (standard error) $\sigma/\sqrt{N}$, où σ est l'écart type de la distribution résultante de la sortie et N le nombre d'itérations.

Exemple 5. Considérons le cas où on cherche à estimer la valeur de π.

Soit un point M de coordonnées (x, y), telles que x et y sont 2 variables aléatoires indépendantes équidistribuées sur l'intervalle $[0, 1]$. Le point M appartient au disque de centre $(0, 0)$ de rayon 1 si $x^2 + y^2 \leqslant 1$.
La probabilité que le point M appartienne au disque est $\frac{\pi}{4}$. Si u et N désignent respectivement le nombre de points tirés dans le disque et le nombre de tirages effectués, alors l'approximation de π est donnée par $\pi = \frac{u}{N} \times 4$. Pour différentes valeurs de N, les résultats suivants sont obtenus :
– avec $N = 1000$ l'approximation de π est 3.112,
– avec $N = 10000$ l'approximation de π est 3.1352,
– avec $N = 100000$ l'approximation de π est 3.13552,
et l'on observe bien que la précision de l'estimation augmente avec le nombre d'échantillons.

2.1.1.3 Avantages et inconvénients de l'approche Monte Carlo

La méthode de Monte Carlo possède deux avantages principaux [86] :

1. Elle est très facile à mettre en place.
2. Elle est utilisable pour tout type de modèle (analytique, code informatique).

Cette approche possède trois inconvénients :

1. Elle ne permet de prendre en compte qu'un nombre limité de cas et donc sous-estime l'intervalle de confiance.
2. La précision de l'intervalle de confiance dépend du nombre de tirages,
3. Elle reste limitée à un faible nombre p d'entrées.

2.1.1.4 Méthode Quasi-Monte Carlo

Le nom "Quasi-Monte Carlo" fut employé pour la première fois dans un rapport de Richtmyer (1951)[87]. A l'origine, les méthodes de Quasi-Monte Carlo devaient permettre d'accélérer les simulations numériques sur des ordinateurs dont la puissance de calcul était relativement limitée. Les fondements théoriques de cette approche ont été développés notamment grace aux travaux de Halton (1960)[88], Hammersley et Handscomb (1964)[82], Haber (1966, 1970)[83, 89], Niederreiter (1972, 1978)[72, 90], Kuipers

et Niederreiter (1974)[84], Cranley et Patterson (1976)[91], Faure (1981, 1982)[92, 93] ou Zinterhof (1987)[94].

Le principal but de la méthode Quasi-Monte Carlo est de trouver de meilleures estimations d'erreur que les erreurs probabilistes des méthodes Monte Carlo et d'avoir une vitesse de convergence plus rapide. La vitesse de convergence en $\frac{1}{\sqrt{N}}$ provient de la nature stochastique de l'échantillon utilisé dans la méthode Monte Carlo classique tandis que la méthode Quasi-Monte Carlo permet d'accélerer la convegence jusqu'à $\frac{1}{N}$. En effet, le tirage des points avec la méthode de Monte Carlo est construit indépendamment, de sorte que l'on observe la formation d'agrégats dans certaines régions du domaine d'intégration (phénomène de sur-échantillonnage), tandis que d'autres régions restent entièrement vides (phénomène de sous-échantillonnage). Le Quasi-Monte Carlo n'est pas basé sur une génération aléatoire, mais sur une intégration sur des ensembles de points engendrés de façon déterministe et possédant des propriétés fortes d'équipartition. La méthode de Quasi-Monte Carlo possède un avantage principal qui réside dans une convergence plus rapide par rapport à la méthode de Monte Carlo classique (Finschi 1996 [95], Pagès et Xiao 1997 [96], Tuffin 1997 [97]). Toutefois, cette approche présente un défaut majeur qui n'existe pas avec la méthode de Monte Carlo dans la mesure où sa vitesse de convergence dépend de la dimension du problème (i.e. la méthode perd de son efficacité lorsque la dimension augmente).

Une autre approche possible pour la propagation d'incertitudes est l'approche ensembliste, en particulier l'analyse par intervalles. Dans les prochaines sections, nous allons présenter et détailler l'analyse par intervalles et ses différents principes de base.

2.1.2 Approche ensembliste : Analyse par intervalles

2.1.2.1 Introduction

Dans certains domaines d'étude, notamment celui de la gestion des risques qui nous concerne, les données utilisées sont souvent de nature incertaine. Plusieurs sources et types d'incertitudes sont rencontrés, comme une méconnaissance de certains paramètres du modèle, difficiles à identifier, une méconnaissance de certaines variables physiques difficiles à observer ou mesurées à la précision technologique des capteurs près, une mauvaise maîtrise des conditions initiales ou tout simplement de l'imprécision numérique liée à la représentation et la manipulation des nombres sur les calculateurs. De façon à introduire l'approche ensembliste, plaçons nous dans le contexte de la gestion de l'imprécision sur les nombres dans les calculateurs, problématique qui fut à l'origine du développement de cette approche. Les nombres réels manipulés par les calculateurs sont souvent représentés par un type de codage à précision limitée appelé flottants. Les opérations mathématiques

qui utilisent ces flottants à la place de réels provoquent un cumul d'erreurs en raison du type de codage qui conduit a travailler sur un arrondi avec un nombre fini de chiffres après la virgule à la place de la valeur exacte.

Dans certaines applications, il est intéressant de connaitre l'impact de ces incertitudes sur la solution obtenue. L'approche utilisée pour résoudre ce type de problèmes, s'appuie sur l'outil ensembliste, qui consiste à représenter les grandeurs par des ensembles de valeurs, et à étendre les opérations sur les réels à ces ensembles. Nous utilisons plus spécifiquement l'analyse par intervalles (Moore R.E., 1979[28] ; Neumaier A., 1990[38]). Au lieu de représenter une grandeur connue avec imprécision par une valeur nominale, l'idée consiste à décrire cette grandeur par son support, en l'occurrence un intervalle, c'est-à-dire l'ensemble des valeurs qu'elle est susceptible de prendre. Par exemple $\pi = 3.14159...$ sera représenté par [3.1415,3.1416] car π appartient à cet intervalle. De manière plus générale, un intervalle permet de modéliser tout type d'incertitude, constante ou dépendante du temps (bruit de mesure, paramètre ou variable physique d'un modèle,...).

Une fois déterminés les supports des entrées incertaines d'un modèle, l'analyse par intervalles permet de propager ces incertitudes dans le modèle étudié et de calculer un intervalle (ou de manière plus générale un ensemble) contenant de manière numériquement garantie toutes les valeurs possibles de la sortie. Plus précisément, l'analyse par intervalles (on parle aussi d'arithmétique ou de calcul par intervalles) définit des règles de calcul sur les bornes des intervalles en entrée pour évaluer celles de la sortie. Au delà de simplement permettre de représenter les incertitudes et les propager dans des relations mathématiques, les points forts de ces méthodes résident dans leur capacité à traiter une grande variété de problèmes mathématiques (optimisation globale, inversion de modèle...), à trouver toutes les solutions à un problème donné, à partitionner l'espace de recherche pour éliminer rapidement de grandes zones ne contenant aucune solution au lieu de le parcourir point par point et finalement à résoudre ce problème en prenant en compte l'imprécision sur la valeur des nombres.

Dans ce chapitre, nous rappellerons dans un premier temps les bases de l'analyse par intervalles et nous nous focaliserons sur la problématique de propagation des incertitudes dans un modèle (d'autres outils seront développés dans le chapitre 3 comme l'inversion ensembliste utile à la réalisation de cartographies). Une fois les opérations ensemblistes élémentaires présentées, le problème fondamental de dépendance, pouvant conduire à des majorations très importantes des bornes calculées lors de l'évaluation d'une fonction intervalle, sera présenté.

2.1.2.2 Arithmétique des intervalles

Par définition, un intervalle est un ensemble fermé et borné de nombres réels (Moore, 1979[28]), (Neumaier, 1990[38]). Si x désigne une variable réelle bornée, alors l'intervalle $[x]$ auquel elle appartient est défini par :

$$[x] = \{x \in \mathbb{R} | x^- \leq x \leq x^+\}. \tag{2.3}$$

Les nombres réels x^- et x^+ sont respectivement les bornes inférieure et supérieure de $[x]$. De manière générale, l'intervalle $[x]$ sera noté comme suit : $[x^-, x^+]$. D'un point de vue modélisation, le fait de considérer la variable bornée x comme incertaine signifie que seul son support $[x]$ est connu (la valeur courante de x étant elle inaccessible).
Les opérations arithmétiques addition $(+)$, soustraction $(-)$, multiplication $(\times)$, division $(/)$ sur les variables réelles peuvent être reformulées dans le cadre de l'analyse par intervalles. Le résultat d'une opération entre deux intervalles est un intervalle qui contient tous les résultats des opérations entre les éléments des deux intervalles. Le résultat d'une opération entre deux intervalles de bornes finies est obtenu en travaillant uniquement sur leurs bornes. Pour chaque opération arithmétique, élément de l'ensemble $\{+, -, \times, /\}$, il est possible d'écrire :

$$[x] \circ [y] = \{x \circ y | x \in [x], y \in [y]\}. \tag{2.4}$$

Dans le tableau 2.1 sont définis les opérateurs de l'arithmétique des intervalles.

Opérations arithmétiques	**Calcul des bornes de l'intervalle obtenu**
Addition	$[x] + [y] = [x^- + y^-, x^+ + y^+]$
Soustraction	$[x] - [y] = [x^- - y^+, x^+ - y^-]$
Multiplication	$[x] \times [y] = [min(x^-y^-, x^-y^+, x^+y^-, x^+y^+), max(x^-y^-, x^-y^+, x^+y^-, x^+y^+)]$
Division	$[x]/[y] = [x] \times [1/y^+, 1/y^-], 0 \notin [y]$

TABLE 2.1: Opérations arithmétiques sur les variables intervalles

L'ensemble des intervalles sera noté $\mathbb{IR}$. Soit $[x] \in \mathbb{IR}$, on définit alors :

- sa borne inférieure : $inf([x]) = x^-$
- sa borne supérieure : $sup([x]) = x^+$
- sa largeur : $w([x]) = x^+ - x^- \geq 0$
- son centre : $mid([x]) = (x^+ + x^-)/2$
- son rayon : $rad([x]) = \dfrac{x^+ - x^-}{2} \geq 0$

2.1.2.3 Vecteurs et matrices d'intervalles

La notion d'intervalle peut être étendue au cas d'un vecteur $\boldsymbol{x}$ constitué de n variables bornées réelles $x_i \in \mathbb{R}, i \in \{1, \ldots, n\}$. Le vecteur intervalle $[\boldsymbol{x}]$ contenant $\boldsymbol{x}$ se définit comme suit :

$$[\boldsymbol{x}] = [[x_1], ..., [x_n]]^T \tag{2.5}$$

où chaque intervalle $[x_i] = [x_i^-, x_i^+]$ est associé à une variable réelle x_i. Le vecteur intervalle $[\boldsymbol{x}]$ définit géométriquement un orthotope aligné avec les axes du repère $(x_1, \ldots, x_n)$, plus familièrement appelé boîte ou pavé, comme le montre la figure suivante.

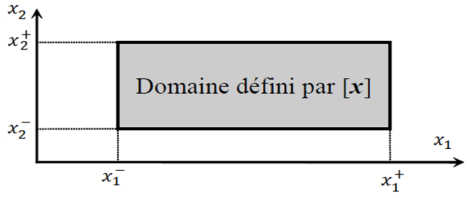

FIGURE 2.2: Vecteur intervalle de dimension 2

Soit $[\boldsymbol{x}] \in \mathbb{IR}^n$, on définit alors :
- ses bornes inférieures : $inf([\boldsymbol{x}]) = [x_1^-, x_2^-, ..., x_n^-]^T$
- ses bornes supérieures : $sup([\boldsymbol{x}]) = [x_1^+, x_2^+, ..., x_n^+]^T$
- sa largeur : $w([\boldsymbol{x}]) = \max_{j=1}^{n}(x_j^+ - x_j^-) \geq 0$
- son centre : $mid([\boldsymbol{x}]) = [mid([x_1]), mid([x_2]), ..., mid([x_n])]^T$

De même, nous notons par $[\boldsymbol{A}] \in \mathbb{IR}^{n \times m}$ une matrice intervalle de n lignes et de m colonnes dont les éléments sont des intervalles $[a_{i,j}], i \in \{1, ..., n\}$ et $j \in \{1, ..., m\}$.
Ainsi le calcul sur les vecteurs et matrices intervalles est le même que dans le cas de nombres réels, à l'exception que les opérateurs arithmétiques sur les réels seront remplacés par leurs correspondants intervalles.
Par exemple, soit $[\boldsymbol{A}] \in \mathbb{IR}^{n \times m}$ une matrice d'éléments $[a_{i,j}]$ et $[\boldsymbol{b}] \in \mathbb{IR}^n$ un vecteur d'éléments $[b_k]$, alors les éléments de $[\boldsymbol{c}] = [\boldsymbol{A}] \times [\boldsymbol{b}]$ sont donnés par : $[c_i] = \sum_{k=1}^{n} [a_{i,k}][b_k]$.

Exemple 6. [98] Soient la matrice intervalle $[\mathbf{A}]$ de dimension 2×2 et le vecteur intervalle $[\mathbf{b}]$ de dimension 2×1

$$[\mathbf{A}] = \begin{bmatrix} [0.5, 1] & [-1, 0] \\ [2, 2.5] & [0, 1] \end{bmatrix}, [\boldsymbol{b}] = \begin{bmatrix} [0, 0.5] \\ [1, 2] \end{bmatrix}$$

alors le produit matrice vecteur intervalles $[\mathbf{A}] \times [\mathbf{b}]$ conduit au vecteur intervalle suivant :

$$\begin{bmatrix} [0.5, 1] & [-1, 0] \\ [2, 2.5] & [0, 1] \end{bmatrix} \times \begin{bmatrix} [0, 0.5] \\ [1, 2] \end{bmatrix} = \begin{bmatrix} [0.5, 1][0, 0.5] + [-1, 0][1, 2] \\ [2, 2.5][0, 0.5] + [0, 1][1, 2] \end{bmatrix} = \begin{bmatrix} [-2, 0.5] \\ [0, 3.25] \end{bmatrix}$$

2.1.2.4 Intervalles et opérations sur les ensembles

Les intervalles peuvent être considérés comme des ensembles ou comme des paires de nombres réels ; les opérations classiques sur les ensembles telles que l'intersection, l'union,... peuvent donc être facilement interprétées par des inégalités conçues à partir de leurs bornes.
Soient $[x]$ et $[y]$ deux intervalles, pour l'inclusion par exemple, on a l'équivalence suivante :

$$[x] \subseteq [y] \Leftrightarrow x^- \geq y^- \wedge x^+ \leq y^+$$

et pour l'intersection on a la relation,

$$[x] \cap [y] = \begin{cases} \phi, \text{ si } x^+ \leq y^- \text{ou } y^+ \leq x^-, \\ [max\{x^-, y^-\}, min\{x^+, y^+\}] \text{ sinon.} \end{cases}$$

L'union de ces deux intervalles est donnée par le plus petit intervalle contenant les éléments de $[x]$ et de $[y]$:

$$[x] \cup [y] = \begin{cases} [min\{x^-, y^-\}, max\{x^+, y^+\}]. \end{cases}$$

2.1.2.5 Fonctions d'inclusion

Nous avons vu dans la section 2.1.2.2 les bases du calcul par intervalles dans le cas des opérations arithmétiques. Dans cette partie, nous allons nous intéresser à l'évaluation de fonctions vectorielles plus générales (ou de champs de vecteurs) dont les variables sont des intervalles, et nous présenterons différentes méthodes pour les construire.
Soit le champ de vecteurs sur les réels suivant :

$$\boldsymbol{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Alors une fonction d'inclusion de $\boldsymbol{f}$ notée $[\boldsymbol{f}]$ est une fonction de $\mathbb{IR}^n$ dans $\mathbb{IR}^m$ vérifiant la propriété d'inclusion suivante pour 2 intervalles $[\boldsymbol{x}]$ et $[\boldsymbol{y}]$:

$$[\boldsymbol{x}] \subset [\boldsymbol{y}] => [\boldsymbol{f}]([\boldsymbol{x}]) \subset [\boldsymbol{f}]([\boldsymbol{y}]).$$

Cette propriété est essentielle car il en découle la propriété suivante (notion de calcul garanti) :

$$\boldsymbol{f}([\boldsymbol{x}]) = \{\boldsymbol{f}(\boldsymbol{x}) | \boldsymbol{x} \in [\boldsymbol{x}]\} \subseteq [\boldsymbol{f}]([\boldsymbol{x}]). \tag{2.6}$$

La fonction d'inclusion $[\boldsymbol{f}]([\boldsymbol{x}])$ n'est pas unique et dépend de la façon avec laquelle la fonction intervalle $[\boldsymbol{f}]$ est écrite, c'est-à-dire de l'expression formelle utilisée.

En général, l'analyse par intervalles vise à utiliser des fonctions d'inclusion peu pessimistes dans le sens où $[\boldsymbol{f}]([\boldsymbol{x}])$ surestime le moins possible $\boldsymbol{f}([\boldsymbol{x}])$.
Notons qu'il peut exister différentes fonctions d'inclusion dites minimales, qui conduisent au plus petit intervalle $[\boldsymbol{f}]^*$ contenant l'ensemble $\boldsymbol{f}([\boldsymbol{x}])$. Sur la figure (2.3) nous évaluons la fonction d'inclusion minimale et non minimale d'un pavé $[\boldsymbol{x}]$ de dimension 2.

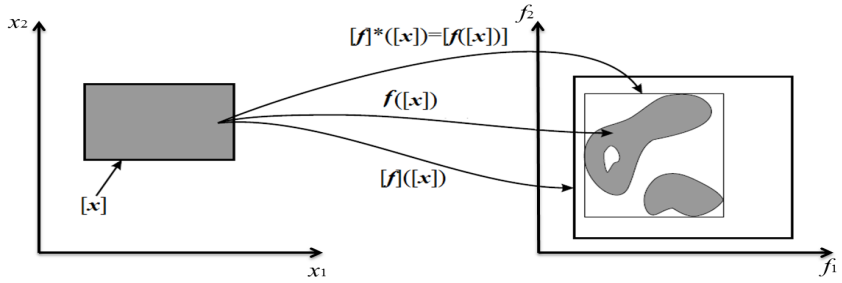

FIGURE 2.3: L'image d'un pavé par les fonctions $\boldsymbol{f}$, $[\boldsymbol{f}]$ et$[\boldsymbol{f}]^*$

Exemple 7. [99] Soit une fonction réelle $f(x) = x^2 - 2x + 1$ avec $[x] = [-1, 3]$, alors l'évaluation des 3 fonctions d'inclusions suivantes, dont les expressions sont équivalentes sur les réels, conduit à 3 intervalles différents :
$[f_1]([x]) = [x]^2 - 2[x] + 1 = [-5, 12]$
$[f_2]([x]) = [x]([x] - 2) + 1 = [-8, 4]$
$[f_3]([x]) = ([x] - 1)^2 = [0, 4]$

Les résultats obtenus montrent que selon la manière avec laquelle la fonction d'inclusion $[f]$ est écrite, l'intervalle résultat est différent. L'intervalle solution le moins pessimiste correspond à la fonction ayant une occurrence unique de la variable $[x]$, l'intervalle obtenu $[f_3]$ correspondant à la fonction d'inclusion minimale $[f]^*$.
Notons que la fonction d'inclusion utilisée pour la fonction puissance est celle spécifiée dans le tableau 2.2 :

$[x]^n$	**Conditions à vérifier**
$[x^{-n}, x^{+n}]$	si $x^- > 0$ ou si n est impair
$[x^{+n}, x^{-n}]$	si $x^+ < 0$ et si n est pair
$[0, \lvert[x]\rvert^n]$	si 0 est élément de $[x]$ et si n est pair

TABLE 2.2: Fonction d'inclusion de $[x]^n$

Fonction d'inclusion des fonctions élémentaires

Toute fonction élémentaire f $\{exp, log, ...\}$, continue et monotone, admet comme fonction d'inclusion minimale,

$$[f]^*([x]) = [min(f(x^-), f(x^+)), max(f(x^-), f(x^+))]. \tag{2.7}$$

Si f n'est pas monotone comme pour les fonctions trigonométriques *sin* et *cos*, la construction de la fonction d'inclusion minimale dépend alors d'une étude globale des variations de la fonction. Dans ce cas, la méthode employée consiste à exploiter la "monotonie par morceaux" de f.

Exemple 8. Sur son domaine de définition $D =]-\infty, +\infty[$, la fonction exponentielle, notée $exp()$ est croissante, ce qui conduit à la fonction d'inclusion minimale : :

$$\forall [x] = [x^-, x^+] \subset D, exp([x]) = [exp(x^-), exp(x^+)]$$

Exemple 9. Soit la fonction cosinus non monotone, nous procédons de la façon suivante

$$[cos]([x]) = [a, b],$$

avec

$$a = \begin{cases} -1 \quad \text{si} \quad \exists k \in \mathbb{Z} | (2k+1)\pi \in [x] \\ min(cos(x^-), cos(x^+)), \quad \text{sinon} \end{cases}$$

et

$$b = \begin{cases} 1 \quad \text{si} \quad \exists k \in \mathbb{Z} | 2k\pi \in [x] \\ max(cos(x^-), cos(x^+)), \quad \text{sinon} \end{cases}$$

La fonction puissance détaillée dans le tableau 2.2 est une autre illustration de la construction de fonctions d'inclusion.

Fonction d'inclusion naturelle

Cette méthode d'évaluation de la fonction d'inclusion est utilisée dans les cas où on a plusieurs variables. Pour la construire il suffit de remplacer chaque variable x_i par l'intervalle auquel elle appartient ($[x_i]$), chaque opérateur arithmétique par l'opérateur intervalle correspondant et les fonctions élémentaires par leur fonction d'inclusion minimale.

Exemple 10. [100] Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ dont l'expression est donnée par :

$$f(x_1, x_2) = \frac{x_2}{x_1 + x_2} + \sin(x_1)\cos(x_1) \tag{2.8}$$

avec $x_1 \in [-1, 2]$ et $x_2 \in [3, 5]$. En appliquant les principes énoncés précédemment, nous obtenons la fonction d'inclusion naturelle suivante :

$$[f_1]([x_1], [x_2]) = \frac{[x_2]}{[x_1] + [x_2]} + [\sin]([x_1]) \times [\cos]([x_1]) \tag{2.9}$$

L'évaluation de cette fonction par intervalle conduit à : $[f_1]([-1,2],[3,5]) = [-0.42, 3.5]$. On peut trouver une autre fonction d'inclusion naturelle $[f_2]$ pour f, en modifiant l'expression (2.8). Nous avons alors :

$$[f_2]([x_1],[x_2]) = \frac{1}{1+[x_1]/[x_2]} + [\sin]([x_1]) \times [\cos]([x_1]) \tag{2.10}$$

Son évaluation donne l'intervalle suivant : $[f_2]([-1,2],[3,5]) = [-0.2415, 2.5]$. Finalement une troisième fonction d'inclusion $[f_3]$ pour f donnée par :

$$[f_3]([x_1],[x_2]) = \frac{1}{1+[x_1]/[x_2]} + \frac{[\sin](2 \times [x_1])}{2}, \tag{2.11}$$

conduit à $[f_3]([-1,2],[3,5]) = [0.1, 2]$.

2.1.2.6 Pessimisme

La contre-partie de la garantie numérique apportée par le calcul par intervalles est le problème de pessimisme, autrement dit de sur-estimation de l'intervalle calculé en sortie. En effet, le résultat d'une suite d'opérations entre deux ou plusieurs intervalles n'est pas forcément minimal ; l'intervalle obtenu est donc pessimiste. Ce problème a pour origine 2 phénomènes essentiellement : la dépendance et l'enveloppement.

Dépendance

Le problème de dépendance, qui est à l'origine de l'inclusion formulée dans la relation (2.6), est lié à la multi-occurrence des variables incertaines dans l'expression de la fonction réelle f. Lorsque l'on construit une fonction d'inclusion, par exemple naturelle, chaque variable bornée se retrouve remplacée par son support intervalle, c'est-à-dire un ensemble, certes identique en ce qui concerne ses bornes, mais donc chaque occurrence est indépendante.
Prenons l'exemple d'une fonction $f(x) = x \circ x$, x élément de $[x^-, x^+]$ et où $\circ$ peut être une opération arithmétique, une fonction élémentaire ou une composition des deux. L'évaluation de la fonction d'inclusion naturelle conduit à calculer $[f]([x]) = [x] \circ [x]$ dont les bornes sont obtenues à l'aide de règles de calculs portant sur x^- et x^+. Le calcul réellement effectué correspond à l'ensemble $\{x \circ y / x \in [x^-, x^+] \wedge y \in [x^-, x^+]\}$ et non $\{[x] \circ [x] / x \in [x^-, x^+]\}$. On combine n'importe quelle valeur de $[x^-, x^+]$ avec n'importe quelle autre valeur de cet intervalle car le fait de travailler sur des ensembles ne permet plus de tenir compte de la dépendance entre les deux occurrences de la variable bornée x. C'est cette non prise en compte de la dépendance entre variables qui génère le pessimisme.

On reprend l'exemple 10, où $[f_1], [f_2], [f_3]$ représentent 3 fonctions d'inclusion différentes de la même fonction f. Nous remarquons que la taille des intervalles obtenus par ces 3 fonctions d'inclusion dépend de l'expression utilisée pour l'écriture de f. Ceci est dû au phénomène de dépendance expliqué précédemment et qui dépend du nombre d'occurrences des variables intervalles. Dans cet exemple, la fonction d'inclusion naturelle $[f_3]$ est la plus précise, même si celle-ci n'est pas minimale. Concernant $[f_1]$, il y a 3 occurrences de x_1 et 2 occurrences de x_2. Pour $[f_2]$, on a 3 occurrences de x_1 et une occurrence de x_2. Dans le cas de $[f_3]$, nous avons 2 occurrences de x_1 et une occurrence de x_2. On peut constater que plus l'occurrence des variables est importante, plus le pessimisme lors des évaluations sera important.

Enveloppement

Un inconvénient majeur de l'utilisation de l'analyse par intervalles réside dans l'effet d'enveloppement dû à la représentation d'un ensemble quelconque par un pavé.

Exemple 11. [101] : On considère des rotations successives d'un pavé $[\boldsymbol{x}] = [x_1] \times [x_2]$ à l'aide de la transformation définie par la matrice

$$\boldsymbol{A} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Une rotation du pavé $[\boldsymbol{x}]$ donne lieu à un **rectangle** de même taille, tracé sur la figure 2.4.

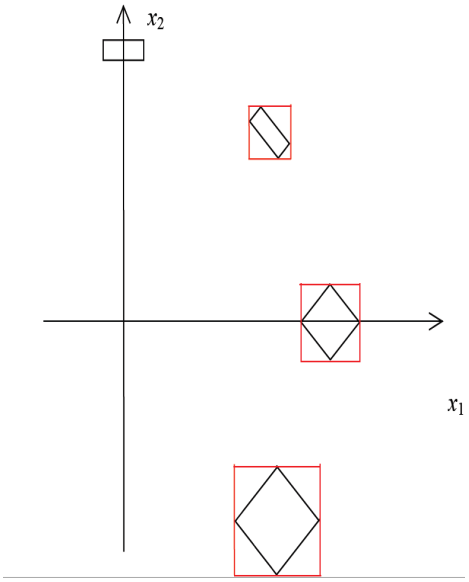

FIGURE 2.4: Effet d'enveloppement pour $\theta = -\frac{\pi}{4}$

La représentation de ce rectangle par un pavé se fait en l'enveloppant par un autre rectangle dont les cotés sont parallèles aux axes du repère, ce qui inclut des valeurs supplémentaires qui ne sont pas solution du problèmes posé. La rotation suivante de ce pavé dont la taille est plus grande que celle du pavé initial donne lieu à un autre rectangle de même taille. Sa représentation par un pavé englobant aligné avec les axes du repère augmente une nouvelle fois la taille de l'ensemble calculé. Et ainsi de suite, la rotation successive d'un pavé génère une suite de pavés de tailles croissantes alors que la rotation est une opération conservatrice. Le phénomène d'enveloppement est particulièrement préjudiciable dans le cas de modèles exprimés sous formes de récurrence, où chaque étape consiste à calculer l'image de l'ensemble précédemment évalué.

Méthodes à base de division uniforme

Soit la fonction $f : \mathbb{R}^n \to \mathbb{R}$. L'évaluation d'une fonction d'inclusion $[f]$ est d'autant plus performante que la largeur de l'argument intervalle $[\boldsymbol{x}]$ est petite :

$w([\boldsymbol{x}]) \to 0 \Rightarrow w([f]([\boldsymbol{x}])) \to 0.$

Pour améliorer la précision, on peut réaliser une sous-division uniforme des supports des variables multi-occurrentes, on évalue la fonction d'inclusion sur chaque partition, puis on en calcule l'union (figure 2.5).

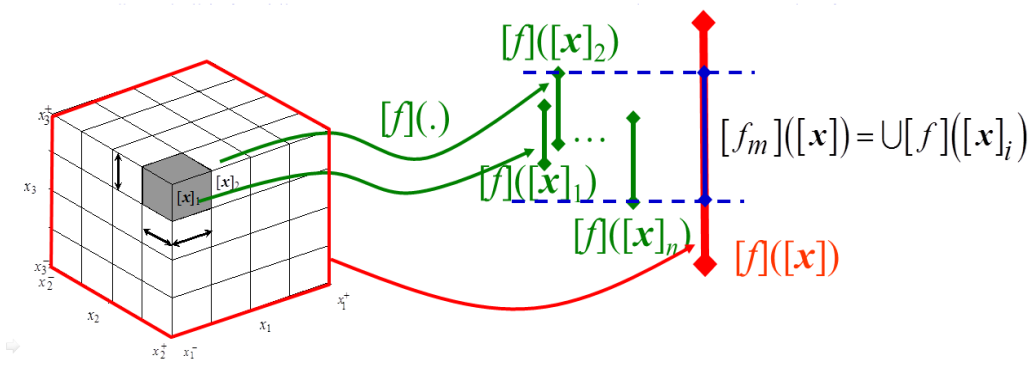

FIGURE 2.5: Sous-division uniforme d'ordre $m = 4$

L'idée est qu'en travaillant sur des intervalles plus petits, le pessimisme lors de l'évaluation de la fonction d'inclusion, sur chaque sous-intervalle $[\boldsymbol{x}]_i$ sera moins important. Il en sera donc de même du résultat final donné par $\bigcup[f]([\boldsymbol{x}]_i)$.

L'inconvénient est que cette approche est coûteuse en temps de calcul. L'idée consiste à coupler cette approche à des fonctions d'inclusion spécifiques, dont on sait qu'elles ont de bonnes propriétés en termes de convergence (lorsque la taille de $[\boldsymbol{x}]$ diminue) vers l'intervalle minimum. En clair, les fonctions d'inclusion qui vont maintenant être détaillées vont permettre de réduire l'ordre de la sous-division uniforme à retenir.

Fonction d'inclusion centrée

Ce type de fonction permet dans certains cas de réduire le pessimisme dû au phénomène de dépendance [102, 103]. L'utilisation de pavés de faibles largeurs augmente l'efficacité de cette méthode. La forme centrée sera aussi utilisée dans le cadre de fonctions à variables multi-occurrentes et dont il est difficile de réduire les occurrences par de simples manipulations algébriques.
Elle s'applique aux fonctions $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ différentiables sur un pavé $[\boldsymbol{x}] \in \mathbb{IR}^n$ et son principe est basé sur le théorème de la valeur moyenne, où on note par $\hat{\boldsymbol{x}}$=$mid([\boldsymbol{x}])$ le centre du pavé $[\boldsymbol{x}]$.

$$\forall \boldsymbol{x} \in [\boldsymbol{x}], \exists \boldsymbol{y} \in [\boldsymbol{x}] | \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\hat{\boldsymbol{x}}) + \boldsymbol{J}(\boldsymbol{y})(\boldsymbol{x} - \hat{\boldsymbol{x}}) \tag{2.12}$$

où $\boldsymbol{J}$ signifie le Jacobien de la fonction $\boldsymbol{f}$ de dimension $n \times m$. Ainsi on peut affirmer que :

$$\forall \boldsymbol{x} \in [\boldsymbol{x}], \boldsymbol{f}(\boldsymbol{x}) \in \boldsymbol{f}(\hat{\boldsymbol{x}}) + [\boldsymbol{J}]([\boldsymbol{x}])([\boldsymbol{x}] - \hat{\boldsymbol{x}}) \tag{2.13}$$

où $[\boldsymbol{J}]$ est la fonction d'inclusion de $\boldsymbol{J}$ et on obtient alors :

$$\boldsymbol{f}([\boldsymbol{x}]) \subseteq \boldsymbol{f}(\hat{\boldsymbol{x}}) + [\boldsymbol{J}]([\boldsymbol{x}])([\boldsymbol{x}] - \hat{\boldsymbol{x}}) \tag{2.14}$$

Donc, la fonction intervalle suivante :

$$[\boldsymbol{f}_c]([\boldsymbol{x}]) \equiv \boldsymbol{f}(\hat{\boldsymbol{x}}) + [\boldsymbol{J}]([\boldsymbol{x}])([\boldsymbol{x}] - \hat{\boldsymbol{x}}) \tag{2.15}$$

est une fonction d'inclusion de $\boldsymbol{f}$.

Fonction d'inclusion de Taylor

Les fonctions d'inclusion de Taylor sont obtenues en employant un ordre de dérivation supérieure à 1 (Neumaier 2003 [104]). L'expression ci-dessous exprime la fonction d'inclusion de Taylor du second ordre :

$$[\boldsymbol{f}_T]([\boldsymbol{x}]) = \boldsymbol{f}(\hat{\boldsymbol{x}}) + \boldsymbol{J}([\hat{\boldsymbol{x}}])([\boldsymbol{x}] - \hat{\boldsymbol{x}}) + \frac{1}{2}([\boldsymbol{x}] - \hat{\boldsymbol{x}})^T [\boldsymbol{H}]([\boldsymbol{x}])([\boldsymbol{x}] - \hat{\boldsymbol{x}}) \tag{2.16}$$

On désigne par $\boldsymbol{H}$ la fonction d'inclusion du Hessien de $\boldsymbol{f}$.
Les fonctions d'inclusion de Taylor peuvent réduire considérablement le pessimisme pour des intervalles de petite taille.

Fonction d'inclusion à test de monotonie

Soit la fonction $f : \mathbb{R}^n \to \mathbb{R}$. La fonction d'inclusion à test de monotonie est une forme à valeur moyenne où on exploite la monotonie de f en testant le signe des intervalles $[df_{xi}]$, où $[df_{xi}]$ est l'extension intervalle naturelle de la dérivée de f par rapport à la variable x_i.

$$[f_{mt}]([x_1], ..., [x_n]) = [f(u_1, ..., u_n), f(v_1, ..., v_n)] + \sum_{i \in S} [df_{xi}]([x_1], ..., [x_n])([x_i] - \hat{x}) \quad (2.17)$$

où la paire (u_i, v_i) est définie par :

$$[u_i, v_i] = \begin{cases} [x_i^-, x_i^+] & \text{si} \quad df_{x_i}^-([x_1], ..., [x_n]) \geq 0 \\ [x_i^+, x_i^-] & \text{si} \quad df_{x_i}^-([x_1], ..., [x_n]) < 0 \quad \text{et} \quad df_{x_i}^+([x_1], ..., [x_n]) \leq 0 \\ [\hat{x}, \hat{x}] & \text{sinon} \end{cases} \quad (2.18)$$

avec S l'ensemble des indices $i \in \{1, ..., n\}$ correspondant aux intervalles $[df_{xi}]$ contenant 0 :

$$df_{x_i}^-([x_1], ..., [x_n]) < 0 < df_{x_i}^+([x_1], ..., [x_n]) \quad (2.19)$$

Dans le cas où la fonction f est croissante par rapport à la variable x_i, ce qui est numériquement garanti pour toutes les valeurs des entrées si $df_{x_i}^-([x_1], ..., [x_n]) \geq 0$, alors la borne inférieure de f s'évalue pour la borne inférieure de l'intervalle $[x_i]$. De même, la borne supérieure de f s'évalue pour la borne supérieure de l'intervalle $[x_i]$. Le raisonnement est identique si $df_{x_i}^-([x_1], ..., [x_n]) < 0$ et $df_{x_i}^+([x_1], ..., [x_n]) \leq 0$, c'est-à-dire lorsque f est monotone et décroissante par rapport à x_i, et dans ce cas les bornes inférieure et supérieure de f s'évaluent en inversant les bornes de $[x_i]$ par rapport au cas précédent. Pour ces deux cas, l'évaluation de la fonction d'inclusion $[f_{mt}]$ n'utilise pas le calcul classique par intervalles, mais une double évaluation, sans pessimisme, de la fonction réelle f. Dans le dernier cas, la monotonie de la fonction f par rapport à x_i ne peut être prouvée, et c'est une évaluation de type fonction d'inclusion centrée qui est réalisée.

Exemple comparatif [98]

Considérons la fonction f définie par :

$$f(x) = x^2 + \sin(x),$$

où la variable $x \in [x] = [\frac{2\pi}{3}, \frac{4\pi}{3}]$.

Les fonctions d'inclusion naturelle, centrée et de Taylor d'ordre deux associées à f sont données respectivement par les relations suivantes :

$[f_n]([x]) = [x]^2 + \sin([x]) = [3.52046, 18.41199]$,
$[f_c]([x]) = f(\pi) + ([x] - \pi)[f']([x]) = [1.62022, 18.11899]$,
$[f_T]([x]) = f(\pi) + ([x] - \pi)[f']([x]) + \frac{([x]-\pi)^2}{2}[f'']([x]) = [4.33706, 16.97362]$.

Par ailleurs, pour cet exemple, comme f est croissante sur l'intervalle $[x]$, il est possible de déterminer sa fonction d'inclusion minimale $[f]^*$:

$[f]^*([x]) = [(x^-)^2 + \sin(x^-), (x^+)^2 + \sin(x^+)] = [5.25251, 16.67994]$.

En se basant sur les résultats obtenus pour ces quatre évaluations, on conclut que la fonction d'inclusion la plus précise parmi f_n, f_c, f_T est la fonction d'inclusion de Taylor pour cet exemple. En effet, celle-ci est la plus proche de la fonction d'inclusion minimale.

2.1.2.7 Implémentation de la méthode de propagation par intervalles

La figure 2.6 illustre la phase de calcul de la propagation d'incertitudes en utilisant l'analyse par intervalles.

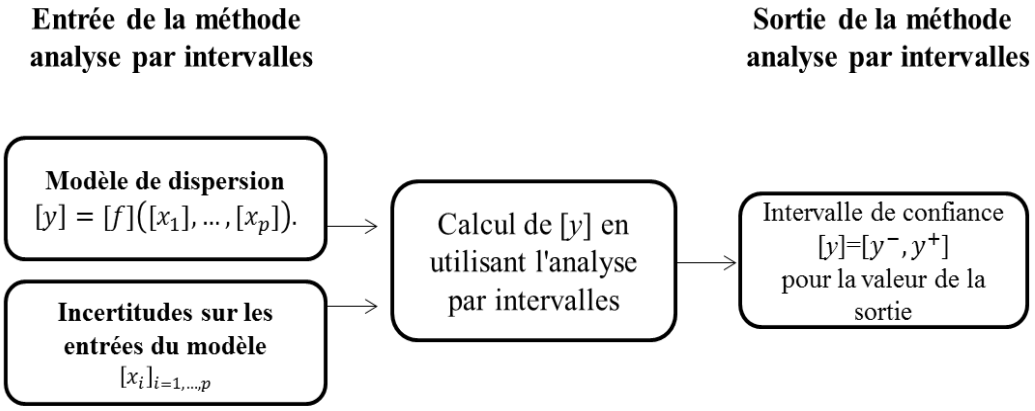

FIGURE 2.6: Phase de calcul de la propagation d'incertitudes

À partir de l'intervalle de confiance calculé par la méthode de l'analyse par intervalles (AI), l'indicateur de la propagation de l'incertitude u_{-AI} est défini, selon le même principe que pour celui de l'approche probabiliste (2.2), par :

$$u_{-AI} = \frac{y^+ - y^-}{2} \times \frac{100}{y_{ValeurNominale}}. \tag{2.20}$$

2.1.2.8 Avantages et inconvénients de l'analyse par intervalles

La méthode de l'analyse par intervalles possède deux avantages principaux :

1. La prise en compte de tous les cas possibles ce qui garantit les résultats.
2. Sa capacité à calculer en un coup l'intervalle de sortie conduisant à des temps de calcul réduits par rapport à une approche de type Monte Carlo.

Cette méthode possède deux inconvénients principaux. Le premier réside dans la nécessité d'avoir une forme analytique du modèle pour en concevoir une fonction d'inclusion, mais il est possible de contourner cet inconvénient avec les modèles par code informatique sous certaines conditions, comme on l'expliquera plus tard dans le chapitre 4 (en étudiant la monotonie du modèle par rapport aux différentes entrées, il est possible d'évaluer les bornes supérieures et inférieures de la sortie sans faire de calcul par intervalles dans le même esprit que la fonction d'inclusion à test de monotonie). Le deuxième inconvénient réside dans la surestimation des résultats en raison du pessimisme.

2.2 Application à la dispersion atmosphérique

2.2.1 Modèle d'effet Gaussien

Afin d'évaluer la gravité du risque lorsqu'un événement indésirable et inattendu se produit, un modèle mathématique et/ou physique peut être utilisé pour calculer l'intensité des effets engendrés par le phénomène dangereux.
Dans cette étude, un modèle d'effets est utilisé pour estimer ou prévoir la concentration de gaz émis sous le vent à partir de sources comme des installations industrielles, les véhicules TMD ou les rejets accidentels de produits chimiques.
Nous considérons le modèle gaussien de rejet continu proposé par Pasquill [105] et déjà présenté dans la section 1.4.1.1. Pour rappel, ce modèle représente les relations entre les entrées du modèle de dispersion atmosphérique (vitesse du vent, conditions de point d'émission, débit de fuite,...) et la concentration c_k de gaz dans l'air en un point donné de coordonnées (x_k, y_k, z_k) [106].

$$\begin{aligned} c_k &= f(x_k, y_k, z_k, u_{ref}, z_{ref}, h, q, a_y, a_z, b_y, b_z, c_y, c_z) \\ &= \frac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} (a_y x_k^{by} + c_y)(a_z x_k^{bz} + c_z)} \times exp\left[-\frac{1}{2}\left(\frac{y_k}{a_y x_k^{by} + c_y}\right)^2\right] \times \\ &\left\{exp\left(-\frac{1}{2}\left(\frac{z_k - h}{a_z x_k^{bz} + c_z}\right)^2\right) + \alpha exp\left(-\frac{1}{2}\left(\frac{z_k + h}{a_z x_k^{bz} + c_z}\right)^2\right)\right\} \end{aligned} \tag{2.21}$$

- Les termes $a_y x_k^{by} + c_y$ et $a_z x_k^{bz} + c_z$ représentent les paramètres de dispersion et dépendent de la distance x_k. Ils représentent respectivement les écarts-types d'un panache suivant une distribution normale dans les dimensions latérales et verticales. Les valeurs de a_y, a_z, b_y, b_z, c_y et c_z peuvent être déterminées pour chaque classe de stabilité atmosphérique définie par Pasquill, en utilisant les tableaux 1.2 et 1.3 de la section 1.4.1.1 [106, 107],
- La variable d'entrée q désigne le débit de fuite,
- La variable d'entrée u_{ref} désigne la vitesse du vent mesurée à une altitude z_{ref},
- La fuite a lieu au point de coordonnées $(0, 0, h)$,
- Le terme α désigne le coefficient de réflexion au sol.

2.2.2 Modélisation des entrées incertaines du modèle

Au lieu de représenter une entrée mal connue par une valeur nominale constante, celle-ci peut être définie comme une variable bornée. En d'autres termes, sa valeur réelle est inconnue, mais elle appartient à un ensemble de valeurs possibles définies comme un intervalle dont les bornes sont elles connues.
Dans la suite, une imprécision ρ_v signifie qu'une variable positive incertaine v est représentée par le support intervalle centré sur v et de demie largeur ρ_v :

$$[v(1 - \rho_v), v(1 + \rho_v)]. \tag{2.22}$$

Cette étude a été appliquée à un exemple d'accident impliquant de l'oxyde nitrique (ou monoxyde d'azote) sous forme gazeuse. Ce gaz est toxique et a une densité relative par rapport à l'air de 1.04, de sorte qu'un modèle gaussien est bien adapté pour modéliser la dispersion d'un tel gaz. Il peut être transporté, par exemple sous forme comprimée et les informations de classification dans le cadre du transport sont donnés par : Numéro ONU : 1660 , Code ADR Classe 2, Code 1 TOC.

Pour une classe de stabilité choisie C, les valeurs nominales des paramètres de dispersion sont les suivantes : $a_y = 0.105, a_z = 0.066, b_y = b_z = 0.915$, et $c_y = c_z = 0$ [106]. Nous supposons que la hauteur du point de fuite est $h = 2m$. La vitesse du vent mesurée est $u_{ref} = 4.58m/s$, à une hauteur de $z_{ref} = 40m$. La valeur nominale théorique du débit de fuite est $q = 2216g/s$. On suppose une réflexion totale du panache ce qui conduit à $\alpha = 1$.

Dans la suite, les imprécisions sur les entrées suivantes sont considérées :
- le débit de fuite q avec une incertitude de $\rho_v = 5\%$,
- la vitesse du vent u_{ref} avec $\rho_v = 2.5\%$,
- les paramètres de dispersion a_y, a_z avec $\rho_v = 2.5\%$ et b_y, b_z avec $\rho_v = 1\%$,

ce qui permet pour chaque entrée de définir un support intervalle de valeurs possibles donné par l'équation (2.22).

L'estimation des niveaux d'incertitudes ρ_v est obtenue grâce à des éléments issus de la bibliographie, des résultats de mesures antérieures, l'expérience ou la connaissance du comportement et des propriétés des matériaux et instruments utilisés,...

Ces intervalles sont directement utilisés pour propager les incertitudes sur la sortie à l'aide de la méthode d'analyse par intervalles. Pour comparer ce résultat avec celui de la simulation de Monte Carlo, nous avons besoin d'une part de générer des valeurs aléatoires pour chaque entrée du modèle contenues dans le même support borné selon une loi de distribution qui est choisie ici uniforme par raison de simplicité. D'autre part, il faut disposer d'un point de comparaison pour évaluer les résultats obtenus par les deux approches probabiliste et ensembliste. C'est là qu'interviennent les indicateurs (2.2) et (2.20) que nous rappelons :

- $u_{-MC} = \dfrac{y_{Max} - y_{Min}}{2} \times \dfrac{100}{y_{ValeurNominale}}$
- $u_{-AI} = \dfrac{y^+ - y^-}{2} \times \dfrac{100}{y_{ValeurNominale}}$.

2.2.3 Propagation des incertitudes

Cas d'étude

L'objectif est de comparer les intervalles de confiance obtenus par la méthode de Monte Carlo (MC) et l'Analyse par Intervalles (IA) en termes de précision et de temps du calcul. Dans notre étude, la concentration de gaz est estimée en 5 points placés sous le vent de la source d'émission de gaz d'oxyde nitrique. Ces points ont été judicieusement choisis de manière à être représentatifs du panache de gaz tout en appartenant à la zone de danger que l'on cherchera à établir dans les chapitres suivants. Pour ces 5 points, les supports intervalles des concentrations en gaz calculés, ainsi que les indicateurs représentatifs de l'amplitude des incertitudes sur ces concentrations sont donnés pour les deux approches. Le nombre d'échantillons pour la méthode de Monte Carlo est ajusté en l'augmentant jusqu'à ce qu'aucun changement significatif dans les bornes supérieures y_{Max} et inférieures y_{Min} ne soit observé. Cela conduit à retenir $N = 100.000$ échantillons, ce qui est un choix raisonnable et classique compte tenu du nombre $p = 6$ d'entrées incertaines du modèle.

Le tableau 2.3 récapitule les valeurs nominales des entrées du modèle utilisées avec les incertitudes ajoutées.

$q \pm 5\%$	$a_y \pm 2.5\%$	$a_z \pm 2.5\%$	$b_y \pm 1\%$	$b_z \pm 1\%$	$u_{ref} \pm 2.5\%$	x	y
2216	0.105	0.066	0.915	0.915	4.58	40	5
						350	8
						460	35
						600	50
						800	45

TABLE 2.3: Entrées du modèle

Dans le modèle de dispersion gaussien existent des entrées multi-occurrentes. C'est pour cette raison que nous présentons la fonction d'inclusion naturelle utilisée dans notre cas d'étude et pour simplifier le codage du modèle et sa présentation, nous avons découpé celle-ci en plusieurs morceaux :

$$u_k = \left(u_{ref} \times \left(\frac{h}{z_{ref}} \right)^{\frac{1}{3}} \right) \tag{2.23}$$

$$A_1 = exp\left(\frac{-0.5 \times (z_k - h)^2}{(a_z \times x_k^{bz})^2} \right) \tag{2.24}$$

$$A_2 = exp\left(\frac{-0.5 \times (z_k + h)^2}{(a_z \times x_k^{bz})^2} \right) \tag{2.25}$$

$$A = A_1 + A_2 \tag{2.26}$$

$$B = exp\left(\frac{-0.5 \times y_k^2}{(a_y \times x_k^{by})^2} \right) \tag{2.27}$$

$$C = \frac{q}{2 \times \pi \times (u_k \times ((a_y \times a_z) \times x_k^{by+b_z}))} \tag{2.28}$$

$$c_k = C \times (B \times A) \tag{2.29}$$

Le tableau 2.4 donne, pour les 5 points considérés, les concentrations en gaz $c_{Nom,k}$ calculées à partir des valeurs nominales des entrées étudiées, les intervalles de confiances estimés, ainsi que les indicateurs $U - MC$ calculés par Monte Carlo et $U - AI$ calculés par l'analyse par intervalles en utilisant une fonction d'inclusion naturelle.

№	$Point(x_k, y_k, z_k)$	$c_{Nom,k}$	$[y_{Min}, y_{Max}]_{MC}$	$[y^-, y^+]_{AI}$	$U - MC$	$U - AI$
1	(40, 5, 2)	10.46	[9.25,11.76]	[8.37,12.99]	11.99%	22.10%
2	(350, 8, 2)	1.22	[1.09,1.37]	[1.06,1.41]	11.47%	14.34%
3	(460, 35, 2)	0.379	[0.34,0.42]	[0.32,0.45]	10.81%	18.33%
4	(600, 50, 2)	0.193	[0.17,0.22]	[0.16,0.23]	11.39%	19.43%
5	(800, 45, 2)	0.187	[0.17,0.21]	[0.16,0.22]	10.42%	16.57%

TABLE 2.4: Les intervalles de confiance et les indicateurs calculés

La figure 2.7 représente la propagation des incertitudes avec la méthode de Monte Carlo et la méthode Analyse par Intervalles pour les 5 points de coordonnées (x_k, y_k, z_k).

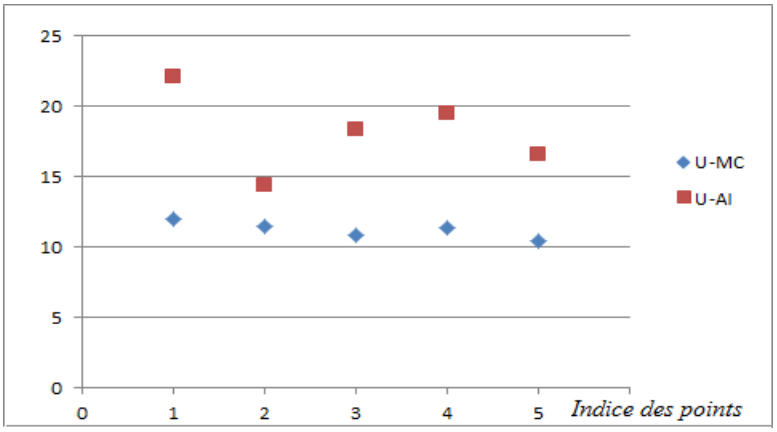

FIGURE 2.7: La propagation de l'incertitude selon les points étudiés

Interprétation des résultats

Avec la simulation de Monte Carlo, les résultats obtenus montrent que l'incertitude sur la sortie de modèle varie entre 10.42 % et 11.99 % selon les différents points étudiés. Avec la méthode d'analyse par intervalles, l'indicateur varie entre 14.34 % et 22.10%.
En ce qui concerne le temps d'exécution avec la méthode AI, le script de calcul a besoin de $3.5ms$ comme temps d'exécution pour obtenir la concentration en un point donné. Avec la méthode MC le temps d'exécution est de $85ms$, on note donc une réduction de 95.8% du temps de calcul pour l'approche AI par rapport à l'approche MC pour le nombre d'échantillon N choisi. Ces résultats montrent également que l'AI fournit des intervalles de confiance plus grands par rapport à MC lorsque certains paramètres du modèle sont incertains. Plusieurs raisons expliquent la différence entre les deux approches.
La première est due à la capacité de la méthode de l'AI à prendre en compte toutes les combinaisons possibles des entrées incertaines et à garantir que toutes les valeurs possibles de la concentration se retrouvent bien dans les intervalles de confiance obtenus. Cependant à cause du problème de pessimisme induit par de multiples occurrences de certaines entrées incertaines du modèle telles que a_y, a_z, b_y et b_z, cette approche conduit en réalité à une surestimation des intervalles de confiance recherchés. La deuxième raison est que la simulation Monte Carlo a besoin de générer aléatoirement chaque entrée du modèle sur son support borné. Parce que le nombre d'échantillons N est fini, la méthode de Monte Carlo est incapable de prendre en compte en les tirant toutes les valeurs possibles des entrées, par ailleurs elle ne peut pas non plus appréhender toutes les combinaisons possibles des valeurs des entrées. En d'autres termes, AI et MC conduisent respectivement à des approximations extérieures et intérieures des intervalles de confiance exacts de la concentration en gaz.

En ce qui concerne le temps d'exécution, la principale raison provient du grand nombre N d'échantillons utilisés par MC. Notons que l'on pourrait s'attendre à un plus grand écart entre les temps de calculs mesurés. Il faut garder à l'esprit que pour une seule simulation du modèle, la méthode probabiliste nécessite moins de calculs puisque l'analyse par intervalles utilisent des règles de calculs, plus ou moins complexes sur les bornes à manipuler en fonctions des opérations à effectuer. De plus, afin d'optimiser les temps de calculs pour MC, les N échantillons sont tirés avant d'être transmis au modèle, qui, capable de travailler matriciellement, va renvoyer les N sorties associées en une seule simulation. Il n'y a donc qu'un seul appel au modèle et il est clair que si nous avions procédé en appelant autant de fois le modèle qu'il y a d'échantillons, le temps de calcul par MC aurait été encore plus important.

Nous présentons dans cette partie la fonction d'inclusion à test de monotonie pour le modèle gaussien. L'objectif est de comparer les résultats par rapport à la fonction d'inclusion naturelle pour les 5 points d'étude. Pour calculer la dérivée partielle de la fonction d'inclusion par rapport aux entrées incertaines, nous avons besoin de calculer les termes suivants en utilisant les equations 2.23 à 2.29, soit :

$$C1 = \frac{1}{2 \times \pi \times (u_k \times ((a_y \times a_z) \times x_k^{by+b_z}))} \tag{2.30}$$

$$C2 = \frac{-q}{2 \times \pi \times \left(\frac{h}{z_{ref}}\right)^{\frac{1}{3}} \times ((a_y \times a_z) \times x_k^{by+b_z})} \tag{2.31}$$

$$D_5 = \frac{y_k^2}{(x_k^{2b_y} \times a_y^2)} \tag{2.32}$$

$$D_1 = -1 + D_5 \tag{2.33}$$

$$D_4 = \frac{(A_1 \times (z_k - h)^2) + (A_2 \times (z_k + h)^2)}{(x_k^{2b_z}) \times a_z^2} \tag{2.34}$$

$$D_2 = -A + D_4 \tag{2.35}$$

d'où :

- $df_q = C_1 \times (A \times B)$
- $df_{u_{ref}} = C_2 \times (A \times B)$
- $df_{a_y} = \frac{D_1}{a_y} \times (C \times (B \times A))$
- $df_{a_z} = \frac{D_2}{a_z} \times (C \times B)$
- $df_{b_y} = ((D_1 \times A) \times (B \times C)) \times log(x_k)$
- $df_{b_z} = (D_2 \times (B \times C)) \times log(x_k)$

Le tableau 2.5 donne les indicateurs sur les 5 points étudiés, $U - AI1$ calculé à l'aide de la fonction d'inclusion naturelle et $U - AI2$ calculé par la fonction d'inclusion à test de monotonie.

№	$Point(x_k, y_k, z_k)$	$U - AI1$	$U - AI2$
1	(40, 5, 2)	22.10%	13.76%
2	(350, 8, 2)	14.34%	13.15%
3	(460, 35, 2)	18.33%	12.00%
4	(600, 50, 2)	19.43%	13.29%
5	(800, 45, 2)	16.57%	10.91%

TABLE 2.5: Les indicateurs calculés

La figure 2.8 illustre la propagation des incertitudes avec la méthode de Monte Carlo et la méthode de l'Analyse par Intervalles selon les fonctions d'inclusion naturelle et à test de monotonie pour les 5 points de coordonnées (x_k, y_k, z_k).

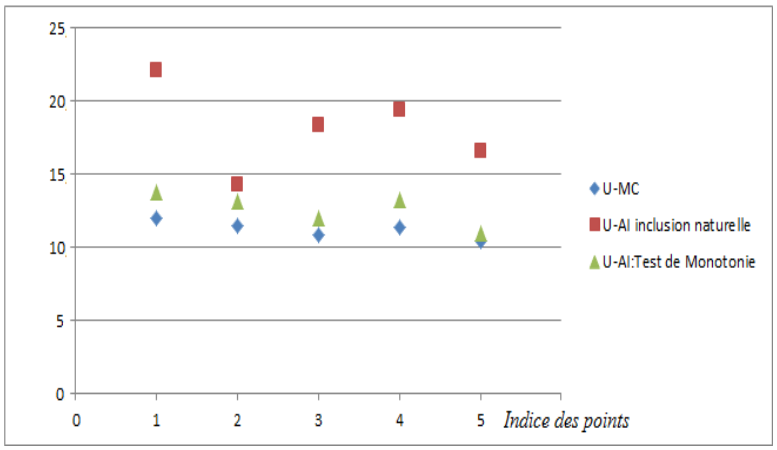

FIGURE 2.8: La propagation de l'incertitude selon les points étudiés

Interprétation des résultats

Avec la simulation de Monte Carlo, les résultats obtenus montrent que l'incertitude sur la sortie du modèle varie entre 10.42 % et 11.99 % selon les différents points étudiés. Avec la méthode de l'analyse par intervalles utilisant la fonction d'inclusion naturelle, l'indicateur varie entre 14.34 % et 22.10% et selon la fonction d'inclusion à test de monotonie, il ne varie plus qu'entre 10.91% et 13.76%. Ces résultats montrent que l'AI selon la fonction d'inclusion à test de monotonie fournit des intervalles de confiance plus petits que par rapport à l'AI selon la fonction d'inclusion naturelle, et plus grands par rapport à la méthode de Monte Carlo lorsque certains paramètres du modèle sont incertains.

La raison principale qui explique la différence des résultats réside dans la sous-division uniforme des supports des variables multi-occurrentes qui va conduire à travailler sur des

supports intervalles plus petits, et dans l'évaluation de la fonction d'inclusion sur chaque partition, qui lorsqu'elle est monotone par rapport à une entrée, va être évaluée à l'aide d'opérations sur les réels (et non sur les intervalles). Ce sont ces deux mécanismes qui vont conduire à réduire le pessimisme. Dans le cas particulier du modèle gaussien qui traduit de manière simple des phénomènes purement physiques, le modèle est monotone par rapport à chacune des entrées incertaines, ce qui permet de calculer la fonction d'inclusion à test de monotonie en un coup (sans division uniforme) et sans aucun pessimisme. C'est donc l'intervalle minimal qui est ici évalué.

Finalement, la réduction du pessimisme dans l'intervalle de confiance est remarquable pour cette application dans le cas de la fonction d'inclusion à test de monotonie, ce qui montre l'intérêt de bien choisir la fonction d'inclusion lors de l'utilisation de la méthode d'analyse par intervalles pour améliorer la précision des résultats obtenus.

2.3 Conclusion

Nous nous sommes intéressés dans ce chapitre à présenter les différentes techniques dont l'utilisation est devenue relativement classique dans le cadre de la propagation des incertitudes.

Dans la première partie, nous avons présenté deux approches : l'approche probabiliste par la méthode de Monte Carlo et l'approche ensembliste basée sur les techniques d'analyse par intervalles. Concernant la méthode de Monte Carlo, nous avons décrit le principe utilisé pour la propagation des incertitudes qui est basé sur un échantillonnage aléatoire qui conduit à une sous estimation de l'intervalle de confiance recherché. Pour la méthode d'analyse par intervalles, nous avons présenté un rappel des règles du calcul par intervalles ainsi que des principaux problèmes rencontrés lors de la manipulation des intervalles. Cette approche souffre essentiellement d'un problème de pessimisme qui conduit a une surestimation de l'intervalle de confiance recherché. Des exemples illustrent que selon la façon dont on écrit la fonction d'inclusion, l'intervalle résultat est différent. Nous avons choisi de présenter plusieurs types de fonctions d'inclusion comme la fonction d'inclusion naturelle, la forme centrée, celle de Taylor,... Nous avons ensuite discuté de l'efficacité de ces différentes fonctions d'inclusion et rappelé que si la fonction d'inclusion naturelle était efficace et simple d'utilisation, les fonctions d'inclusion plus sophistiquées amélioraient la précision de l'intervalle résultat, et ce d'autant plus que les supports intervalles pour les entrées sont petits. L'intervalle solution le moins pessimiste correspond de manière générale à la fonction d'inclusion ayant des occurrences moindres des variables et travaillant sur des intervalles de faible taille.

La deuxième partie de ce chapitre a été consacrée à une application à la dispersion atmosphérique pour estimer l'intervalle de confiance sur la concentration en gaz libéré sous le vent. Les résultats obtenus par la méthode Monte Carlo ont été comparés à ceux obtenus par la méthode de l'analyse par intervalles. L'analyse par intervalles et la méthode de Monte Carlo conduisent respectivement à des approximations sur-estimées et sous-estimées de l'intervalle de confiance recherché sur la concentration en gaz. En termes de temps de calcul, l'analyse par intervalles est plus rapide par rapport à la simulation Monte Carlo. C'est pour ces raisons qu'il sera intéressant de proposer une nouvelle approche dans le chapitre 4 qui vise à améliorer la précision de l'intervalle de confiance au regard de la simulation Monte Carlo avec un nombre d'échantillons plus restreint afin de réduire le temps d'exécutions. Nous allons nous concentrer dans le chapitre suivant sur des méthodes permettant la génération de cartographies. Celle-ci reposeront en partie sur les méthodes de propagation d'incertitudes dans un modèle qui viennent d'être rappelées.

Chapitre 3

Réalisation des cartographies : méthodes et applications

Lorsqu'un accident de TMD survient, plusieurs questions sont posées par les acteurs qui gèrent les risques TMD comme par exemple :

– Quel est le type de matière dangereuse transportée et quels sont les dangers associés ?
– Quelles sont les distances d'effets possibles et les zones de danger et de sécurité ?
– Quels sont les enjeux qui doivent être protégés ou évacués dans un premier temps ?

Les réponses à ces questions aident à évacuer la population menacée dans un périmètre de sécurité, à confiner les établissements sensibles et à mettre en place des barrières pour dérouter les véhicules afin d'éviter d'autres accidents et pour réduire les conséquences possibles. La dispersion accidentelle de gaz toxique constitue un scénario d'accident où il peut être intéressant de connaître les distances d'effets pour mieux dimensionner la réponse. Plusieurs modèles analytiques et logiciels de simulation couplés à un système d'information géographique (SIG) estiment les distributions spatiales de ces effets à l'intérieur desquelles les cibles sont recensées [108]. Ces logiciels basés sur le principe de l'égalité de diffusion des effets autour du point d'occurrence de l'accident se limitent souvent à représenter les distances des effets dans les cas les plus simples sous la forme d'un cercle (indépendamment de la direction du vent), ou d'un panache dont l'évaluation demande plus de ressources de calculs.

Les résultats de simulation obtenus par ces logiciels sont dépendants d'un très grand nombre de paramètres et de variables physiques ou chimiques, et l'inconvénient réside dans les incertitudes liées à ces entrées qui vont conduire à des incertitudes au niveau des cartographies réalisées. Si les résultats obtenus au niveau de la distribution spatiale du gaz toxique ne sont pas fiables en raison de la non prise en compte de ces incertitudes, cela peut conduire à une réponse inadéquate, par exemple une évacuation insuffisante de la zone impactée ou au contraire à une sollicitation exagérée des moyens de secours.

3.1 Problème à résoudre

L'objectif principal de ce chapitre est de produire une carte indiquant les zones de sécurité et de danger pour l'organisation de l'intervention d'urgence par exemple. Pour un problème de dispersion de produit toxique, la zone de danger (respectivement de sécurité) correspond à la zone géographique où la concentration en gaz est supérieure (respectivement inférieure) à un seuil choisi (effet létaux ou irréversibles sur la santé). Mais comme la réalisation des cartographies est basée sur des modèles d'effets où existent un certain nombre de paramètres ou de variables d'entrée connus avec imprécision, il est nécessaire de prendre en compte les incertitudes sur les entrées des modèles, afin d'obtenir des cartes représentatives de ce qui se passe sur le terrain. En raison de ces incertitudes une troisième zone située entre les zones de sécurité et de danger va apparaitre : c'est la zone d'incertitudes où la concentration peut éventuellement dépasser la valeur seuil, sans que ce soit forcément le cas. Cela dépend des valeurs réellement prises (mais inconnues) par les entrées incertaines parmi l'ensemble de leurs valeurs possibles. Les modèles de dispersion utilisés calculent la concentration de gaz en différents points du panache en fonction des équations associées. Pour réaliser une cartographie, c'est-à-dire trouver les positions où la concentration s'avère être inférieure ou supérieure à une valeur seuil donnée on doit inverser les modèles étudiés. Deux modèles d'effets sont utilisés d'un point de vue applicatif. Le premier est un modèle analytique (gaussien) et le second est un modèle par code (SLAB). Concernant la réalisation de cartographies à l'aide de modèles d'effets incertains, l'approche probabiliste est applicable à n'importe quel type de modèle (analytique, code source), tandis que l'approche ensembliste nécessite une forme analytique du modèle pour en concevoir une fonction d'inclusion, donc elle est avant tout applicable pour le modèle gaussien. Il est aussi possible de l'appliquer dans le cas de modèles par code sous certaines conditions, essentiellement de monotonie, au détriment de certains de ses avantages, comme nous le verrons dans le chapitre 4.

3.2 Résolution avec la méthode probabiliste

Globalement, la méthode probabiliste est spécifique au problème posé, c'est-à-dire que la manière de résoudre le problème d'inversion varie en fonction du type du modèle, des grandeurs d'entrée et de la façon dont les sorties calculées sont renvoyées. Nous allons donc détailler cette approche spécifiquement pour chacun des deux modèles utilisés.

3.2.1 Modèle analytique

Le modèle gaussien de dispersion atmosphérique calcule initialement la concentration en gaz en un point donné selon l'équation suivante :

$$\begin{aligned} c_k &= f(x_k, y_k, z_k, u_{ref}, z_{ref}, h, q, a_y, a_z, b_y, b_z, c_y, c_z) \\ &= \frac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} (a_y x_k^{by} + c_y)(a_z x_k^{bz} + c_z)} \times exp\left[-\frac{1}{2}\left(\frac{y_k}{a_y x_k^{by} + c_y}\right)^2\right] \times \\ &\left\{exp\left(-\frac{1}{2}\left(\frac{z_k - h}{a_z x_k^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\frac{z_k + h}{a_z x_k^{bz} + c_z}\right)^2\right)\right\} \end{aligned} \tag{3.1}$$

L'objectif est de réaliser une cartographie horizontale en 2D à $z_k = 2m$ fixé (hauteur proche des voix respiratoires des victimes potentielles). On doit donc rechercher les valeurs des abscisses x_k et ordonnées y_k telles que la concentration aux points de coordonnées (x_k, y_k, z_k) soit inférieure à une valeur seuil notée c_s pour la zone de sécurité, supérieure pour la zone de danger.

Tout d'abord la variable x_k est discrétisée sur tout le support de l'étude : $x_{k,k=1,\ldots,n}$ de sorte que $x_{k+1} = x_k + \triangle x$ avec $\triangle x$ une constante choisie et $x_1 = 0$ (point d'origine de l'émission). Supposons dans un premier temps que les autres entrées soient parfaitement connues (sans incertitude) et recherchons la frontière entre les deux zones précitées. En partant de l'expression analytique précédente, pour une valeur x_k fixée, la procédure d'inversion consiste à rechercher formellement les valeurs symétriques de la dernière inconnue y_k vérifiant l'égalité $f(x_k, y_k, z_k, u_{ref}, z_{ref}, h, q, a_y, a_z, b_y, b_z, c_y, c_z) = cs$. Ce qui conduit à :

$$\begin{aligned} &exp\left[-\frac{1}{2}\left(\frac{y_k}{a_y x_k^{by} + c_y}\right)^2\right] \times \left\{exp\left(-\frac{1}{2}\left(\frac{z_k - h}{a_z x_k^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\frac{z_k + h}{a_z x_k^{bz} + c_z}\right)^2\right)\right\} = \\ &c_s \times \frac{2\pi u_{ref} h^{0.33} (a_y x_k^{by} + c_y)(a_z x_k^{bz} + c_z)}{q z_{ref}^{0.33}} \end{aligned} \tag{3.2}$$

puis à :

$$\left(\frac{y_k}{a_y x_k^{by} + c_y}\right)^2 = -2 \times \ln\left[c_1 \times \frac{2\pi u_{ref} h^{0.33} (a_y x_k^{by} + c_y)(a_z x_k^{bz} + c_z)}{q z_{ref}^{0.33}\left\{exp\left(-\frac{1}{2}\left(\frac{z_k - h}{a_z x_k^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\frac{z_k + h}{a_z x_k^{bz} + c_z}\right)^2\right)\right\}}\right] \tag{3.3}$$

Une solution existe si le terme de droite est positif, et elle est alors donnée par :

$$y_k = \pm(a_y x_k^{by} + c_y) \times \sqrt{-2 \times \ln\left[c_1 \times \frac{2\pi u_{ref} h^{0.33} (a_y x_k^{by} + c_y)(a_z x_k^{bz} + c_z)}{q z_{ref}^{0.33}\left\{exp\left(-\frac{1}{2}\left(\frac{z_k - h}{a_z x_k^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\frac{z_k + h}{a_z x_k^{bz} + c_z}\right)^2\right)\right\}}\right]} \tag{3.4}$$

Le panache calculé étant symétrique par rapport à l'axe des abscisses, nous ne traiterons dans la suite que la solution positive par soucis de simplicité et pour limiter la quantité de calculs.

Cherchons maintenant à faire le même travail mais en prenant en compte les incertitudes sur les entrées du modèle de dispersion. En utilisant une approche de type Monte Carlo et en choisissant une taille d'échantillon N, N valeurs $y_{k,i}$, $i = 1, ..., N$ pour N combinaisons aléatoires des entrées incertaines $(q, u_{ref}, a_z, a_y, b_z, b_y)$ du modèle sont calculées. La valeur $y_{k,i}$ représente donc l'ordonnée du point $(x_k, y_{k,i}, z_k)$ où la concentration c_k atteint la valeur seuil c_s pour la ième combinaison des entrées tirées. Elle peut être supérieure ou inférieure à c_s en ce même point pour les autres combinaisons. Nous sommes donc en train d'évaluer la zone d'incertitudes qui va être déterminée en évaluant le maximum $y_{k,max}$ et le minimum $y_{k,min}$ des N valeurs $y_{k,i}$, $i = 1, ..., N$. Concrètement, les valeurs limites $y_{k,max}$ et $y_{k,min}$ expriment les frontières entre les zones de sécurité, danger et d'incertitudes. Parmi les N tirages effectués, une localisation en un point (x_k, y_k) avec $y_k < y_{k,min}$ correspond à une concentration en gaz nécessairement supérieure à c_s. Si $y_k > y_{k,max}$, la concentration en gaz est forcément inférieure à c_s. Si $y_{k,min} < y_k < y_{k,max}$, il est impossible de conclure avec certitude.

Donc une fois ce travail accompli pour tous les x_k, les deux ensembles ordonnés de points $(x_k, y_{k,min})$ et $(x_k, y_{k,max})$ définissent, pour une taille N d'échantillons suffisante, les frontières y_{Max} et y_{Min} de trois zones (figure 3.1) :

- la zone de sécurité où la concentration en gaz doit être inférieure à c_s pour toutes les valeurs possibles des entrées incertaines du modèle,
- la zone de danger où la concentration en gaz doit être supérieure à c_s pour toutes les valeurs possibles des entrées incertaines du modèle,
- la zone d'incertitudes où la concentration en gaz peut être supérieure ou inférieure à c_s en fonction des valeurs possibles des entrées incertaines du modèle.

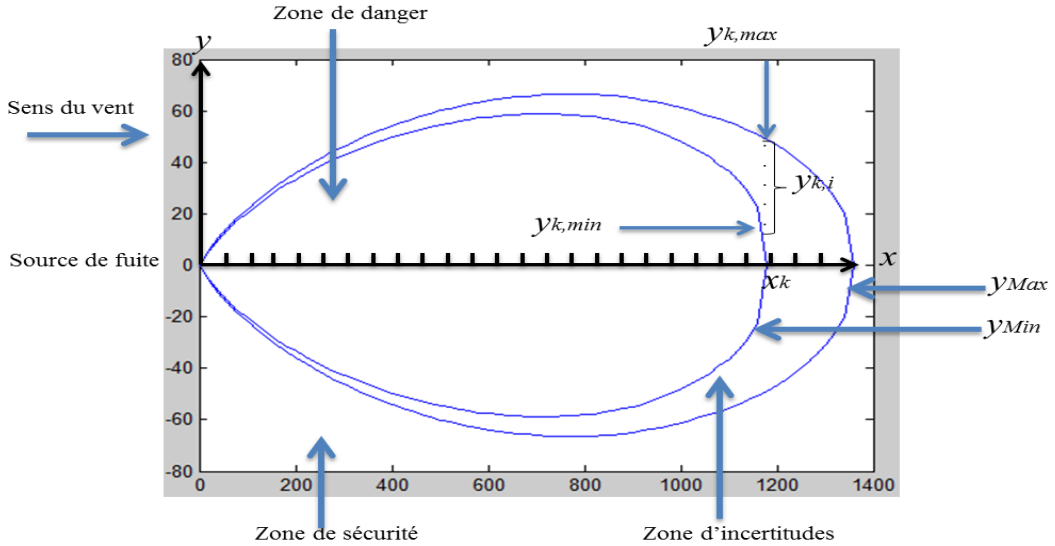

FIGURE 3.1: Zones déterminées par l'inversion probabiliste du modèle gaussien

3.2.2 Modèle complexe par code

Le modèle complexe de dispersion atmosphérique SLAB estime en sortie les concentrations volumiques moyennées sur le temps dans un plan horizontal z_k fixé par l'utilisateur, et calculées selon une grille de points de coordonnées (x_k, y_k) que le logiciel fixe automatiquement, comme on l'a vu dans la section 1.4.1.2.
L'axe des abscisses est discrétisé avec un pas variable (plus faible à proximité de la source) et pour chaque valeur x_k obtenue, la concentration est évaluée pour 6 ordonnées différentes $y_{k,j}$, $j = 1, ..., 6$, avec $y_{k,j} = BBC_k.N_j$ et $N_1 = 0, N_2 = 0.5, N_3 = 1, N_4 = 1.5, N_5 = 2, N_6 = 2.5$.
Pour rappel, BBC_k est une variable interne du modèle qui désigne la demi-largeur effective du nuage à chaque position x_k sous le vent.
Les concentrations calculées pour une simulation sont renvoyées sous la forme du tableau 3.1.

x_k	BBC_k	$y_{k,1}$	$y_{k,2}$	$y_{k,3}$	$y_{k,4}$	$y_{k,5}$	$y_{k,6}$
x_1	BBC_1	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}
x_2	BBC_2	c_{21}	c_{22}	c_{23}	c_{24}	c_{25}	c_{26}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
x_n	BBC_n	c_{n1}	c_{n2}	c_{n3}	c_{n4}	c_{n5}	c_{n6}

TABLE 3.1: Tableau des concentrations renvoyées par SLAB

Pour un jeu des grandeurs d'entrée (une simulation), l'objectif est de trouver pour une valeur x_k donnée la valeur de y_k telle que la concentration au point (x_k, y_k) soit égale à c_s, en clair inverser le tableau 3.1 de concentrations renvoyé par SLAB.
Pour chaque x_k , on recherche la valeur de concentration $c_{k,j}$ immédiatement inférieure, c'est à dire la valeur de l'indice j tel que $c_{k,j} < c_s <= c_{k,j-1}$. Ensuite à partir des données du tableau 3.1, une interpolation linéaire est réalisée afin d'en déduire la valeur y_k, soit :

$$a = \frac{c_{k,j} - c_{k,j-1}}{y_{k,j} - y_{k,j-1}} \quad \text{et} \quad b = c_{k,j-1} - (a \times y_{k,j-1}) \tag{3.5}$$

ce qui conduit à : $y_k = \dfrac{c_s - b}{a}$.

Certes l'interpolation réalisée aurait été nettement plus précise si le nombre (fixé à 6) d'ordonnées utilisées avait été plus élevé. Mais l'objectif était de ne pas modifier le code SLAB, simplement d'utiliser les données renvoyées pour réaliser la cartographie.

Prenons maintenant en compte les incertitudes sur les entrées du modèle. Pour cela, N simulations pour N combinaisons aléatoires des entrées incertaines du modèle sont exécutées avec la méthode de Monte Carlo pour calculer à chaque fois une nouvelle valeur

interpolée $y_{k,i}$, $i = 1, ..., N$. Notons que les valeurs discrétisées de x_k sont identiques pour toutes les simulations alors que les autres grandeurs renvoyées $(BBC_k, y_{k,j}, c_{k,j})$ varient en fonction des valeurs tirées des entrées.

Finalement le maximum $y_{k,max}$ et le minimum $y_{k,min}$ de $y_{k,i}$, $i = 1, ..., N$ sont déterminés respectivement. $y_{k,max}$ et $y_{k,min}$ expriment les frontières entre les zones de sécurité, danger et d'incertitudes. Une localisation en un point (x_k, y_k) avec $y_k < y_{k,min}$ correspond à une concentration en gaz supérieure à c_s. Si $y_k > y_{k,max}$, la concentration en gaz est inférieure à c_s. Si $y_{k,min} < y_k < y_{k,max}$, la concentration en gaz peut être inférieure ou supérieure à c_s en fonction des incertitudes.

Comme pour la modèle analytique trois zones ont été définies (figure 3.2) :
- la zone de sécurité où la concentration en gaz doit être inférieure à c_s pour toutes les valeurs possibles des entrées incertaines du modèle,
- la zone de danger où la concentration en gaz doit être supérieure à c_s pour toutes les valeurs possibles des entrées incertaines du modèle,
- la zone d'incertitude où la concentration en gaz peut être supérieure ou inférieure à c_s en fonction des valeurs possibles des entrées incertaines du modèle.

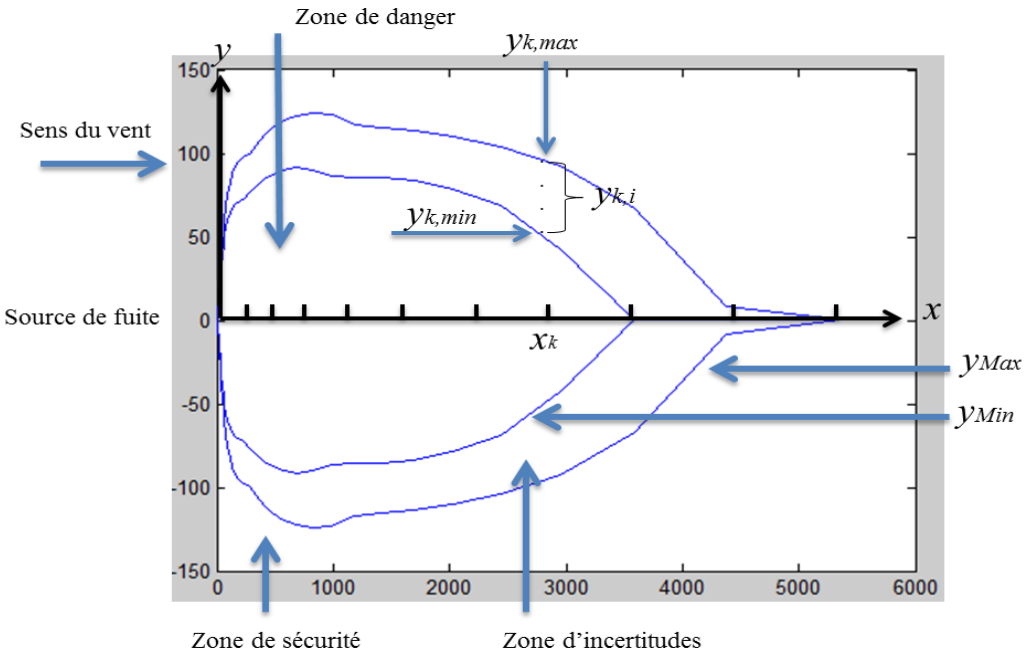

FIGURE 3.2: Zones déterminées par l'inversion probabiliste du modèle SLAB

3.3 Résolution avec la méthode ensembliste

3.3.1 Méthode pour résoudre le problème de cartographie

La condition générale pour réaliser les cartographies selon l'approche ensembliste est que le modèle analytique doit être d'une part dérivable par rapport à chacune de ses entrées incertaines, et d'autre part qu'il doit être possible de construire des fonctions d'inclusion des équations du modèle et de leurs dérivées par rapport à chacune des entrées incertaines.
Le calcul des zones de danger et de sécurité peut alors être posé sous la forme d'un ensemble de contraintes à satisfaire (CSP) exprimant l'ensemble des points tels que la concentration en gaz soit inférieure à un seuil donné (zone de sécurité) ou bien supérieure à ce même seuil (zone de danger). En imposant un support intervalle approprié pour la concentration, le problème d'inversion du modèle de dispersion intervalles par rapport aux variables de localisation géographique conduit à calculer complètement ce CSP à l'aide d'une méthode d'inversion ensembliste afin d'obtenir une union de boites définissant la zone recherchée.

3.3.1.1 Problème de satisfaction de contraintes par intervalles (ICSP)

Les problèmes de satisfaction de contraintes par intervalles (ICSP) expriment un cadre simple et formel pour représenter et résoudre des problèmes dans de nombreux domaines d'ingénierie [109–111].

Definition 3.1. (Contrainte).
Une contrainte S est une relation sur un ensemble de variables réelles $v_1, v_2, ..., v_n$. Plusieurs types de contraintes existent comme par exemple les contraintes floues ([Zadeh, 1971][112] et [Dubois and Prade, 1980][30]), probabilistes [113],... et notamment sur les variables à supports intervalles.

Exemple 12. [100] Soit le vecteur réel $\boldsymbol{v}= [x, y, z]^T$, la contrainte sphérique S_{sphere} pour le vecteur $\boldsymbol{v}\in \mathbb{R}^3$ s'écrit :

$$S_{sphere} : \|\boldsymbol{v}\|_2 = 1 \Rightarrow x^2 + y^2 + z^2 = 1, \tag{3.6}$$

où $\|\boldsymbol{v}\|_2$ est la norme euclidienne du vecteur $\boldsymbol{v}$.

Un problème de satisfaction de contraintes est caractérisé par un ensemble de variables, un ensemble de domaines associés et un ensemble de contraintes sur ces variables.
Une contrainte C est définie par :

- l'ensemble $V = \{v_1, v_2, ...v_n\}$ des variables intervenant dans cette contrainte,
- la relation $s = f(v_1, v_2, ...v_n)$ reliant la variable de sortie s aux variables d'entrée v_i,
- l'ensemble $D = \{D_1, D_2, ..., D_n, D_f\}$ des domaines associés à chacune des variables d'entrée et à la variable de sortie.

Definition 3.2. Un CSP est défini par un ensemble de contraintes à satisfaire :

$$\{C_1, C_2, ..., C_m\}$$

Les contraintes qui nous intéresseront par la suite sont celles manipulant des variables appartenant à des supports intervalles. Les CSP à résoudre se présenteront plus spécifiquement sous la forme du triplet $(V, D, \boldsymbol{f})$ où :

- V est l'ensemble des variables et est divisé en deux sous ensembles, celui des composantes du vecteur $\boldsymbol{u}$ de $\mathbb{R}^n$ représentant les inconnues du problème d'inversion à résoudre, et celui des composantes du vecteur $\boldsymbol{p}$ de $\mathbb{R}^p$ représentant les grandeurs incertaines (c'est-à-dire mal connues) du problème,
- D est l'ensemble des supports intervalles des $n + p$ variables d'entrée et des m sorties des différentes composantes de $\boldsymbol{f}$,
- $\boldsymbol{f}$ est un champ de vecteur de $\mathbb{R}^{(n+p)}$ dans $\mathbb{R}^m$.

Les supports des grandeurs inconnues u_i peuvent éventuellement être non bornés et représentent l'espace de recherche (par exemple $[0, +\infty[$ pour une grandeur dont on sait seulement qu'elle est positive) et l'objectif est de rechercher les valeurs de ces variables consistantes avec le CSP. Les grandeurs p_i représentent des données connues (certes avec imprécision) du problème et, contrairement aux variables u_i, l'objectif n'est pas de les déterminer, elles paramètrent simplement le CSP. Leurs supports sont généralement bornés et plus petits que ceux des variables inconnues.

L'ensemble des contraintes du CSP à résoudre peuvent se mettre sous la forme :

$$CSP : \begin{cases} f_1(u_1, ...u_n, p_1, ..., p_p) = s_1 \\ . \\ . \\ . \\ f_m(u_1, ...u_n, p_1, ..., p_p) = s_m \end{cases} \tag{3.7}$$

Avec $u_i \in [u_i]_{i=1,...,n}$, $p_i \in [p_i]_{i=1,...,p}$ et $s_i \in [s_i]_{i=1,...,m}$

Notion de consistance

La résolution d'un CSP consiste donc à trouver les valeurs admissibles à affecter aux variables inconnues permettant de satisfaire l'ensemble des contraintes. Une valeur est dite consistante avec un CSP, si elle vérifie toutes les contraintes de ce CSP.

Reprenons le cas de l'exemple 12, un CSP pour la contrainte sphérique est défini par

$$\begin{cases} x^2 + y^2 + z^2 = 1 \\ (x, y, z) \in [x] \times [y] \times [z]. \end{cases} \tag{3.8}$$

Si on prend $[x] = [-1, 1], [y] = [-1, 1]$ et $[z] = [-1, 1]$, alors l'ensemble $\mathbb{S}_1$ des points consistants avec le CSP énoncé en 3.8, est donné par l'intersection entre l'hypercube $[-1, 1]^{\times 3}$ et la surface de la sphère centrée sur l'origine et de rayon 1 notée $D_{r=1}$, ce qui conduit à

$$\mathbb{S}_1 = D_{r=1} = \left\{ \ (x, y, z) \in [-1, 1]^{\times 3} | x^2 + y^2 + z^2 = 1 \ \right\}. \tag{3.9}$$

Tandis que si on prend $(x, y, z) \in [-\frac{1}{2}, \frac{1}{2}]^{\times 3}$, l'ensemble des points consistants pour la contrainte devient

$$\mathbb{S}_2 = D_{r=1} \cap [-\frac{1}{2}, \frac{1}{2}]^{\times 3} = \varnothing. \tag{3.10}$$

où $D_{r=1}$ est l'ensemble des points situés sur la surface d'une sphère en trois dimensions de rayon $r = 1$.

Les résultats obtenus avec le test de consistance montrent que tous les points du pavé $[-\frac{1}{2}, \frac{1}{2}]^{\times 3}$ sont inconsistants avec la contrainte $x^2 + y^2 + z^2 = 1$.

On remarque que la consistance d'un point est une notion assez simple. Elle se traduit par la satisfaction de l'égalité qui compose la contrainte.

3.3.1.2 Inversion ensembliste pour la résolution d'un CSP intervalle

Le problème du calcul de l'image réciproque d'un ensemble par une fonction est reconnu comme étant un problème de grande importance. Le calcul par intervalles offre une approche pour résoudre ce problème. Cette approche a été formalisée, développée et utilisée pour résoudre des problèmes d'estimations dans le travail de Jaulin [114] et Moore [115].

L'inversion ensembliste consiste à trouver l'ensemble des solutions satisfaisant un ensemble de contraintes données par :

$$\begin{bmatrix} . \\ . \\ s_k \\ . \\ . \end{bmatrix} = \begin{bmatrix} . \\ . \\ f_k(u_k, p_k) \\ . \\ . \end{bmatrix} \tag{3.11}$$

que l'on réécrira de manière plus compacte

$$\boldsymbol{s} = \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p})$$

où :
- le champ de vecteur $\boldsymbol{f}$ regroupe l'ensemble des relations des contraintes,
- le vecteur $\boldsymbol{s}$ contient les différentes sorties s_k,
- les vecteurs $\boldsymbol{u}$ et $\boldsymbol{p}$ regroupent les entrées inconnues et les entrées incertaines du modèle.

Par exemple dans le cas du problème de cartographie basé sur le modèle gaussien, $\boldsymbol{f}$ sera réduit à une unique relation donnée par l'expression (3.1), la sortie $\boldsymbol{s}$ correspondra à la concentration en gaz c_k, le vecteur $\boldsymbol{u}$ sera composé des coordonnées x_k et y_k et le vecteur $\boldsymbol{p}$ sera constitué de toutes les autres entrées incertaines ou connues du modèle $(u_{ref}, z_{ref}, h, a_y, a_z, b_y, b_z, c_y, c_z, q, z)$.
L'objectif est de déterminer l'ensemble des valeurs du vecteur $\boldsymbol{u}$, noté $S_{u,\exists p}$, dont les images par le champ de vecteurs $\boldsymbol{f}$ appartiennent à l'ensemble d'arrivée imposé $[\boldsymbol{s}]$ pour au moins une valeur de $\boldsymbol{p}$.

$$S_{u,\exists p} = \boldsymbol{f}^{-1}(\boldsymbol{s}) = \{\boldsymbol{u} \in [\boldsymbol{u}] / \exists \boldsymbol{p} \in [\boldsymbol{p}] \wedge \exists \boldsymbol{s} \in [\boldsymbol{s}] \wedge \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p}) = \boldsymbol{s}\} \tag{3.12}$$

On note par S_{ext} et S_{int} deux ensembles de valeurs (vides au début) définis comme des unions de boîtes de solutions $[\boldsymbol{u}]_i$.

Brièvement, la méthode d'inversion ensembliste consiste à :

- réaliser un pavage (découpage) du domaine de recherche $[\boldsymbol{u}]$ en pavés plus petits $[\boldsymbol{u}]_i$,
- déterminer les pavés image $[\boldsymbol{s}]_i = [\boldsymbol{f}]([\boldsymbol{u}]_i, [\boldsymbol{p}])$ de chaque $[\boldsymbol{u}]_i$.

Après le calcul du pavé $[\boldsymbol{f}]([\boldsymbol{u}]_i, [\boldsymbol{p}])$ par la fonction d'inclusion $[\boldsymbol{f}]$, la consistance des éléments du pavé $[\boldsymbol{u}]_i$ est testée à l'aide du principe d'élimination. Ainsi $[\boldsymbol{u}]_i$ (figure 3.3) est défini comme :

- Satisfaisant si $[\boldsymbol{f}]([\boldsymbol{u}]_i, [\boldsymbol{p}])$ est inclus dans le domaine $[\boldsymbol{s}]$, (tout point de $[\boldsymbol{u}]_i$ est alors solution du CSP 3.12 puisque ce calcul est garanti), ce qui conduit à transférer $[\boldsymbol{u}]_i$ dans les listes de boites S_{ext} et S_{int}.
- Insatisfaisant si l'intersection entre $[\boldsymbol{f}]([\boldsymbol{u}]_i, [\boldsymbol{p}])$ et $[\boldsymbol{s}]$ est vide, (aucun point de $[\boldsymbol{u}]_i$ n'est solution), ce qui conduit à éliminer entièrement le pavé $[\boldsymbol{u}]_i$ des solutions.
- Indéterminé sinon, ce qui signifie que seulement une partie de $[\boldsymbol{u}]_i$ est solution. La démarche consiste alors :
  1. soit à ranger le pavé $[\boldsymbol{u}]_i$ dans la liste S_{ext} s'il satisfait aux tolérances imposées (par exemple le pavé est jugé suffisamment petit pour que la précision demandée soit atteinte),
  2. sinon à découper le pavé $[\boldsymbol{u}]_i$ en 2 par bissection et à recommencer ce test sur les deux sous-pavés obtenus.

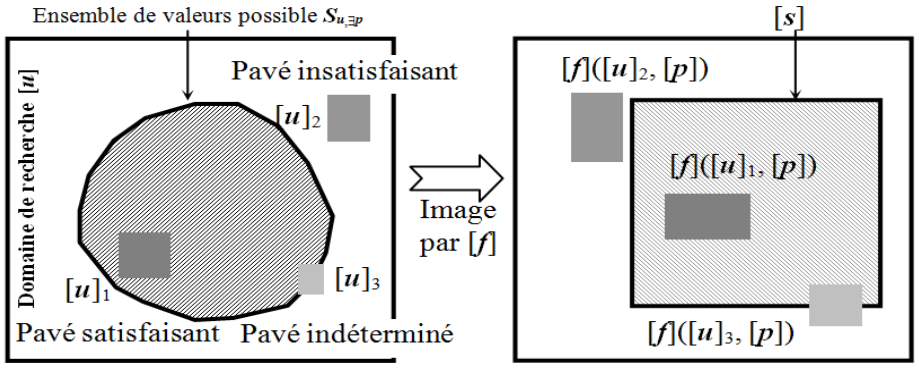

FIGURE 3.3: Inversion ensembliste et principe d'élimination

A l'issue, on obtient deux listes de pavés S_{ext} et S_{int} qui définissent respectivement :
– une approximation extérieure de $S_{u,\exists p}$: $S_{ext} \supset S_{u,\exists p}$
– une approximation intérieure de $S_{u,\exists p}$: $S_{int} \subset S_{u,\exists p}$
Toutes les solutions au problème d'inversion se trouvent dans l'approximation extérieure et tout point de l'approximation intérieure est solution.

On peut montrer [59] que, tous les points de S_{int} satisfont en réalité toutes les contraintes pour toutes les valeurs possibles de $\boldsymbol{p}$ que dans ces conditions, l'ensemble S_{int} est aussi une approximation intérieure (c'est-à-dire ne se compose que d'éléments) de l'ensemble de valeurs suivant $S_{u,\forall p}$:

$$S_{u,\forall p} = \{\boldsymbol{u} \in [\boldsymbol{u}] / \forall \boldsymbol{p} \in [\boldsymbol{p}] \wedge \exists \boldsymbol{s} \in [\boldsymbol{s}] \wedge \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p}) = \boldsymbol{s}\} \tag{3.13}$$

ce qui conduit à la relation d'ordre suivante :

$S_{int} \subset S_{u,\forall p} \subset S_{u,\exists p} \subset S_{ext}$

Cela signifie que plus important est le pavé $[\boldsymbol{p}]$, plus grand est l'ensemble solution S_{ext}, mais plus mince devient S_{int} parce que les contraintes sont plus difficiles à satisfaire pour toutes les valeurs de $\boldsymbol{p}$. Le fait de pouvoir travailler avec un quantificateur existentiel ou universel sur $\boldsymbol{p}$ est intéressant, car il permet de retenir l'une ou l'autre des deux approximations comme solution en fonction des objectifs visés. Pour le problème de cartographie par exemple, pour déterminer la zone où des actions doivent être mises en œuvre (cumul des zones de danger et d'incertitudes), c'est-à-dire les points où la concentration peut dépasser la valeur seuil pour certaines valeurs des grandeurs incertaines (quantificateur existentiel sur $\boldsymbol{p}$) et comme il est préférable de la surestimer (plutôt que le contraire), c'est donc S_{ext} qui sera retenue. Au contraire, pour déterminer la zone de sécurité ne comprenant que des points où la concentration est plus faible quelles que soient les incertitudes (quantificateur universel sur $\boldsymbol{p}$), et comme on préfère la sous-estimer, on retiendra S_{int}.

Une méthode de réduction (ou de contraction) est différente dans le sens où elle s'adresse au problème suivant : pour un pavé $[\boldsymbol{u}]_i$ donné, l'objectif est de déterminer un autre pavé $[\boldsymbol{u}]'_i$ inclus dans le précédent, le plus petit possible, tel que toutes les solutions éventuellement contenues dans $[\boldsymbol{u}]_i$ le soient aussi dans $[\boldsymbol{u}]'_i$. Cette approche limite l'explosion du nombre de pavés, cependant rien n'assure qu'il y aura effectivement réduction.

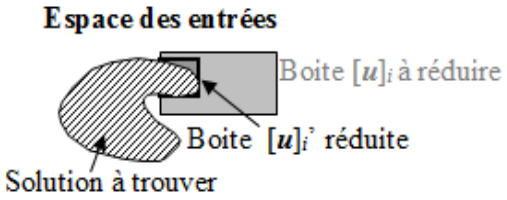

FIGURE 3.4: Principe de réduction

De façon à compenser les points faibles respectifs des approches par bissection et par réduction, la solution communément utilisée consiste alors à coupler les deux [111, 114], en contractant par exemple au maximum un pavé $[\boldsymbol{u}]_i$, puis lorsque cette réduction n'est plus significative, en le bissectant puis en relançant l'algorithme sur les nouveaux pavés obtenus.

Exemple 13. Considérons le CSP à résoudre défini par les deux contraintes suivantes :

$$\boldsymbol{f}(\boldsymbol{u}, p) = \begin{bmatrix} (u_1 - 1)^2 + p \times u_2 \\ u_1^2 + u_2^2 \end{bmatrix}$$

avec les supports intervalles suivants : $[\boldsymbol{u}] = \begin{bmatrix} [-5, 5] \\ [-5, 5] \end{bmatrix}, [p] = [0.9, 1.1], [\boldsymbol{s}] = \begin{bmatrix} [0, 2] \\ [1, \infty] \end{bmatrix}$

L'approximation intérieure (en vert)S_{int} de $S_{u,\forall p}$ (3.13) et l'approximation extérieure (en jaune) S_{ext} de $S_{u,\exists p}$ (3.12) sont présentés sur la figure 3.5. On remarque que S_{ext} et S_{int} encadrent correctement le domaine de solution exact calculé théoriquement et délimité par l'enveloppe en bleu. La précision de cet encadrement dépend des tolérances choisies dans l'algorithme d'inversion. Plus elles seront élevées, plus les pavés solution seront petits et nombreux, et meilleures seront les approximations intérieures et extérieures. La zone frontière (pavés en rouge, éléments de S_{ext} sans appartenir à S_{int}) montre l'influence du paramètre incertain p sur les sorties. Son épaisseur dépend à la fois de l'incertitude sur p et des tolérances choisies.

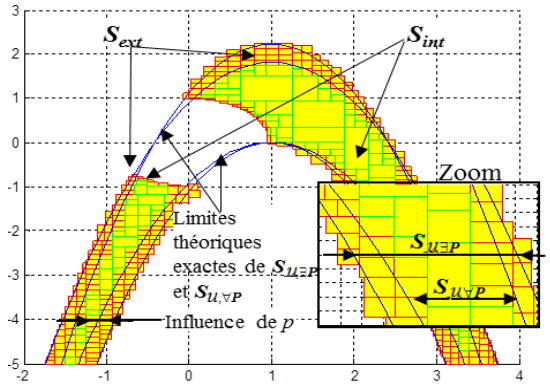

FIGURE 3.5: Ensembles de valeurs possibles S_{ext} et S_{int}

3.3.2 Méthode générique pour résoudre d'autres types de problèmes

L'approche générique CSP couplée avec l'inversion ensembliste [111, 116–119] permet de résoudre d'autres types de problèmes inverses comme la détection et localisation de défauts [120, 121] sur les installations industrielles, ou l'estimation (d'état ou paramétrique) de grandeurs non observées sur les procédés utilisés. Dans les sections suivantes, nous nous focaliserons sur le TMD en surveillant par exemple le système d'instrumentation mesurant la concentration en gaz, en estimant le débit de fuite inconnu ou en localisant parmi l'ensemble des camions géolocalisés transportant le produit incriminé celui accidenté provoquant l'émission de gaz toxique.

3.3.2.1 Problème de détection et localisation de défauts

L'évaluation des risques commence par la détection d'un phénomène dangereux ; cette détection peut être basée sur un réseau de capteurs. Il est alors nécessaire de vérifier régulièrement la validité des informations communiquées par le réseau de capteurs, pour déterminer s'ils fonctionnent correctement et enclencher une opération de maintenance le cas échéant. Si un ou plusieurs capteurs renvoient des informations erronées lors de la survenue d'un accident, cela va conduire à des conséquences sur les zones calculées de risque. On peut pour cela imaginer des tests réalisés en lâchant dans des conditions connues un gaz traceur inoffensif et récupérer les informations mesurées pour en vérifier la cohérence. La détection de défauts dans un réseau de capteurs repose sur la génération et l'évaluation d'un ensemble de relations de redondance analytiques obtenues en structurant les équations du modèle d'effets et les mesures de capteurs sous la forme de relations sensibles aux défauts qui doivent être détectés. Ces relations peuvent être considérées comme un ensemble de contraintes qui doivent être satisfaites dans le cas de l'absence de défaut. La détection de défaut dans un réseau de capteurs consiste finalement à tester si les concentrations en gaz mesurées sont cohérentes avec le modèle de

dispersion de référence, qui conduit à calculer l'ensemble solution pour ce problème de satisfaction de contrainte (CSP), afin de vérifier si celui-ci est "vide" ou non. Un ensemble solution "vide" implique qu'au moins une contrainte ne peut être satisfaite et que donc au moins l'un des capteurs utilisés fournit de l'information erronée. Il est clair qu'il est nécessaire de prendre en compte les incertitudes dans le modèle d'effets lors du test de cohérence pour différencier un problème de modélisation de la présence réelle d'un défaut sur un capteur. Une fois une anomalie détectée, on peut rechercher plus précisément la ou les contraintes non satisfaites pour essayer de localiser le capteur incriminé. On parle dans la littérature de méthodes FDI (pour Fault Detection and Isolation).
Un modèle vu comme un ensemble de contraintes est exprimé comme suit :

$$\boldsymbol{s} = \begin{bmatrix} . \\ . \\ s_k \\ . \\ . \end{bmatrix} = \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p}) = \begin{bmatrix} . \\ . \\ f_k(\boldsymbol{u}_k, \boldsymbol{p}_k) \\ . \\ . \end{bmatrix} \tag{3.14}$$

Par exemple dans le cas où le modèle d'effets utilisé est le modèle gaussien de dispersion, l'étape de génération de relations de redondance qui consiste à structurer les équations et les mesures du modèle va conduire à construire un CSP en empilant autant de relations f_k du modèle gaussien définies par (3.1) que de capteurs (3.14). Comme chaque grandeur incertaine de $\boldsymbol{p}$ est définie comme une variable bornée, sa valeur réelle est inconnue, mais elle appartient à un ensemble de valeurs réalisables défini comme un intervalle dont les bornes sont connues. Déterminer si les mesures de capteurs sont cohérentes avec le modèle de référence conduit à rechercher les valeurs inconnues de $\boldsymbol{u}$ satisfaisant ce CSP, c'est à dire l'ensemble $S_{u,\exists p}$ (ou une estimation) solution de ce problème de satisfaction des contraintes intervalles, afin de vérifier si celui-ci est vide ou non [122]. Un défaut est détecté si au moins une contrainte présente dans (3.14) ne peut être satisfaite ; c'est-à-dire si aucune valeur possible de $\boldsymbol{u}$ et $\boldsymbol{p}$ appartenant respectivement à $[\boldsymbol{u}]$ et $[\boldsymbol{p}]$ existe de telle sorte que $\boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p})$ appartient à $[\boldsymbol{s}]$. Si toutes les contraintes sont satisfaites, le réseau de capteurs est supposé être sans défaut, même si un défaut peut être effectivement masqué par les incertitudes du modèle. D'un point de vue pratique, les tests de cohérence peuvent être obtenus d'abord par le calcul de S_{ext} qui traite le CSP (3.12). La méthode d'inversion ensembliste est en mesure de rechercher toutes les solutions du CSP (3.12). Néanmoins évaluer complètement $S_{u,\exists p}$ n'est pas nécessaire dans le contexte FDI, car il suffit de trouver une boîte solution afin de conclure que $S_{u,\exists p}$ contient probablement au moins une solution. En outre, si cette boite appartient aussi à S_{int}, une information plus fiable est obtenue car il est garanti numériquement que $S_{u,\exists p}$ n'est pas vide.

3.3.2.2 Problème de localisation de la source de danger

Quand un accident entraînant l'émission d'un gaz toxique se produit, la position de la source de danger (le moyen de transport) peut être inconnue dans un premier temps. Il est alors possible de localiser à distance la source de danger en déterminant à partir du modèle d'effets et des emplacements géographiques connus des moyens potentiels de transport qui transportent des marchandises dangereuses, lequel est cohérent avec les mesures disponibles de l'intensité du phénomène dangereux. Le problème se pose de manière similaire à celui de la détection et localisation de défauts, sauf que le réseau de capteurs est considéré comme sans défaut. L'identification de la source de danger conduit à calculer l'ensemble solution pour plusieurs problèmes de satisfaction de contraintes associés (3.12) chacun à une source de danger possible afin de vérifier si celui-ci est "vide" ou non. Un moyen de transport est identifié comme la source de fuite si une émission à partir de ce lieu permet d'expliquer les mesures obtenues par les différents capteurs positionnés sous le vent, auquel cas existe une solution pour son CSP. En d'autres termes, toutes les contraintes sont satisfaites pour cette source si au moins une boîte solution de S_{ext} est calculée, compte tenu des incertitudes de modèle.

3.4 Application à la dispersion atmosphérique

3.4.1 Réalisation de cartographies à l'aide d'un modèle analytique (gaussien)

Cette étude a été appliquée à un exemple d'accident impliquant sous forme gazeuse de l'oxyde nitrique (NO). Ce gaz est toxique et a une densité relative à l'air de 1.04, de sorte qu'un modèle gaussien (3.1) est bien adapté pour modéliser la dispersion d'un tel gaz. Dans cette partie applicative, on a utilisé les mêmes données que celles présentées dans l'application du chapitre 2 (2.2.1) concernant la classe de stabilité de l'atmosphère, les valeurs nominales des entrées et les incertitudes correspondantes.

Le réseau de capteurs se compose de six capteurs indiqués par des points rouges sur la figure 3.6 et qui couvrent une zone dont l'axe x' des abscisses s'étend de 0 et 2500 mètres et l'axe des ordonnées y' est compris entre 0 et 1400 mètres. L'origine (0, 0) correspond au coin inférieur gauche de la zone de travail. Cette zone contient trois camions qui transportent de l'oxyde nitrique, positionnés sur différentes routes à un instant donné et indiqués sur cette même carte par des identifiants (camion 1, 2, 3). Le vent souffle quand à lui vers l'Est.

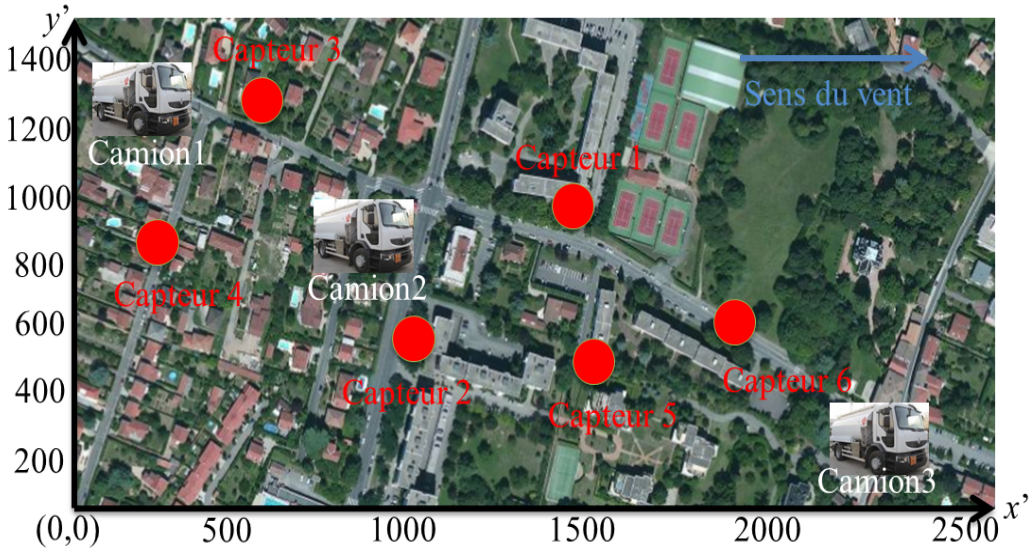

FIGURE 3.6: Cas d'étude

Deux systèmes de coordonnées sont utilisés. Le premier est le système de référence de la carte (x', y') dont l'origine $(0, 0)$ est le coin inférieur gauche comme indiqué sur la figure 3.6, et qui permet de positionner les camions et les capteurs dans la zone de travail. Le second noté (x, y), est utilisé pour le modèle gaussien, avec comme origine la position du camion qui est la source de la fuite et avec un axe x aligné avec la direction du vent. Le passage de l'un à l'autre des deux systèmes de coordonnées est effectué par une simple transformation de translation et de rotation si le vent souffle dans une direction différente.

Sur la figure 3.7, le système de coordonnées (x, y) correspond au cas où la source supposée est le camion 2 et la direction du vent est l'Est.

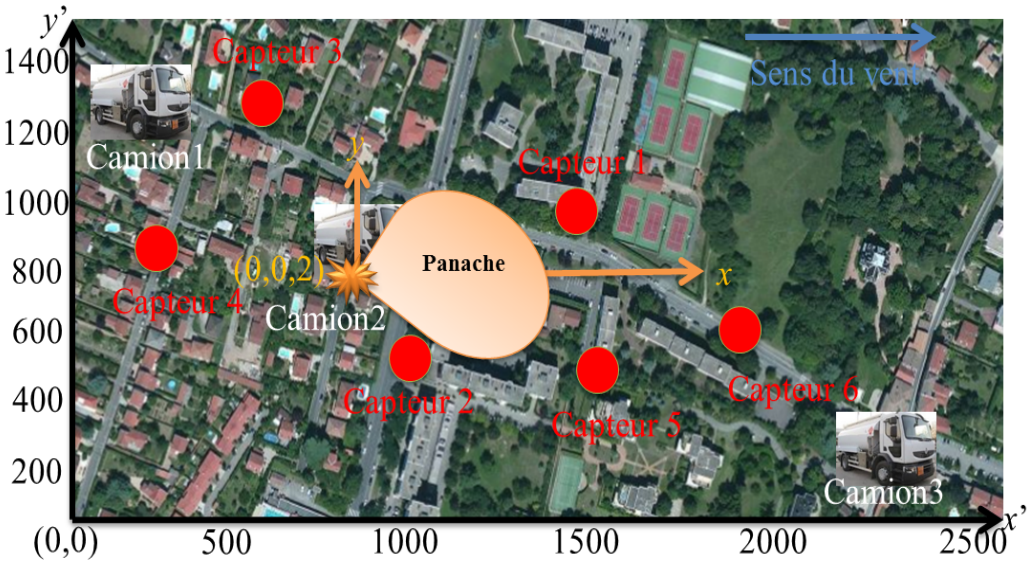

FIGURE 3.7: Systèmes de coordonnées

L'objectif est de déterminer les zones géographiques dans lesquelles la concentration en gaz est inférieure (zone de sécurité) ou supérieure (zone de danger et d'incertitudes) à un seuil réglementaire donné afin d'éviter par exemple des effets létaux ou irréversibles sur la santé. Dans cette étude, nous déterminons la région géographique dans laquelle la concentration en gaz est supérieure au seuil d'effets irréversible $c_1 = 0.123 g/m^3$[107]. Selon les entrées du modèle, le gaz toxique considéré ne constitue pas une situation de danger dans le cas où la concentration en gaz ne dépasse jamais c_1. Pour rappel, les incertitudes du modèle touchant les 6 grandeurs d'entrée $u_{ref}, q, a_y, a_z, b_y, b_z$ sont indiquées au début de la section 2.2.2.

3.4.1.1 Génération des cartographies avec la méthode probabiliste

Cette section traite le problème d'inversion via la simulation de Monte Carlo, c'est-à-dire l'évaluation des zones de danger, de sécurité et d'incertitudes.
Le nombre d'échantillons N est ajusté jusqu'à ce qu'aucune modification significative dans les zones calculées ne soient observées. Cela conduit à $N = 100.000$ échantillons avec la simulation de Monte Carlo, ce qui est un choix raisonnable et classique compte tenu du nombre $p = 6$ d'entrées incertaines du modèle. Le temps d'exécution est de l'ordre de 513.52 secondes, ce qui est conséquent.

D'un point de vue pratique, les limites de chaque zone sont calculées en répétant les opérations décrites dans la section 3.2.1 et les résultats obtenus sont donnés sur la figure 3.8.

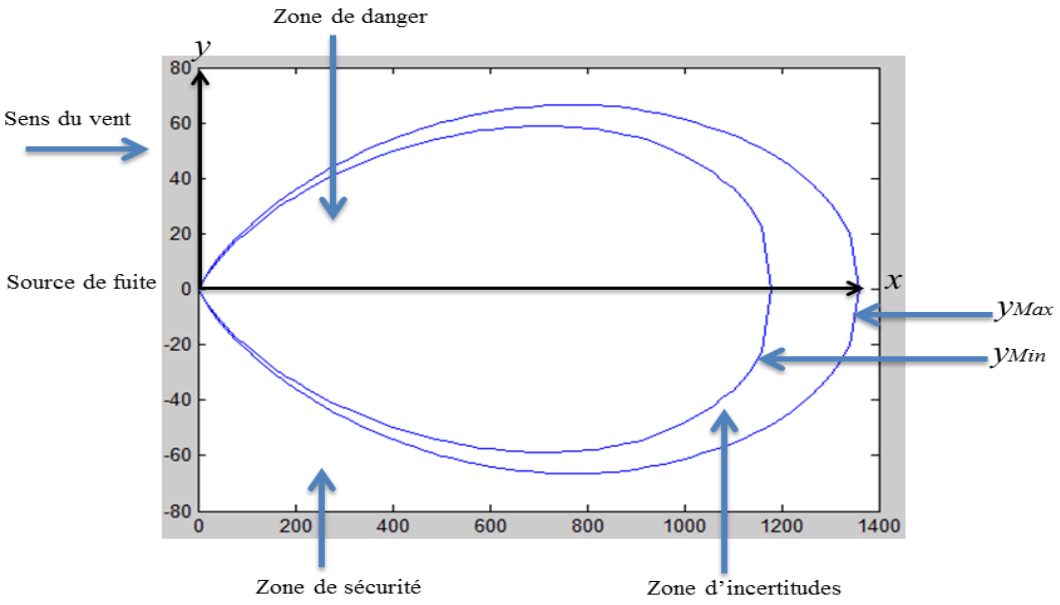

FIGURE 3.8: Panache de NO simulé par l'inversion probabiliste

3.4.1.2 Génération des cartographies avec la méthode ensembliste

La génération des cartographies par l'approche ensembliste consiste dans un premier temps à exprimer le problème sous la forme du CSP suivant :

$$s = f(\boldsymbol{u}, \boldsymbol{p}) \quad \text{avec} \quad \begin{pmatrix} \boldsymbol{u} = [x, y]^T \\ \boldsymbol{p} = [u_{ref}, z_{ref}, h, a_y, a_z, b_y, b_z, c_y, c_z, q, z]^T \end{pmatrix}$$

avec $x \in [1, 10000]$ et $y \in [-1000, +1000]$. Les incertitudes de modèle concernent les 6 grandeurs d'entrée $u_{ref}, q, a_y, a_z, b_y, b_z$ comme indiqué au début de la section 2.2.2, les autres entrées $(z_{ref}, h, c_y, c_z, z)$ sont connues et f sera réduite à une unique relation donnée par l'expression (3.1).
Comme l'objectif est ici de déterminer la zone de danger, où la concentration en gaz est supérieure au seuil d'effets irréversibles c_1, le support de la variable de sortie s représentant cette grandeur est imposé à $[s] = [c_1, \infty[$.
La résolution de ce CSP est ensuite réalisée par inversion ensembliste, ce qui fournit deux ensembles de boites S_{ext} et S_{int}.

Notons qu'il est numériquement garanti que la concentration en gaz de tous les points situés dans l'approximation intérieure S_{int} (boites encadrées en vert sur la figure 3.9) est supérieure à c_1 et cette propriété est respectée pour toutes les sources d'incertitudes c'est-à-dire pour toutes les situations appréhendées par le modèle de dispersion. De plus tous les points qui peuvent conduire à une concentration de gaz supérieure à c_1 pour au moins une combinaison particulière des entrées incertaines du modèle sont nécessairement contenus dans l'approximation extérieure S_{ext} (exprimée par les boites colorées en jaune sur la figure 3.9). Dans ces conditions, puisque S_{ext} surestime l'ensemble des points $S_{u,\exists p}$ où la concentration en gaz peut dépasser c_1 pour au moins une valeur possible de $\boldsymbol{p}$, son complément donne une sous-estimation de la zone de sécurité. Le domaine S_{int} fournit quant à lui une sous-estimation de la zone de danger (c'est-à-dire de $S_{u,\forall p}$ où on est sûr de dépasser c_1) et l'union de pavé $S_{ext} \backslash S_{int}$ située entre S_{ext} et S_{int} (boites encadrées en rouge) donne une surestimation de la zone d'incertitude (où on peut éventuellement dépasser c_1 en fonction des valeurs prises par les grandeurs incertaines). Le nombre des pavés traités par la méthode d'inversion ensembliste est de 6709. Le temps de calcul est de 53.36 secondes.

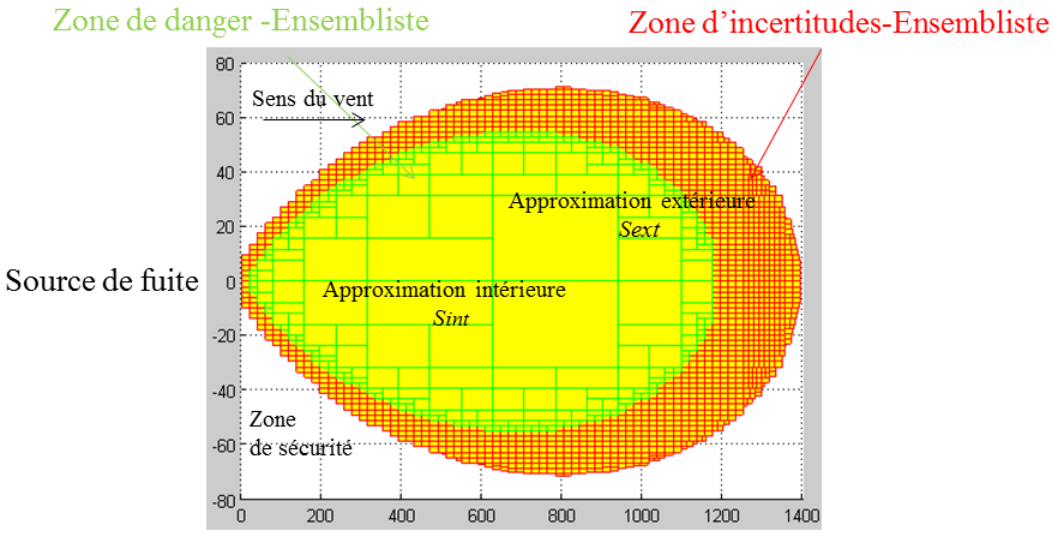

FIGURE 3.9: Panache de NO simulé par l'inversion ensembliste

3.4.1.3 Comparaison des cartographies obtenues par les deux méthodes

Nous pouvons facilement comparer les zones obtenues en fusionnant les résultats de l'inversion ensembliste avec les résultats de l'inversion probabiliste (figure 3.10). L'approximation intérieure S_{int} calculée avec l'inversion ensembliste donne une sous-estimation de la zone de danger alors que la zone à l'intérieur de la courbe y_{Min} obtenue par la simulation de Monte Carlo conduit à une surestimation. De la même manière, l'inversion ensembliste et l'inversion probabiliste calculent respectivement une surestimation $Sext \backslash Sint$ et une sous-estimation (domaine entre les courbes y_{Min} et y_{Max}) de la zone d'incertitudes. Plusieurs raisons expliquent la différence entre les deux approches. La première est due au principe même du fonctionnement de la méthode d'inversion ensembliste qui, en travaillant sur des fonctions d'inclusion, conduit à construire deux approximations intérieure S_{int} et extérieure S_{ext} de la zone de danger. La deuxième raison est que l'inversion probabiliste ne permet pas de prendre en compte toutes les combinaisons possibles de valeurs des entrées incertaines du modèle, ce qui conduit à sous-estimer la zone d'incertitudes.

En effet, il faudrait un nombre d'échantillons infini pour obtenir toutes les valeurs possibles de l'ordonnée $y_{k,i}$ recherchée et atteindre ses valeurs minimale $y_{k,min}$ et maximale $y_{k,max}$ exactes. Il s'ensuit que cette approche surestime la zone de danger. Le point intéressant est que dans ces conditions, les zones exactes de danger et d'incertitudes recherchées sont encadrées par les zones approchées calculées par l'inversion ensembliste et l'inversion probabiliste. Les deux méthodes dépendent de paramètres de réglage (tolérances sur les pavés solution pour l'inversion ensembliste et nombre de tirages pour l'inversion probabiliste). L'écart constaté entre les zones calculées par les deux méthodes donne donc de l'information sur la qualité du réglage opéré puisque celles-ci devraient

converger pour une précision très grande. Si cet écart est significatif, l'une des deux méthodes (au moins) ne donne pas satisfaction en termes de précision et les paramètres de réglage se doivent d'être ajustés. Dans le cadre de l'évaluation des risques dans le transport de matières dangereuses, il est a priori plus intéressant d'avoir une approximation extérieure de la zone où les risques sont élevés (zones cumulées de danger et d'incertitudes) avec une bonne précision au lieu d'en obtenir seulement une partie. Une sous-estimation peut conduire à une évacuation insuffisante ou la mise en place de barrières inadéquates de la zone dangereuse.

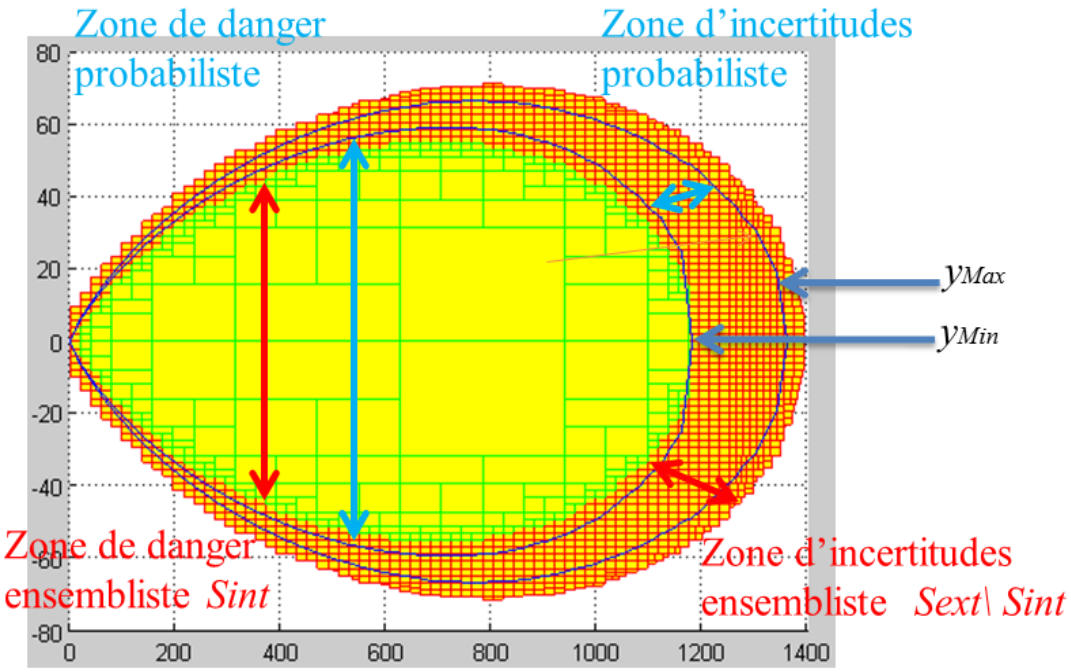

FIGURE 3.10: Panache de NO simulé par l'inversion ensembliste et probabiliste

3.4.1.4 Application au projet Geofencing

Dans le cadre du projet Geofencing où s'inscrit ce travail et présenté dans la section 1.2, une contribution dans la réalisation d'un démonstrateur de suivi et de supervision des TMD a été développée. Par le biais d'un web service, les informations issues de la "tour de contrôle" (géolocalisation GPS des camions, produits transportés,...) sont traitées dans le cas de la survenue d'un accident TMD afin de générer les cartographies à l'aide de l'approche ensembliste. Les zones calculées peuvent être reproduites sur une carte satellite afin de fournir des informations pratiques pour les services d'urgence (de plus amples détails sont fournis en Annexe A). Les figures 3.11 et 3.12 montrent la zone cumulée de danger et d'incertitudes calculée dans les sections précédentes sur une carte représentant une partie de l'agglomération grenobloise.

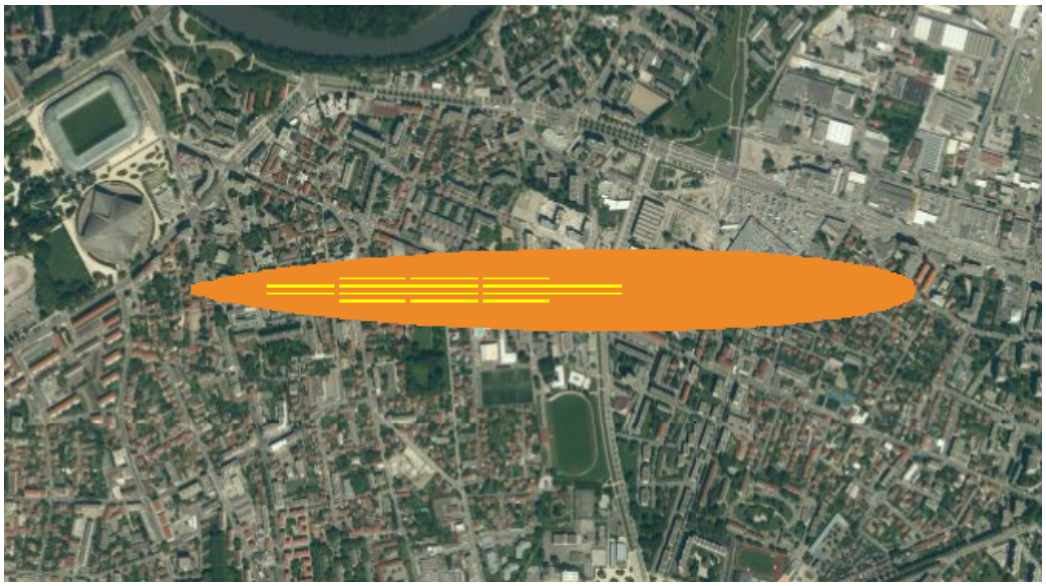

FIGURE 3.11: Zone cumulée de danger et d'incertitudes

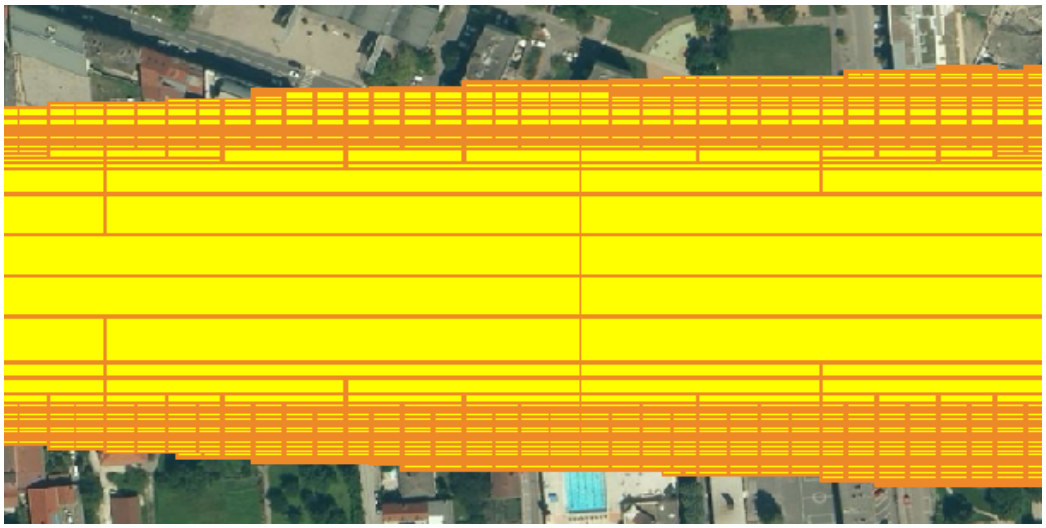

FIGURE 3.12: Zoom sur la zone calculée

3.4.2 Autres exemples d'application à l'aide d'un modèle analytique

3.4.2.1 Validité du modèle gaussien

Le modèle gaussien est valable pour la détermination de la dispersion atmosphérique pour les points géographiques qui se trouvent dans la direction du vent c'est-à-dire lorsque l'inégalité $x > 0$ est vérifiée (dans le système de coordonnées (x, y) propre au modèle). En d'autres termes ce modèle n'est pas valable pour les capteurs qui sont derrière le point de fuite comme indiqué sur la figure 3.13. Donc, nous pouvons en déduire qu'il existe deux types de contraintes en fonction de la position du capteur par rapport au point de fuite :

– Pour un capteur qui est en face du point de fuite, la contrainte de modélisation de la concentration en gaz est donnée par le modèle gaussien (3.1).
– Pour un capteur qui se trouve derrière le point de fuite, la contrainte devient $c_k = 0$.

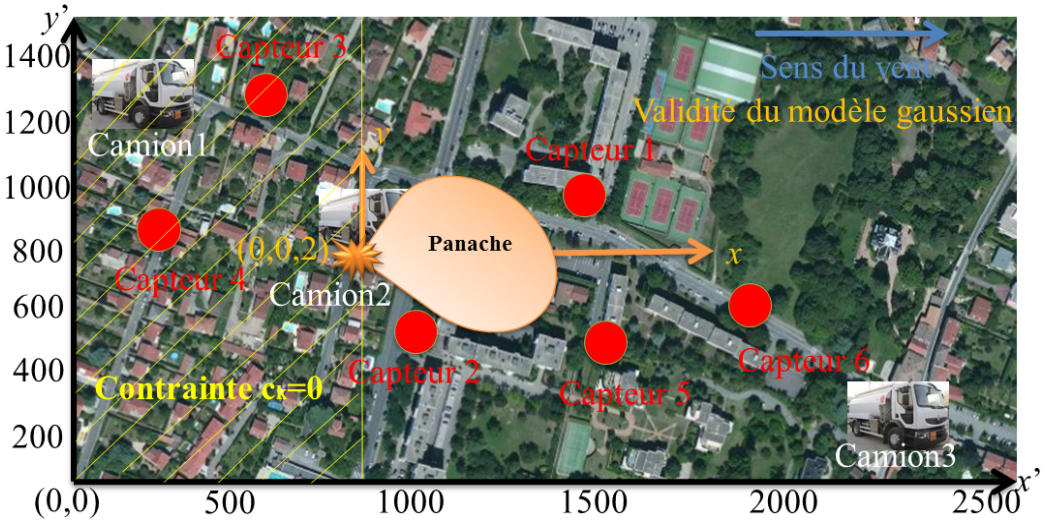

FIGURE 3.13: Validité du modèle gaussien

Le type de contrainte $c_k = 0$ est très intéressant pour un problème formulé sous la forme d'un CSP, car par exemple, si les mesures de concentration en gaz de capteurs qui sont derrière un camion donné sont différentes de 0, nous pouvons déduire automatiquement que ce camion n'est pas la source de la fuite (sous l'hypothèse qu'il n'y ait qu'un seul camion accidenté dans la zone à un instant donné).

3.4.2.2 Application au problème de détection et localisation de défauts

Pour détecter s'il existe une anomalie dans le réseau de capteurs, un gaz traceur est relâché à partir d'une position connue. Sa concentration est mesurée par les différents capteurs dispersés dans la zone de travail. L'objectif consiste à surveiller ces $n = 6$ capteurs de concentration en gaz, en vérifiant si les mesures sont cohérentes avec le modèle de dispersion (3.1) compte tenu des différentes sources d'incertitudes.
Comme indiqué dans la section 3.3.2.1, ce problème conduit à calculer partiellement l'ensemble de valeurs $S_{u,\exists p}$ sachant que le CSP à résoudre (3.14) est composé de n contraintes $f_k(), k \in 1, ..., n$ comme suit :

$$\begin{bmatrix} c_1 \\ . \\ . \\ c_n \end{bmatrix} = \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p}) \quad \text{avec} \quad \left(\overset{u=q}{\underset{\boldsymbol{p}=[u_{ref},z_{ref},h,a_y,a_z,b_y,b_z,c_y,c_z,x_1,y_1,z_1,...,x_n,y_n,z_n]^T}{}} \right)$$

Le terme c_k représente la mesure de concentration en gaz du $k$ème capteur, f_k exprime la contrainte utilisée pour estimer cette concentration à la position de ce capteur dont les coordonnées par rapport au point d'émission sont notées par (x_k, y_k, z_k).
Supposons par simplicité de présentation que le point d'émission du gaz traceur corresponde à la même position que celle du camion n°2 indiqué sur la figure 3.13, alors le CSP associé sera concrètement composé pour cet exemple des six contraintes suivantes :

$$
CSP : \begin{cases}
c_1 = \dfrac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} (a_y x_1^{by} + c_y)(a_z x_1^{bz} + c_z)} \times exp\left[-\frac{1}{2}\left(\dfrac{y_1}{a_y x_1^{by} + c_y}\right)^2\right] \times \\
\left\{exp\left(-\frac{1}{2}\left(\dfrac{z_1 - h}{a_z x_1^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\dfrac{z_1 + h}{a_z x_1^{bz} + c_z}\right)^2\right)\right\} \\
c_2 = \dfrac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} (a_y x_2^{by} + c_y)(a_z x_2^{bz} + c_z)} \times exp\left[-\frac{1}{2}\left(\dfrac{y_2}{a_y x_2^{by} + c_y}\right)^2\right] \times \\
\left\{exp\left(-\frac{1}{2}\left(\dfrac{z_2 - h}{a_z x_2^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\dfrac{z_2 + h}{a_z x_2^{bz} + c_z}\right)^2\right)\right\} \\
c_3 = 0 \\
c_4 = 0 \\
c_5 = \dfrac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} (a_y x_5^{by} + c_y)(a_z x_5^{bz} + c_z)} \times exp\left[-\frac{1}{2}\left(\dfrac{y_5}{a_y x_5^{by} + c_y}\right)^2\right] \times \\
\left\{exp\left(-\frac{1}{2}\left(\dfrac{z_5 - h}{a_z x_5^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\dfrac{z_5 + h}{a_z x_5^{bz} + c_z}\right)^2\right)\right\} \\
c_6 = \dfrac{q z_{ref}^{0.33}}{2\pi u_{ref} h^{0.33} (a_y x_6^{by} + c_y)(a_z x_6^{bz} + c_z)} \times exp\left[-\frac{1}{2}\left(\dfrac{y_6}{a_y x_6^{by} + c_y}\right)^2\right] \times \\
\left\{exp\left(-\frac{1}{2}\left(\dfrac{z_6 - h}{a_z x_6^{bz} + c_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\dfrac{z_6 + h}{a_z x_6^{bz} + c_z}\right)^2\right)\right\}
\end{cases} \tag{3.15}
$$

Les positions des capteurs et du point d'émission étant supposées parfaitement connues, les différentes sources d'incertitudes portent uniquement sur $u_{ref}, a_y, a_z, b_y, b_z$ et restent les mêmes que celles définies en section (2.2.2).
Le problème de détection de défauts conduit à résoudre ce CSP où la variable inconnue est le flux de sortie q dont le domaine de définition a été posé à [0,2500], les autres entrées $(h, z_{ref}, c_y, c_z, x_1, ..., z_n)$ étant connues. Si aucune valeur possible de $u(q)$ ne peut être trouvée, cela signifie qu'il y a incohérence entre les informations mesurées et le modèle de dispersion incertain. Au moins une contrainte du CSP (3.15) ne peut être satisfaite, ce qui conduit à conclure qu'un défaut affecte l'un des capteurs. Pour réaliser la localisation, il est possible d'appliquer la même démarche sur des sous-ensembles de contraintes du CSP complet et en fonction des CSP partiels sans solution de localiser le capteur défaillant à l'aide de tables de signatures (stratégies Dedicated Observer Scheme DOS ou Generalized Observer Scheme GOS [123, 124]).

Les tableaux 3.2 et 3.3 montrent un exemple de détection de défauts pour une émission de gaz traceur. Les six capteurs indiqués sur la figure 3.13 sont considérés. L'objectif consiste à réaliser six tests de cohérence différents, en appliquant à chaque fois un défaut spécifique à l'un des capteurs et en vérifiant si celui-ci est bien détecté. Pour chaque test un biais est ajouté sur un seul capteur et la cohérence entre cette mesure erronée et les autres qui sont saines est évaluée. Le tableau 3.2 spécifie, pour chaque capteur, la valeur mesurée et la plage d'incertitude (intervalle de confiance) sur celle-ci, lorsqu'ils fonctionnent normalement. Le tableau 3.3 indique pour chacun des 6 tests le biais ajouté pour simuler un défaut sur le capteur concerné. Tous les défauts indiqués dans le tableau 3.3 sont détectés malgré les incertitudes de modèle, ce qui signifie que pour chaque biais considéré, l'ensemble de solutions pour le CSP par intervalles à résoudre est vide. Les capteurs 3 et 4 sont derrière la source de gaz traceur par rapport au sens du vent, leurs mesures de concentration doivent être égales à 0. Ainsi, un biais sur ces capteurs égal à une petite quantité ϵ différente de 0 est nécessairement détecté.

Capteurs	Mesures	Plages d'incertitudes
Capteur n^o 1	0.8236	[0.7824, 0.8648]
Capteur n^o 2	0.0051	[0.0049, 0.0054]
Capteur n^o 3	0	0
Capteur n^o 4	0	0
Capteur n^o 5	0.1814	[0.1723, 0.1905]
Capteur n^o 6	0.0330	[0.0314, 0.0347]

TABLE 3.2: Mesures de concentration en gaz.

Tests	Biais	Ensemble de solutions calculé	Conclusion
Capteur n^o 1	0.1647	Vide	Défaut détecté
Capteur n^o 2	0.0014	Vide	Défaut détecté
Capteur n^o 3	ϵ	Vide	Défaut détecté
Capteur n^o 4	ϵ	Vide	Défaut détecté
Capteur n^o 5	0.0363	Vide	Défaut détecté
Capteur n^o 6	0.0099	Vide	Défaut détecté

TABLE 3.3: Résultats de la détection de défauts

3.4.2.3 Localisation de la source de la fuite

Lorsque les capteurs détectent la présence d'oxyde nitrique, l'objectif est de localiser la source de la fuite, c'est-à-dire déterminer le camion à l'origine de l'émission. Ce problème consiste à vérifier pour chaque source potentielle de la fuite s'il y a des valeurs possibles pour le débit de fuite inconnu q, qui soient cohérentes avec les mesures des capteurs et le modèle de dispersion (3.1). Ainsi, ce problème de localisation conduit à résoudre plusieurs

problèmes de satisfaction de contraintes associés aux différents camions présents dans la zone de travail. En effet, un CSP différent va être construit pour chaque camion puisque si certaines entrées sont identiques (paramètres de dispersion par exemple, vitesse du vent mesurée,...), d'autres comme la position relative des capteurs par rapport au point d'émission (le camion) changent. Ce problème conduit à calculer l'ensemble solution $S_{u,\exists p}$ sachant que pour un camion donné, le CSP (3.14) est composé de $n = 6$ contraintes $f_k(), k \in 1, ..., n$ comme suit, n désignant le nombre de capteurs.

$$\begin{bmatrix} c_1 \\ . \\ . \\ c_n \end{bmatrix} = \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{p}) \quad \text{avec} \quad \left(\overset{u=q}{\underset{\boldsymbol{p}=[u_{ref},z_{ref},h,a_y,a_z,b_y,b_z,c_y,c_z,x_1,y_1,z_1,...,x_n,y_n,z_n]^T}{}} \right)$$

Les triplets (x_i, y_i, z_i) définissent les positions spatiales des capteurs par rapport au camion testé qui représente l'origine du système de coordonnées associé au modèle gaussien (figure 3.7). Ainsi, pour un capteur donné, ces coordonnées changent pour chaque source de danger testée. Les autres composantes incertaines ou connues du vecteur $\boldsymbol{p}$ ont déjà été détaillées dans la section (2.2.2) et restent inchangées. A titre d'information, la répartition des $n = 6$ capteurs et la localisation des trois camions considérés sont dessinées sur la figure 3.7. Les plages d'incertitudes sur les mesures de concentration en gaz sont données dans le tableau 3.4. Les capteurs 3 et 4 sont derrière le camion accidenté à localiser par rapport au sens du vent, leurs mesures de concentrations valent donc 0.

Capteurs	Plages d'incertitudes
Capteur n^o 1	[0.7824, 0.8648]
Capteur n^o 2	[0.0049, 0.0054]
Capteur n^o 3	0
Capteur n^o 4	0
Capteur n^o 5	[0.1723, 0.1905]
Capteur n^o 6	[0.0314, 0.0347]

TABLE 3.4: Mesures des capteurs.

Un camion peut être la source de la fuite si l'ensemble solution de son CSP intervalle $S_{u,\exists p}$ n'est pas vide. Notons que, la prise de décision ne nécessite pas de calculer entièrement cet ensemble solution, l'obtention d'une première boite pour l'approximation extérieure ou intérieure permet de conclure. A l'opposé, un CSP sans solution indique que le camion associé n'est pas incriminé et donc n'est pas la source du danger.
Comme trois camions sont en mesure d'être à l'origine de l'émission de gaz, 3 CSP doivent être traités. Les résultats sont présentés dans le tableau 3.5, sachant que le vrai camion générant l'émission est le deuxième. Pour les camions 1 et 3, la méthode d'inversion ensembliste ne trouve pas de valeurs admissibles pour le débit de fuite. Pour le camion

2, des boites solutions du CSP associé sont trouvées de sorte que cela implique que ce camion est la source de danger.

Identificateur de camion	Ensemble de solutions calculé	Conclusion
Camion 1	vide	pas de fuite
Camion 2	non vide	fuite
Camion 3	vide	pas de fuite

TABLE 3.5: Résultats de localisation

3.4.3 Réalisation de cartographies à l'aide d'un modèle complexe par code (SLAB)

3.4.3.1 Cas d'étude

Cette application a pour but de traiter le cas d'un accident impliquant de l'ammoniac (NH3) qui est un gaz toxique plus lourd que l'air, sous sa forme liquéfiée, dans des conditions normales de température et de pression. L'objectif est de déterminer les zones géographiques dans lesquelles la concentration en gaz est inférieure (zone de sécurité) ou supérieure (zones de danger et d'incertitudes) à un seuil réglementaire donné en utilisant un modèle complexe par code qui aborde ce type de problématique comme SLAB.

SLAB est composé d'une trentaine de paramètres et variables d'entrée qui sont répartis en 4 groupes :

- le premier groupe contient les paramètres liés aux propriétés physico-chimiques du gaz émis (masse molaire du gaz...),
- le deuxième groupe contient les paramètres liés aux caractéristiques de déversement,
- le troisième groupe exprime les paramètres de champ (distance maximale sous le vent...),
- le quatrième groupe représente les paramètres liés aux conditions météorologiques ambiantes (vitesse du vent, température...).

Les valeurs des entrées ci-dessous correspondent à celles utilisées dans le deuxième cas d'étude détaillé dans le guide SLAB [2] et qui correspond à un rejet d'ammoniac. Certaines entrées ont toutefois été ajustées, comme la distance xffm augmentée de 2800 mètres à 5000 mètres pour devenir compatible avec le seuil de toxicité choisi.

Paramètre	Description	Valeur
wms	molecular weight of source gas (kg)	1.7031E-02
cps	vapor heat capacity, const. p. (j/kg-k)	2.0459E+03
tbp	boiling point temperature	2.3957E+02
cmed0	liquid mass fraction	8.1000E-01
dhe	heat of vaporization (j/kg)	1.1700E+06
cpsl	liquid heat capacity (j/kg-k)	4.6118E+03
rhosl	liquid source density (kg/m3)	6.0300E+02
spb	saturation pressure constant (k)	2.9760E+03
spc	saturation pressure constant (k)	0.0000E+00
ts	temperature of source gas (k)	2.3957E+02

TABLE 3.6: Les propriétés physico-chimiques du gaz

Paramètre	Description	Valeur
idspl	spill type	2
qs	mass source rate (kg/s)	1.0787E+02
as	source area (m2)	9.3000E-01
tsd	continuous source duration (s)	3.8100E+02
qtis	instantaneous source mass (kg)	0.0000E+00
hs	source height (m)	1.0000E+00

TABLE 3.7: Les caractéristiques de déversement

Paramètre	Description	Valeur
tav	concentration averaging time (s)	1.0000E+01
xffm	maximum downwind distrace (m)	5.0000E+03
zp(1)	concentration measurement height (m)	0.0000E+00
zp(2)	concentration measurement height (m)	1.0000E+00
zp(3)	concentration measurement height (m)	0.0000E+00
zp(4)	concentration measurement height (m)	0.0000E+00

TABLE 3.8: Les paramètres de champ

Paramètre	Description	Valeur
z0	surface roughness height (m)	3.0000E-03
za	ambient measurement height (m)	2.0000E+00
ua	ambient wind speed (m/s)	4.5000E+00
ta	ambient temperature (k)	3.1000E+02
rh	relative humidity (percent)	2.1300E+01
stab	atmospheric stability class value	4.5185E+00
ala	inverse monin-obukhov length (1/m)	2.2100E-02

TABLE 3.9: Les propriétés météorologiques

En se basant sur le travail de S. Pagnon [3] concernant l'étude du logiciel SLAB et plus spécifiquement l'analyse de sensibilité des grandeurs d'entrées sur la distance d'effets, nous avons décidé de ne mettre des incertitudes que sur les entrées les plus influentes sur la variation de la sortie.

Selon S. Pagnon, certaines entrées ont un effet monotone sur l'estimation des distances d'effets comme la température, et certaines variables ont (ou pourraient avoir) un effet

non-monotone sur l'estimation des distances d'effet comme la vitesse de vent. Les résultats obtenus ont confirmé l'influence très importante et prédominante de la taille de brèche, en d'autres termes le débit de fuite. Les autres paramètres ont une faible influence sur la sortie du modèle.
Les entrées incertaines dans ce cas d'étude sont :

Entrées	Incertitudes
Température (ta)	$\pm 1\%$
Vitesse du vent (ua)	$\pm 8\%$
Débit de fuite (qs)	$\pm 10\%$
L'humidité (rh)	$\pm 5\%$
Rugosité ($z0$)	$\pm 5\%$

TABLE 3.10: Les incertitudes sur les entrées du logiciel SLAB

Les champs de concentration sont comparés aux valeurs seuil de toxicité issues des fiches de toxicité aigüe de référence (Valeurs Seuils de Toxicité Aigues ou VSTAF). Les effets sélectionnés sont les effets irréversibles. Le seuil des effets irréversibles pour l'ammoniac pour une durée d'exposition de 10 s est $c_s = 866$ppm.

3.4.3.2 Génération des cartographies avec la méthode probabiliste

Cette section traite le problème de cartographie pour le cas d'étude détaillé dans la section précédente. L'inversion du modèle numérique SLAB pour déterminer les zones de danger et d'incertitudes est réalisée à l'aide de l'approche probabiliste développée dans la section (3.2.2).
SLAB a été développé dans le langage de programmation Fortran et comprend entre 3500 et 4000 lignes de code. Ce logiciel ne possède pas d'interface facilitant l'automatisation des calculs. Dans notre étude, notamment celui de la gestion des risques qui nous concerne et en particulier pour le développement d'une application web et la réalisation des cartographies, nous avons entrepris de re-coder entièrement SLAB en langage de programmation JAVA de façon à avoir plus de souplesse. Ce travail de transformation de code est passé par les phases de développement, de vérification et de test de code. Nous avons pu interfacer SLAB directement avec le méthode de Monte Carlo pour plus de rapidité.
Pour déterminer le nombre d'échantillons, celui ci est augmenté jusqu'à ce qu'aucune modification significative dans les zones calculées ne soient observées. Cela conduit à $N = 100.000$ échantillons avec la simulation de Monte Carlo, ce qui est un choix raisonnable et classique compte tenu du nombre $p = 5$ d'entrées incertaines du modèle. Le temps d'exécution pour générer la cartographie avec ce nombre d'échantillons est de 250 secondes.

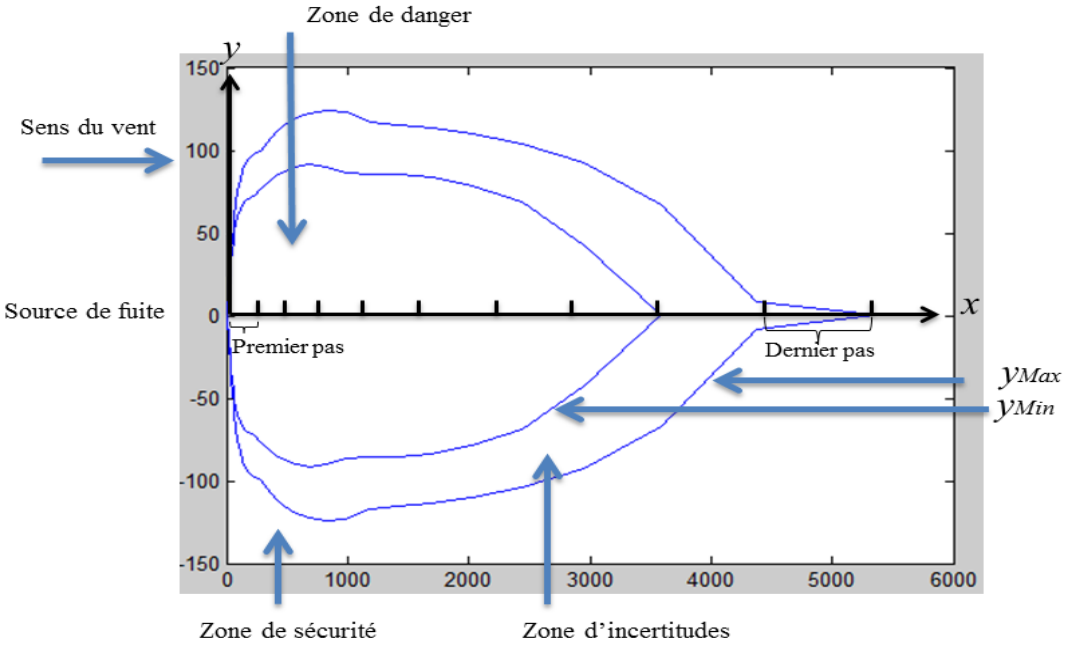

FIGURE 3.14: Panache de NH3 simulé par l'inversion probabiliste

Finalement, en observant les résultats obtenus, nous pouvons en déduire que nous avons réussi à déterminer les zones de danger, de sécurité et d'incertitudes en procédant à l'inversion du modèle numérique SLAB avec la méthode probabiliste. Pour ce qui concerne la précision des résultats, elle reste raisonnable avec une interpolation sur seulement six points sur l'axe y, mais pour améliorer la précision il faudrait augmenter ce nombre de points.

Remarque. Au niveau de la fermeture du panache sur l'axe des abscisses, on note la présence d'un pic important sur la courbe limite y_{Max} délimitant la zone d'incertitudes. Ce pic est dû à la discrétisation à pas non constant (et imposé par SLAB) de la variable x_k représentant les abscisses. Ce pas, très faible à proximité de la source, augmente ensuite avec x_k et est par exemple de l'ordre de 975 mètres sur l'extrémité droite du panache. Ne souhaitant pas modifier le modèle SLAB en lui-même, nous avons opté ici par la solution la plus simple, qui consiste à fermer la courbe sur la valeur discrétisée de x_{k+1} qui suit la dernière x_k où il a été possible d'interpoler une solution y_k (3.5). Il est aussi possible sinon d'extrapoler la courbe y_{Max} sur les derniers points pour obtenir une fermeture plus réaliste.

3.5 Conclusion

Dans ce chapitre, nous avons abordé la problématique concernant la génération des cartographies des zones de danger et de sécurité lorsqu'un accident TMD, qui survient,

entraine la dispersion d'un gaz toxique. Pour réaliser ces cartographies, deux modèles d'effets sont utilisés pour évaluer et comparer à des seuils réglementaires la concentration en gaz estimée dans la zone impactée. Le premier est un modèle analytique (gaussien) et le deuxième est un modèle plus sophistiqué par code (SLAB). La méthode probabiliste et la méthode ensembliste sont utilisées pour propager les incertitudes affectant les différentes entrées des modèles d'effets et les propager sur les cartographies obtenues. Cela se traduit par l'estimation d'une troisième zone, appelée zone d'incertitudes, située entre les zones de danger et de sécurité, où il est impossible de dire, compte tenu des incertitudes sur les entrées, si la concentration en gaz dépasse réellement ou non le seuil préconisé. L'approche probabiliste est applicable à n'importe quel type de modèle (analytique, code source), tandis que l'approche ensembliste nécessite une forme analytique du modèle dérivable et continue (plus précisément des fonctions d'inclusion du modèle et de ses dérivées).

Concernant le modèle gaussien, la réalisation des cartographies avec les deux approches a montré que l'inversion ensembliste génère un majorant des zones cumulées de danger et d'incertitudes alors que la seconde en calcule un minorant. Dans le cadre de l'évaluation des risques dans le transport de matières dangereuses, il est a priori plus intéressant d'avoir une approximation majorant la zone où les risques sont élevés avec une bonne précision au lieu d'en obtenir seulement une partie. Une sous-estimation peut conduire à une évacuation insuffisante ou la mise en place de barrières inadéquates.

L'intérêt d'utiliser les deux approches réside dans l'encadrement des zones de danger et d'incertitudes exactes. En effet, en ajustant les paramètres de réglages de chacune des deux approches pour améliorer la précision des zones estimées, les zones de danger et d'incertitudes évaluées convergent vers les zones exactes. L'encadrement obtenu apporte donc des informations complémentaires sur la précision des approximations calculées. Pour ce qui concerne le problème d'inversion à traiter, l'approche probabiliste est spécifique au modèle utilisé, alors que l'approche ensembliste reste générique en le formulant sous la forme d'un CSP intervalle qui sera résolu à l'aide d'une méthode d'inversion ensembliste. Notons que pour cette dernière approche, il est aussi possible de résoudre d'autres types de problèmes inverses comme la détection et localisation de défauts dans les capteurs qui mesurent la concentration et la localisation de la source de fuite.

Concernant le modèle SLAB, ce chapitre a montré qu'il est possible de propager les incertitudes sur la sortie de modèles par code en utilisant l'inversion probabiliste, d'où l'intérêt de rechercher des moyens pour améliorer les performances de l'approche probabiliste en terme de précision et de temps de calcul, ce qui sera l'objet du chapitre suivant.

Chapitre 4

Amélioration des performances : méthodes et application

L'approche probabiliste est souvent utilisée dans le cas où on cherche à simuler des modèles (notamment complexes) incertains et qui peuvent avoir un nombre important d'entrées pouvant conduire à un temps de calculs très élevé. Ce temps de calculs dépend en premier lieu de la durée nécessaire pour réaliser une simulation du modèle et sur laquelle il est difficile d'agir autrement qu'en construisant un modèle plus simple en faisant un compromis entre coûts en calculs et précision. On parle alors de réduction de modèle (analyse en composantes principales,...) ou de construction de (méta)modèles (régression, interpolation,...) ; ce qui ne sera pas le choix effectué ici. Il dépend ensuite de la taille de l'échantillon qui est généralement fonction du nombre d'entrées incertaines du modèle et qui nécessite de lancer un grand nombre de simulations. C'est sur la taille de cet échantillon que nous souhaitons agir pour limiter les temps de calcul, tout en maintenant une précision raisonnable dans les zones évaluées. Notons que l'approche ensembliste peut elle aussi être concernée par cette problématique puisque l'inversion ensembliste utilisée consiste, dans le cas où on a besoin d'inverser le modèle, à bissecter suivant certaines directions de l'espace de recherche dont la dimension est définie par le nombre d'entrées incertaines.

Le contexte de l'urgence dans la gestion des risques TMD rend nécessaire l'utilisation de modèles « adaptés » pour modéliser la dispersion atmosphérique du gaz toxique. On entend par modèles « adaptés », des modèles assurant une précision des zones évaluées et des temps de calculs raisonnables. L'objet de ce quatrième chapitre est donc de proposer une méthodologie générale, applicable à des modèles aussi bien analytiques que par code informatique, afin de réaliser les cartographies et d'améliorer les performances en termes de temps du calculs et de précision par rapport aux résultats obtenus au chapitre précédent. Cette méthodologie est divisée en 3 phases (figure 4.1) :

- **Phase 1** : Cette phase consiste à analyser préalablement les modèles d'effets, en fonction de leur nature, en vue d'être plus efficace lors des phases suivantes.
  Classiquement parmi les moyens de réduire le temps de calculs, le plus naturel consiste à diminuer le nombre d'entrées considérées incertaines, ce qui mécaniquement conduit à une réduction de la taille de l'échantillon. Il peut donc être intéressant de réaliser une analyse de sensibilité globale pour déterminer quelles sont les entrées incertaines les moins influentes qu'il est raisonnable de fixer à une valeur nominale tout en limitant le nombre de situations en sortie non prises en compte. Notons que les indices de sensibilité utilisés dépendront de la nature du modèle manipulé. Dans le cas d'une approche purement probabiliste, les indices bien connus de Sobol seront utilisés. En revanche pour des modèles par intervalles, des indices de sensibilité intervalles originaux ont été développés pour tenir compte des spécificités de ces modèles. Dans le cas de modèles de type code informatique dont il n'est pas possible de construire une fonction d'inclusion, une étude de la monotonie du modèle par rapport à chacune de ses entrées incertaines est réalisée. Les résultats de cette étude permettront de déterminer les entrées pour lesquelles il sera tout de même possible de travailler avec des supports intervalles pour propager les incertitudes en calculant l'intervalle minimal (2.7) sans utiliser d'opérateurs intervalles. Dans le cas de modèles où une fonction d'inclusion est disponible, une analyse de la multi-occurrence des entrées est réalisée pour déterminer quels sont les supports intervalles les plus intéressants à subdiviser pour mieux maîtriser le pessimisme engendré par le calcul par intervalles.
- **Phase 2** : Cette phase a pour objectif de proposer une nouvelle approche pour la prise en compte et la propagation des incertitudes mixant les approches probabiliste et ensembliste pour pouvoir profiter de leurs atouts respectifs. Issue dans son principe de la simulation de Monte-Carlo, **la première originalité** est que cette approche consiste à tirer des ensembles de valeurs (c'est-à-dire des supports intervalles) plutôt que de simples valeurs pour appréhender plus de situations possibles et propager ces incertitudes sur la sortie tout en utilisant un nombre de tirages réduit. **La seconde originalité** est que cette approche est utilisable pour tout type de modèles de dispersion spatialisés et statiques. Dans le cas d'un modèle analytique dont on peut construire une fonction d'inclusion, toutes les entrées incertaines sélectionnées suite à l'analyse de sensibilité seront tirées sous forme de petits supports intervalles, l'évaluation de la fonction d'inclusion par l'analyse par intervalles permettant le calcul du support de la concentration en gaz. Dans le cas d'un modèle par code, un travail similaire est effectué en tirant des supports intervalles pour les entrées par rapport auxquelles le modèle est monotone (accessoirement par morceaux sur les supports tirés) et de simples valeurs pour les entrées incertaines restantes.
- **Phase 3** : Cette dernière phase consiste à réaliser les cartographies en inversant les modèles d'effets à partir des résultats de la propagation d'incertitudes obtenus lors de

la phase 2.

Le principe général, proche de celui détaillé en section 3.2 pour la simulation de Monte-Carlo, est de réaliser une interpolation sur les supports intervalles calculés en sortie en différents points du panache pour déterminer plus précisément où la concentration en gaz en susceptible d'atteindre la valeur seuil choisie. Cette façon de procéder convient pour tous les types de modèle, néanmoins lorsque celui-ci est analytique et suffisamment simple à manipuler, il est possible de procéder formellement à l'inversion du modèle pour gagner en précision.

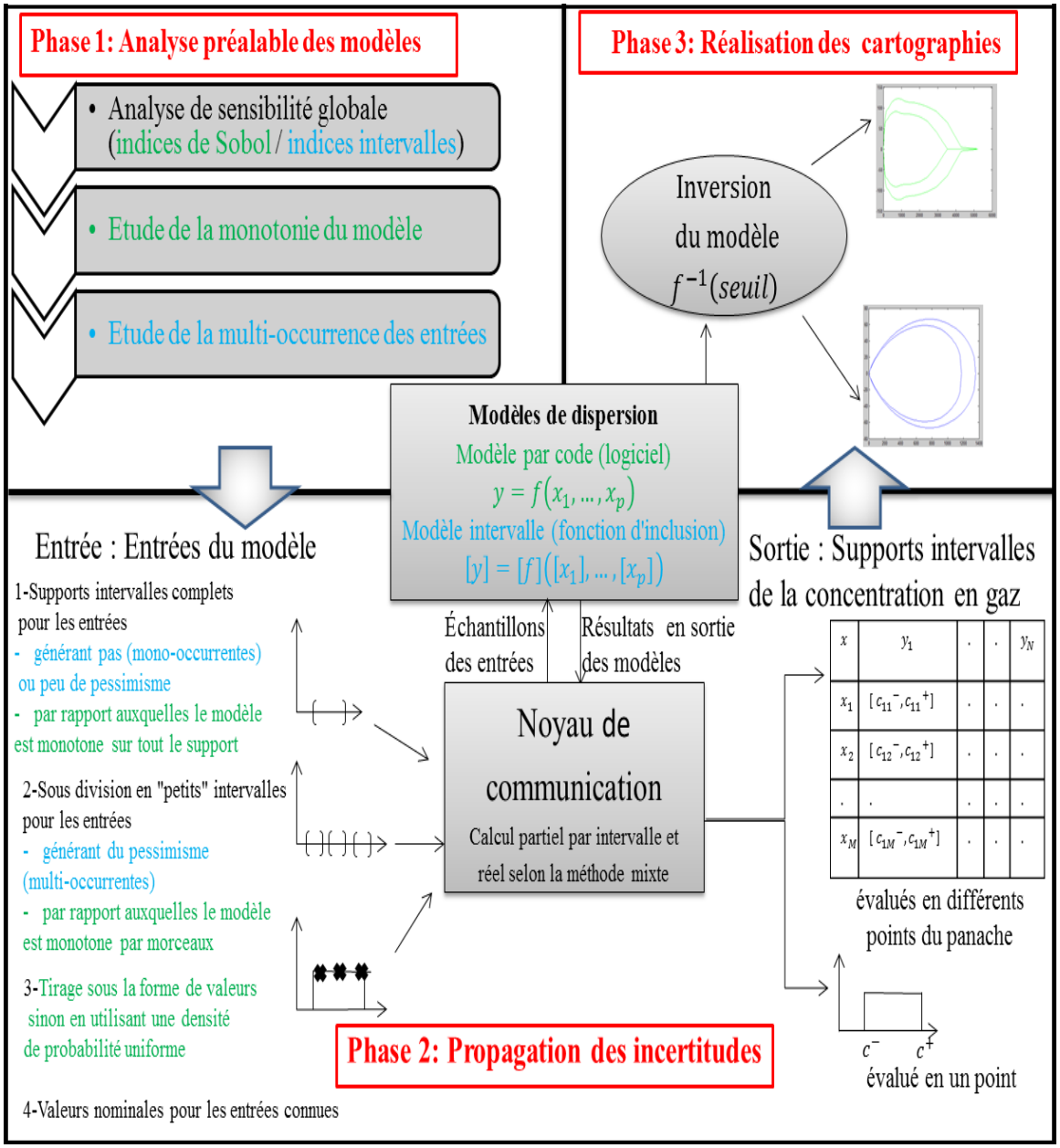

FIGURE 4.1: Architecture de la méthodologie proposée

Nous allons maintenant détailler plus précisément chacune de ces 3 phases.

4.1 Première phase : Analyse préalable des modèles

4.1.1 Analyse de sensibilité

L'objectif de l'analyse de sensibilité est de quantifier l'influence de l'incertitude attachée aux entrées du modèle sur la sortie estimée par le modèle. Plus particulièrement, l'intérêt pratique réside dans la détermination des entrées (ou groupes d'entrées) influentes. L'analyse de sensibilité consiste alors à calculer les indices de sensibilité pour chacune des entrées, ce qui permet de classer ces dernières en fonction de leur influence sur la sortie, ou plus concrètement connaître les entrées dont les variations ont le plus d'influence sur la variation de la sortie. Les entrées les plus influentes sont celles sur lesquelles l'incertitude doit être réduite en priorité si l'objectif est de réduire l'incertitude sur la sortie lors de la phase d'utilisation du modèle. Connaître les entrées les plus influentes permet de réduire l'incertitude sur la sortie du modèle et alors d'améliorer la qualité de la réponse de celui-ci, soit en cherchant à mieux décrire ces grandeurs, soit en adaptant la structure du modèle pour réduire l'effet des variations sur ces entrées. À l'inverse, les entrées les moins influentes peuvent être fixées à une valeur nominale, ce qui permet de simplifier le modèle coûteux en temps de calcul et ainsi d'obtenir un modèle plus léger avec moins d'entrées incertaines sans trop dégrader la réponse. Dans le cas d'un code informatique par exemple, il sera possible d'éliminer les parties du code qui n'ont aucune influence sur la valeur et la variabilité de la réponse. Parmi les différentes méthodes d'analyse de sensibilité globale, nous privilégierons la méthode reposant sur le calcul des indices de sensibilité de Sobol.

4.1.1.1 Indices de sensibilité de Sobol

Principe général

Plaçons-nous dans le cas d'une fonction vectorielle f de la forme :

$$y = f(x_1, ..., x_p)$$

Pour apprécier l'influence d'une variable d'entrée x_i sur la variation de la sortie y, l'idée est d'étudier comment la variance de y diminue lorsque la variable x_i est fixée à une valeur arbitrairement choisie x_i^* :

$$V(y|x_i = x_i^*) \tag{4.1}$$

Si cette diminution est importante par rapport à la variance totale $V(y)$, on en déduit que l'entrée x_i est influente sur la variation de y [125].
Pour l'instant, cet indice est local puisque sa valeur dépend du choix de la valeur x_i^* de x_i. Pour obtenir un comportement global, on considère l'espérance de cette quantité pour toutes les valeurs possibles de x_i^* :

$$E[V(y|x_i)] \tag{4.2}$$

Ainsi plus la variable x_i est importante vis-à-vis de la variance de y , plus cette quantité sera petite. En tenant compte de l'expression de la variance totale donnée par :

$$V(y) = V(E[y|x_i]) + E[V(y|x_i)],$$

nous pouvons utiliser de manière équivalente la quantité :

$$V(E[y|x_i])$$

qui sera plus grande lorsque la variable x_i est influente vis-à-vis de la variance de y [Jacques 2011[70]]. L'interprétation est la suivante. En fixant la variable x_i, la sortie moyenne du modèle en fonction de toutes les autres entrées est calculée. On réitère ensuite pour d'autres valeurs de x_i. Si la variation des sorties moyennes en fonction de x_i est importante, cela signifie que x_i est influente.
Afin d'utiliser un indice normalisé, l'indice de sensibilité du premier ordre qui quantifie la part de la variance de y due à la variable x_i est alors donné par l'expression suivante :

$$S_i = \frac{V(E[y|x_i])}{V(y)}. \tag{4.3}$$

Estimation de l'indice Sobol

Il existe différentes méthodes pour estimer cet indice, ainsi que ceux d'ordres supérieurs évaluant l'influence combinée de plusieurs entrées. Une manière classique d'estimer l'indice de Sobol du premier ordre S_i consiste à se munir d'un N-échantillon $X_{(N)} = (x_{k1}, ..., x_{kp})_{k=1,...N}$ de réalisations des variables d'entrée $(x_1, ..., x_p)$. L'espérance de y, $E[y] = f_0$, et sa variance, $V(y) = V$, sont estimées par :

$$f_0 = \frac{1}{N}\sum_{k=1}^{N} f(x_{k1}, ..., x_{kp}) \ , \ V = \frac{1}{N}\sum_{k=1}^{N} f^2(x_{k1}, ..., x_{kp}) - f_0^2 \tag{4.4}$$

L'estimation des indices de sensibilité de premier ordre consiste à estimer l'espérance de la variance conditionnelle :

$$V_i = V(E[y|x_i]) = E[E[y|x_i]^2] - E[E[y|x_i]]^2 = U_i - E[y]^2 \tag{4.5}$$

Sobol propose d'estimer la quantité U_i, c'est-à-dire l'espérance du carré de l'espérance de y conditionnellement à x_i, comme une espérance classique, mais en tenant compte du conditionement à x_i en faisant varier entre les deux appels à la fonction f toutes les variables sauf la variable x_i. Ceci nécessite deux échantillons de réalisations des variables d'entrée, que nous notons $X^1_{(N)}$ et $X^2_{(N)}$:

$$U_i = \frac{1}{N}\sum_{k=1}^{N} f(x^{(1)}_{k1}, ..., x^{(1)}_{k(i-1)}, x^{(1)}_{ki}, x^{(1)}_{k(i+1)}, ..., x^{(1)}_{kp}) \times f(x^{(2)}_{k1}, ..., x^{(2)}_{k(i-1)}, x^{(1)}_{ki}, x^{(2)}_{k(i+1)}, ..., x^{(2)}_{kp})$$

Les indices de sensibilité de premier ordre sont alors estimés par :

$$S_i = \frac{V_i}{V} = \frac{U_i - f_0^2}{V} \tag{4.6}$$

De même, les indices d'ordres supérieurs sont estimés en fixant un sous-ensemble des variables d'entrées et en retirant la part des variances estimées pour les ordres inférieurs. Par exemple l'indice du 2ième ordre est donné par :

$$S_{ij} = \frac{U_{ij} - f_0^2 - V_i - V_j}{V}$$

Cet indice S_{ij} exprime la sensibilité de la variance de y due à l'interaction des variables x_i et x_j (sans prendre en compte l'effet des variables seules).
Les indices totaux sont obtenus en fixant toutes les entrées sauf x_i. L'expression de l'indice d'ordre total est alors :

$$S_{Ti} = 1 - \frac{V(E[y|x_{\sim i}])}{V(y)} = 1 - \frac{V_{\sim i}}{V}$$

où nous notons $x_{\sim i}$ l'ensemble des entrées privé de x_i et où $V_{\sim i}$ est la variance de l'espérance de y conditionnellement à toutes les variables sauf x_i. $V_{\sim i}$ est alors estimée comme V_i, sauf qu'au lieu de faire varier toutes les variables sauf x_i, nous ne faisons varier uniquement que x_i. Ainsi, pour estimer

$$V_{\sim i} = E[E[y|x_{\sim i}]^2] - E[E[y|x_{\sim i}]]^2 = U_{\sim i} - E[y]^2, \tag{4.7}$$

on estime $U_{\sim i}$ par :

$$U_{\sim i} = \frac{1}{N}\sum_{k=1}^{N} f(x^{(1)}_{k1}, ..., x^{(1)}_{k(i-1)}, x^{(1)}_{ki}, x^{(1)}_{k(i+1)}, ..., x^{(1)}_{kp}) \times f(x^{(1)}_{k1}, ..., x^{(1)}_{k(i-1)}, x^{(2)}_{ki}, x^{(1)}_{k(i+1)}, ..., x^{(1)}_{kp})$$

et on obtient :

$$S_{Ti} = 1 - \frac{U_{\sim i} - f_0^2}{V} \tag{4.8}$$

La somme de tous les indices d'ordre 1 à p vaut théoriquement 1. La somme de tous les indices d'ordre 1 à p contenant une même variable x_i vaut théoriquement l'indice total correspondant. Au cours de ce travail de thèse, nous nous sommes principalement intéressés aux indices du premier ordre S_i.

Exemple 14. Nous considérons le modèle linéaire à entrées indépendantes suivant :

$$y = 9x_1 + 6x_2 + 3x_3 + x_4$$

avec les variables x_i qui suivent une loi uniforme sur $[0, 1]$.
Pour rappel, pour une loi uniforme sur un support borné $[x^-, x^+]$, l'écart type de l'entrée x vaut $\sigma = \sqrt{V(x)} = \dfrac{x^+ - x^-}{2\sqrt{3}}$.

Les indices de Sobol peuvent facilement être calculés analytiquement. Sachant que :
$V(x_i) = v = \dfrac{(1-0)^2}{12}$ et que $V(y) = (9^2 + 6^2 + 3^2 + 1)v = 127v$,
on en déduit que les 4 indices de Sobol définis par $S_i = \dfrac{V[E(y|x_i)]}{V(y)}$ valent respectivement :

$$S_1 = \frac{81v}{127v} = 0.64 \quad S_2 = \frac{36v}{127v} = 0.28 \quad S_3 = \frac{9v}{127v} = 0.07 \quad S_4 = \frac{1v}{127v} = 0.01.$$

Les indices de sensibilité du premier ordre en utilisant la méthode d'estimation expliquée précédemment donnent des résultats concordants (tableau 4.1).

variable	**indice d'ordre 1**
x_1	0.636
x_2	0.281
x_3	0.067
x_4	0.005

TABLE 4.1: Indices de sensibilité du premier ordre pour le modèle linéaire

L'entrée qui a le plus d'influence sur la variance de la sortie est la variable x_1.

4.1.1.2 Indicateur de sensibilité intervalle

Problématique

Un modèle intervalle est un modèle à part entière, différent d'un modèle probabiliste même si tous les deux peuvent reposer sur une expression analytique similaire. Dans le contexte de la propagation des incertitudes des entrées sur la sortie, le modèle intervalle

conduit à une stratégie d'évaluation du pire des cas alors que le modèle probabiliste prend en compte la distribution de la sortie. De plus, le phénomène de pessimisme introduit par le calcul par intervalles peut, en fonction de la fonction d'inclusion utilisée, conduire à rendre une variable multioccurrente plus ou moins influente. Or ce phénomène ne peut être pris en compte lors de l'estimation des indices de Sobol par une stratégie classique d'échantillonnage de type Monte Carlo. Pour cela se pose la question de savoir s'il est possible de construire un indice purement intervalle, dédié aux modèles intervalles, qui permettrait de mieux représenter leurs spécificités, notamment dans la prise en compte du pessimisme et la façon dont fonctionnent ces modèles.

Principe de l'indice de sensibilité intervalle

Considérons la fonction d'inclusion $[f]$ de la fonction f suivante :

$$[y] = [f]([x_1], ..., [x_p]) \supset \{y | y = f(x_1, ..., x_p) \wedge x_1 \in [x_1] ... \wedge x_p \in [x_p]\}.$$

Le point de départ suit le même raisonnement que pour l'indice de Sobol, à savoir fixer une des variables d'entrée x_i. D'un point de vue ensembliste, cela signifie laisser toutes les autres variables varier sur l'intégralité de leur supports intervalles respectifs tout en maintenant x_i contraint à appartenir à un sous-ensemble $[x_i]_j$ de $[x_i]$, sous-entendu $[x_i]_j$ est un intervalle suffisamment « petit » pour considérer que l'on fait peu varier x_i, sans être réduit à un unique point pour appréhender un grand nombre de situations d'un coup.
L'idée consiste ensuite à interpréter la quantité $V(y|x_i)$ (4.1) d'un point de vue ensembliste où la notion de variance va être remplacée par la notion de taille d'un intervalle pour représenter la variation de la sortie. On calcule donc :

$$[y]_{i,j} = [f]([x_1], ..., [x_{i-1}], [x_i]_j, [x_{i+1}], ..., [x_p]),$$

puis les largeurs des intervalles contenant y obtenus respectivement sur tout l'intervalle $[x_i]$ et sur le sous intervalle $[x_i]_j$:

$$w([y]) = y^+ - y^- \quad \text{et} \quad w([y]_{i,j}) = y^+_{i,j} - y^-_{i,j}$$

Plus $w([y]_{i,j})$ sera faible par rapport à $w([y])$, plus la variable d'entrée x_i est influente lorsqu'elle ne varie que sur $[x_i]_j$ au lieu de tout le support $[x_i]$.

Pour avoir une vision globale de la sensibilité de y par rapport à x_i, il suffit de réaliser cette opération sur un certain nombre de sous-intervalles $[x_i]_j$. On peut par exemple réaliser une sous-division uniforme de $[x_i]$ d'ordre m un entier strictement positif comme

présenté sur la figure (2.5). Une sous-division uniforme d'ordre m consiste à décomposer $[x_i]$ en sous intervalles $[x_i]_j$ disjoints, de même taille, dont l'union vaut $[x_i]$:

$$[x_i]_j = \left[x_i^- + (j-1)\frac{w([x_i])}{m}, x_i^- + j\frac{w([x_i])}{m} \right], j = 1, ..., m. \tag{4.9}$$

On peut évaluer $w([y]_{i,j})$ pour tous les intervalles de la sous-division uniforme pour obtenir une analyse complète, ou sur un nombre restreint en procédant à un tirage de N intervalles $[x_i]_j$ différents parmi m.

Finalement, sur le même principe que pour calculer $E[V(y|x_i)]$ (4.2), l'indice intervalle est normalisé en divisant par $w([y])$, en moyennant pour avoir une vision d'ensemble et en retranchant à la valeur 1 pour avoir un indice proche de 1 dans le cas d'une variable influente (proche de 0 dans le cas contraire) comme pour l'indice de Sobol :

$$S_i = 1 - \frac{\sum_{j=1}^{N} w([y]_{i,j})}{N w([y])}.$$

On peut modifier cet indice en remplaçant au dénominateur la taille de l'intervalle de sortie $w([y])$ calculé pour toutes les incertitudes possibles (c'est-à-dire en évaluant $[f]$ sur tout le support $[x_i]$) par la taille de l'intervalle obtenu en faisant l'union des $[y]_{i,j}$ (obtenus en évaluant $[f]$ sur tous les sous intervalles $[x_i]_j$). L'indice obtenu, noté S_i^U s'écrit :

$$S_i^U = 1 - \frac{\sum_{j=1}^{N} w([y]_{i,j})}{N w(\bigcup_{j=1}^{N} [y]_{i,j})}.$$

Remarque : il s'agit plus généralement du plus petit intervalle contenant cette union, celle-ci pouvant ne pas être un intervalle dans le cas où les $[x_i]_j$ sont tirés (présence de trous).

Il en résulte deux points intéressants qui peuvent être considérés comme des avantages à exploiter l'indice intervalle.
Le premier point tient de la différence entre les deux indices intervalles, qui vient du fait qu'en travaillant sur de petits intervalles $[x_i]_j$, le pessimisme dû à la variable x_i diminue (s'il existait). Comparer ces deux indices (ou de manière équivalente $w([y])$ et $w(\bigcup_{j=1}^{N} [y]_{i,j})$ donne donc une indication supplémentaire sur le pessimisme généré par cette variable. Plus S_i^U (respectivement $w(\bigcup_{j=1}^{N} [y]_{i,j})$ a diminué par rapport à S_i (respectivement $w([y])$), plus la variable x_i génère du pessimisme lors du calcul de $[y]$. Une évaluation de ce pessimisme consiste donc à calculer le rapport :

$$p_i = \frac{w([y])}{w(\bigcup_{j=1}^{N} [y]_{i,j})}, \tag{4.10}$$

plus grand que 1 dès qu'il y a du pessimisme généré par la variable x_i, égal à 1 sinon.

Un autre point est qu'on a à la fois une information globale (la valeur de l'indice), mais aussi locale sur la façon dont chaque entrée pèse sur la variation de la sortie en fonction des valeurs prises par cette entrée (provenant du fait que l'on découpe le support de chaque entrée). C'est intéressant dans une certaine mesure car cela donne de l'information sur comment fixer les variables les moins influentes.
Un dernier point important à préciser est que l'indicateur par intervalles consiste à tirer des sous-ensembles de valeurs plutôt que des points (comme pour l'indice de Sobol), ce qui permet d'appréhender plus de situations et d'estimer plus rapidement et plus précisément ces indices en réalisant moins de tirages.

Exemple 15. Nous considérons à nouveau le modèle linéaire à entrées indépendantes suivant :

$$y = 9x_1 + 6x_2 + 3x_3 + x_4$$

avec les variables x_i qui appartiennent au support intervalle [0,1].
Un intervalle $[x]$ peut s'écrire sous la forme : $[x] = [x^-, x^+] = mid([x]) + \frac{w([x])}{2} \times [-1, 1]$.
Dans le cas d'une sous-division d'ordre m, les sous-intervalles $[x_i]_j$ obtenus pour chaque découpage sont de taille $w([x_i]_j) = \frac{w([x_i])}{m} = \frac{1}{m}$ (principe d'une sous-division uniforme d'un intervalle $[0, 1]$ (4.9)). Puisque la largeur de l'intervalle $[x_1]_j$ est identique sur chaque découpage et que le modèle est linéaire et sans pessimisme par rapport à la variable x_1, les intervalles $[y]_{1,j}$ calculés pour la sortie ont tous la même largeur :

$$[y]_{1,j} = 9[x_1]_j + 10\left(\frac{1}{2} + \frac{[-1,1]}{2}\right) = \left(9.mid([x_1]_j) + \frac{10}{2}\right) + \left(\frac{9}{m} + 10\right)\frac{[-1,1]}{2}$$

d'où $w([y]_{1,j}) = \frac{9}{m} + 10$.
Sachant, en effectuant le même calcul, que $w([y]) = 19$, il est alors aisé de calculer théoriquement l'indice associé à la variable x_1 :

$$S_1 = 1 - \frac{m(\frac{9}{m} + 10)}{19m} = \frac{9(m-1)}{19m}$$

On voit ici l'influence de l'ordre m de la sous-division uniforme sur la valeur de l'indice. En effectuant le même travail sur les autres variables d'entrée et en supposant que m est suffisamment grand, les indices de sensibilité tendent respectivement vers les valeurs suivantes :

$$S_1 = \frac{9}{19} = 0.47 \quad S_2 = \frac{6}{19} = 0.32 \quad S_3 = \frac{3}{19} = 0.16 \quad S_4 = \frac{1}{19} = 0.05$$

L'estimation des 4 indices intervalles S_i (ou S_i^U de manière équivalente puisqu'il n'y pas de pessimisme) donne les résultats suivants pour une sous-division d'ordre $m = 100$:

Variable	**Indice de sensibilité intervalle**
x_1	0.47
x_2	0.31
x_3	0.16
x_4	0.05

TABLE 4.2: Indices de sensibilité intervalle pour le modèle linéaire

Notons que pour cet exemple particulier car linéaire par rapport aux entrées, la largeur $w([y]_{i,j})$ est constante quelque soit j, l'indice de sensibilité intervalle S_i est donc invariant suivant le sous-intervalle $[x_i]_j$ tiré et l'estimation des indices conduirait aux mêmes résultats que l'on tire $N = m$ sous-intervalles $[x_i]_j$ où que l'on en tire qu'une partie (voire un seul $N = 1$). Le classement est le même que pour Sobol, mais numériquement les valeurs obtenues des indices sont naturellement différentes. Même s'ils peuvent donner des tendances semblables, par construction les valeurs obtenues n'ont aucune raison d'être identiques. Les indices de Sobol se basent sur le moment d'ordre 2 (la variance) pour représenter la variation de la sortie alors que l'indice intervalle utilise la largeur d'intervalle : deux notions qui reflètent la même chose (c'est grand si la sortie varie beaucoup) sans être numériquement identiques.

4.1.1.3 Interprétation de l'indice de sensibilité intervalle

Prenons le modèle d'Ishigami un peu moins simple, non linéaire par rapport aux variables d'entrée et qui est un benchmark proposé dans [126] :

$$y = sin(x_1) + 7sin^2(x_2) + \frac{x_3^4}{10}sin(x_1)$$

avec les variables x_i qui suivent une loi uniforme sur $[-\pi, \pi]$.
Utilisons l'indice intervalle pour, en réalisant l'analyse de sensibilité, déduire des informations sur le pessimisme généré lors du calcul par intervalle. Pour ce faire, on s'intéresse à différentes formes d'inclusion.

Cas d'étude 1

Toutes les variables intervalles sont mono-occurentes, il n'y a donc aucun pessimisme lors de l'évaluation de ce modèle et c'est donc l'intervalle minimal $[y]$ qui est calculé :

$$[y] = [f_1]([\boldsymbol{x}]) = sin([x_1])(1 + \frac{[x_3]^4}{10}) + 7sin^2([x_2]).$$

Les deux indices S_i et S_i^U donnent naturellement les mêmes résultats.

Variable	Indice de sensibilité intervalle
x_1	0.54
x_2	0.24
x_3	0.55

TABLE 4.3: Indices de sensibilité intervalle pour le modèle d'Ishigami

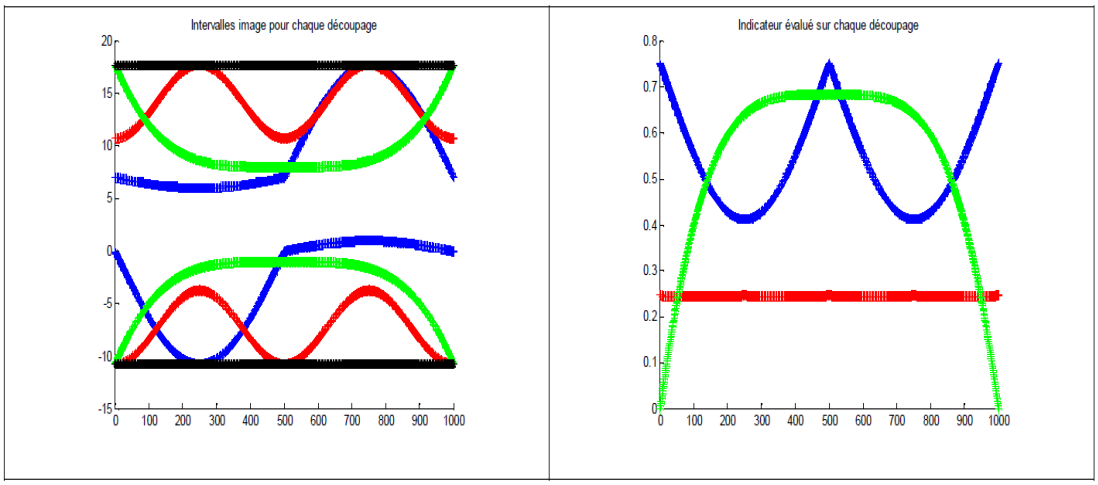

FIGURE 4.2: Intervalle image et indicateur évalué sur chaque découpage

Remarque. En abscisse est indiqué l'indice j de chaque sous-intervalle $[x_1]_j$. Comme la sous-division uniforme d'ordre $m = 1000$ est régulière, débute de la borne inférieure pour aller vers la borne supérieure, et que les sous-intervalle $[x_i]_j$ sont pris dans l'ordre (et non tirés aléatoirement ici), l'axe des abscisses correspond aussi à une image du découpage de l'intervalle $[x_i]$.

On observe cette fois-ci que les indices ne sont plus invariants selon le découpage, ce qui signifie que localement en fonction des valeurs prises par une variable d'entrée donnée, son influence sur la variation de la sortie change. Ainsi la variable x_3 est influente pour des valeurs prises au centre de son support intervalle. En effet, autour de 0, la largeur de l'intervalle $[y]_{3,j}$ calculé (*en vert*) est réduite de moitié par rapport à la largeur de $[y]$ (en noir) et donc le fait de fixer x_3 influence fortement la variation de y. Inversement, x_3 devient de moins en moins influente en se rapprochant des bornes $\pm\pi$ (valeurs pour lesquelles l'intervalle $[y]_{3,j}$ calculé est très proche de $[y]$ et où donc le fait de fixer x_3 n'influence pas la variation de y).

La valeur de l'indice S_3 correspond à moyenner cette interprétation sur tous les découpages. On observe alors que la variable x_1 (*en bleu*), tout en ayant localement une influence différente sur la sortie, a une influence globale similaire à x_3. Pour finir, la variable x_2 (*en rouge*) conduit à des intervalles $[y]_{2,j}$ proches de $[y]$ et donc à un indice relativement faible.

On a ainsi à la fois une information globale, mais aussi locale sur la façon dont chaque entrée pèse sur la variation de la sortie en fonction des valeurs prises par cette entrée. C'est intéressant dans une certaine mesure car cela donne de l'information sur comment fixer les variables les moins influentes. Dans le tableau suivant, sont présentés l'intervalle de sortie calculé et sa largeur lorsque l'une des entrées est fixée à une valeur nominale.

Cas	**Variable fixée**	$[y]$	$w([y])$
	Aucune	[-10.74,17.74]	28.48
(a)	x_1 fixée à 0	[0.00,7.00]	7.00
(c)	x_1 fixée à $\pi/2$	[1.00,17.74]	16.74
(b)	x_1 fixée à π	[0.00,7.00]	7.00
(a)	x_2 fixée à 0	[-10.74,10.74]	21.48
(a)	x_3 fixée à 0	[-1.00,8.00]	9.00
(d)	x_3 fixée à π	[-10.74,17.74]	28.48

TABLE 4.4: Intervalle de sortie et sa largeur quand certaines entrées sont fixées

On observe qu'en fixant une à une les 3 variables au centre (valeur nulle, cas (a)), on perd le moins de solutions pour la variable x_2 trouvée comme étant la moins influente, ce qui est cohérent. Vu l'indice de sensibilité pratiquement constant sur tout le support, on peut fixer x_2 n'importe où sur $[x_2]$. Pour ce qui concerne les deux autres variables, l'observation des figures précédentes (figures 4.2) montre que la variable x_1 est plus influente autour de $\{-\pi,0,\pi\}$ et moins autour de $\{-\pi/2,\pi/2\}$ et la variable x_3 est plus influente autour de 0 et moins autour de $\{-\pi,\pi\}$. Fixer une variable jugée globalement influente est une mauvaise idée à la base. Si on la fixe à une valeur où localement elle est influente, on perd effectivement beaucoup de solutions (jusqu'à 75% du support $[y]$, cas (b)). Si on la fixe à une valeur où localement elle est peu influente, on limite logiquement la perte de solution comme pour $x_1 = \pi/2$ (cas (c)). Pour la variable x_3, on tombe sur un cas vraiment particulier où l'indice est nul autour de $\{-\pi,\pi\}$, ce qui conduit de manière cohérente à ne perdre aucune solution si on fixe x_3 autour de ces deux valeurs (cas (d)). La conclusion est que l'indice intervalle donne des informations intéressantes sur comment fixer certaines entrées.

Les variables intervalles étant mono-occurrentes, l'indicateur de pessimisme p_i conduit naturellement aux résultats suivants :

Variable	**Indicateur de pessimisme**
x_1	1.00
x_2	1.00
x_3	1.00

TABLE 4.5: Indicateur de pessimisme pour le modèle d'Ishigami

Cas d'étude 2

On utilise une fonction d'inclusion où la variable x_2 est multi-occurrente, et où cette fois cette variable génère du pessimisme dans le calcul par intervalle :

$$[y] = [f_2]([\boldsymbol{x}]) = sin([x_1])(1 + \frac{[x_3]^4}{10}) + 7sin([x_2])sin([x_2]).$$

Les résultats pour l'indice S_i sont les suivants :

Variable	Indice de sensibilité intervalle
x_1	0.43
x_2	0.39
x_3	0.44

TABLE 4.6: Indices de sensibilité intervalle pour le modèle d'Ishigami

Les valeurs des indices ont logiquement changées par rapport au cas précédent (ne serait ce que parce que $[y]$ a lui-même changé). Le pessimisme généré par la variable x_2 la rend plus influente et elle est remontée pratiquement au même niveau que les deux autres (même si on a toujours la tendance $x_3 \approx x_1 > x_2$).
Le second indice S_i^U donne les résultats suivants :

Variable	Indice de sensibilité intervalle
x_1	0.43
x_2	0.24
x_3	0.44

TABLE 4.7: Indices de sensibilité intervalle pour le modèle d'Ishigami

On observe comme attendu que les indices S_i^U des variables ne générant pas de pessimisme sont identiques à S_i et plus faibles sinon comme c'est le cas pour la variable x_2 où on est passé de 0,39 à 0,24. L'indicateur de pessimisme p_2 confirme que la variable x_2 génère un intervalle $[y]$ en sortie 25% plus grand que l'intervalle minimal.

Variable	Indicateur de pessimisme
x_1	1.00
x_2	1.25
x_3	1.00

TABLE 4.8: Indicateur de pessimisme pour le modèle d'Ishigami

Graphiquement, on observe bien le pessimisme lié à la variable x_2 en comparant l'intervalle surestimé $[y]$ (en noir) et les intervalles $[y]_{2,j}$ en rouge dont l'union se rapprochant de l'intervalle minimal ne couvre pas $[y]$ (partie encadrée par l'ellipse).

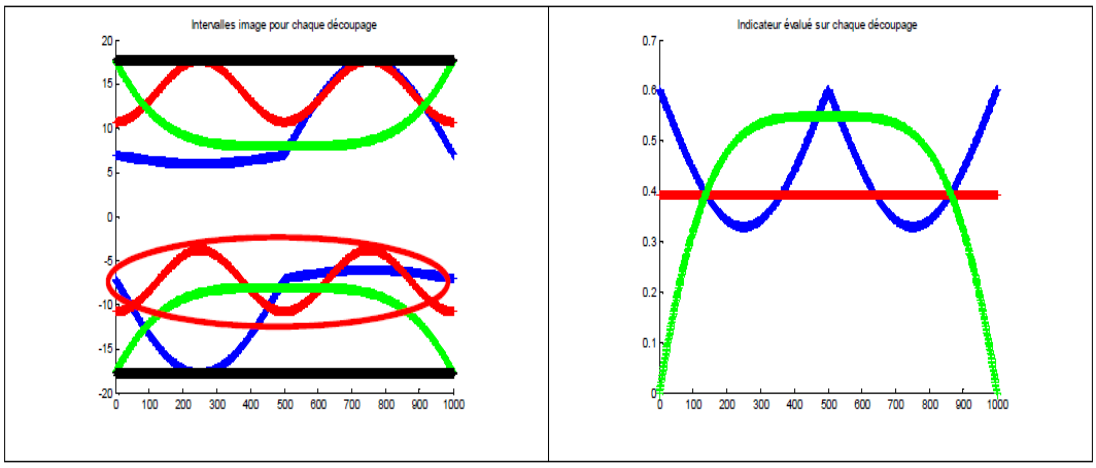

FIGURE 4.3: Intervalle image et indicateur évalué sur chaque découpage

Comparaison entre les indices par intervalles et Sobol

Dans cette partie nous analysons et comparons les résultats obtenus par les 2 indices en termes de classement des différentes entrées en fonction de leur influence dans le modèle d'Ishigami. Seules les tendances nous intéressent puisque l'on a déjà montré que numériquement, les indices par intervalles et de Sobol conduisaient à des valeurs numériquement différentes. L'estimation des 3 indices de Sobol du 1er ordre (4.6) et totaux (4.8) par la méthode expliquée précédemment donne les résultats suivants :

Variable	Indice de sensibilité d'ordre 1	Indice de sensibilité total
x_1	0.315	0.544
x_2	0.445	0.450
x_3	0.004	0.234

TABLE 4.9: Indices de sensibilité de Sobol pour le modèle d'Ishigami

Pour comparer, on utilise les indices de sensibilité intervalles calculés dans le cas d'étude 1 où toutes les variables étaient mono-occurentes (il n'y a donc pas de pessimisme dans les calculs par intervalles). On observe des tendances différentes, la variable x_3 étant la variable la plus influente par intervalle, et la moins influente avec Sobol (plus précisément, elle n'a pratiquement pas d'influence seule, mais elle a une influence en interaction avec x_1 qui reste cependant plus faible que pour les autres entrées).

Cette différence peut s'expliquer à l'aide des 2 points technique et théorique suivants. Techniquement, une première différence est la suivante. L'indice de Sobol consiste finalement à calculer une variance en évaluant un écart du terme U_i par rapport à la moyenne f_0 sur l'échantillon, donc un écart par rapport à un comportement moyen global (4.6). L'indice intervalle procède en calculant localement des largeurs de l'intervalle de sortie,

soit un écart par rapport au centre de chaque $[y]_{i,j}$ qui définit un comportement "moyen" cette fois-ci local.

Sur un plan théorique, la seconde différence tient dans la façon de modéliser les incertitudes et ensuite propager celles-ci dans un modèle. L'approche probabiliste permet la prise en compte de la distribution des entrées incertaines et de la sortie alors que l'approche ensembliste évalue le pire des cas (approche min/max).

Il s'ensuit que moyenner le pire des cas (indice intervalle) n'est forcément pas moyenner sur tous les tirages où on prend en compte la distribution (indice de Sobol), ce qui peut conduire à des classements certes différents concernant l'influence des différentes variables, mais cohérents avec la nature du modèle utilisé. Ainsi comme l'indiquent les figures 4.2 et le tableau 4.4, fixer x_3 de manière quelconque (c'est-à-dire ailleurs qu'autour de $\pm\pi$) conduira effectivement à considérablement réduire le support de sortie $[y]$ pour un modèle intervalle.

En général, le choix de l'indice de sensibilité utilisé est lié au type de modèle : l'indice Sobol pour le modèle probabiliste (analytique ou par code), l'indice intervalle pour le modèle intervalle avec S_i quand on travaille sur tout le support en un coup, S_i^U quand on travaille sur des sous-intervalles comme cela sera le cas pour la méthode mixte proposée pour la phase 2 dans la section 4.2.1.

4.1.2 Application de l'analyse de sensibilité

Les indices de Sobol et par intervalles présentés dans les sections précédentes sont utilisés pour réaliser l'analyse de sensibilité globale des modèles de dispersion utilisés.

4.1.2.1 Outils développés

Nous avons développé un outil informatique qui aide à réaliser l'analyse de sensibilité globale pour les modèles de dispersion atmosphérique analytique (gaussien) et pour le modèle complexe par code (SLAB).

Cet outil a nécessité un développement en trois étapes :

– La première étape a consisté à développer en java le code permettant de calculer les deux indices par intervalles et les indicateurs de pessimisme dans le cas où un modèle intervalle est utilisé, ainsi que d'estimer les indices de Sobol du premier ordre ou totaux dans le cas d'un modèle probabiliste (analytique ou par code).

- La deuxième étape a consisté à reprogrammer le logiciel SLAB qui est écrit en Fortran 77 en JAVA dans le but d'avoir plus de souplesse et de diminuer le temps nécessaire à une simulation du modèle pour permettre une approche par échantillonnage avec un temps de calculs raisonnable. Nous avons éliminé par exemple les multiples écritures des tableaux de résultats dans des fichiers textes qui ralentissaient énormément les temps d'exécution. De plus, nous avons implémenté le modèle gaussien en JAVA sous une forme adaptée au calcul par intervalles (fonction d'inclusion) et sous la forme probabiliste.
- La troisième étape a consisté à développer une passerelle (contrôleur) reliant les logiciels développés lors des étapes précédentes (modèles et calculs des indices de sensibilité) .

L'outil comprend trois phases principales illustrées sur la figure 4.4 :

Phase d'entrée :

- Définition du type de modèle (gaussien ou SLAB, probabiliste ou intervalle) pour déterminer le type d'indice à calculer,
- Définition des paramètres caractérisant les entrées incertaines du modèle (bornes inférieures et supérieures, lois de distribution (uniforme, normale)),
- Choix des paramètres d'échantillonnage (nombre d'échantillons N, ordre de la sous-division uniforme m).

Phase de traitement :

- Le système génère N échantillons selon une méthode d'échantillonnage aléatoire (avec la possibilité dans le cas intervalles de prendre tous les sous-intervalles de la division uniforme choisie, ou d'en tirer une partie),
- Pour chaque échantillon, les valeurs (ou supports) de la sortie sont calculées en fonction du type de modèle étudié.
- Ces valeurs (ou supports) sont ensuite utilisées pour calculer les indices de sensibilité (1er ordre, totaux,...) selon la méthode de Sobol ou les indices intervalles.

Phase de sortie :

- Les indices de sensibilité calculés sont renvoyés pour chaque entrée incertaine prise en compte.

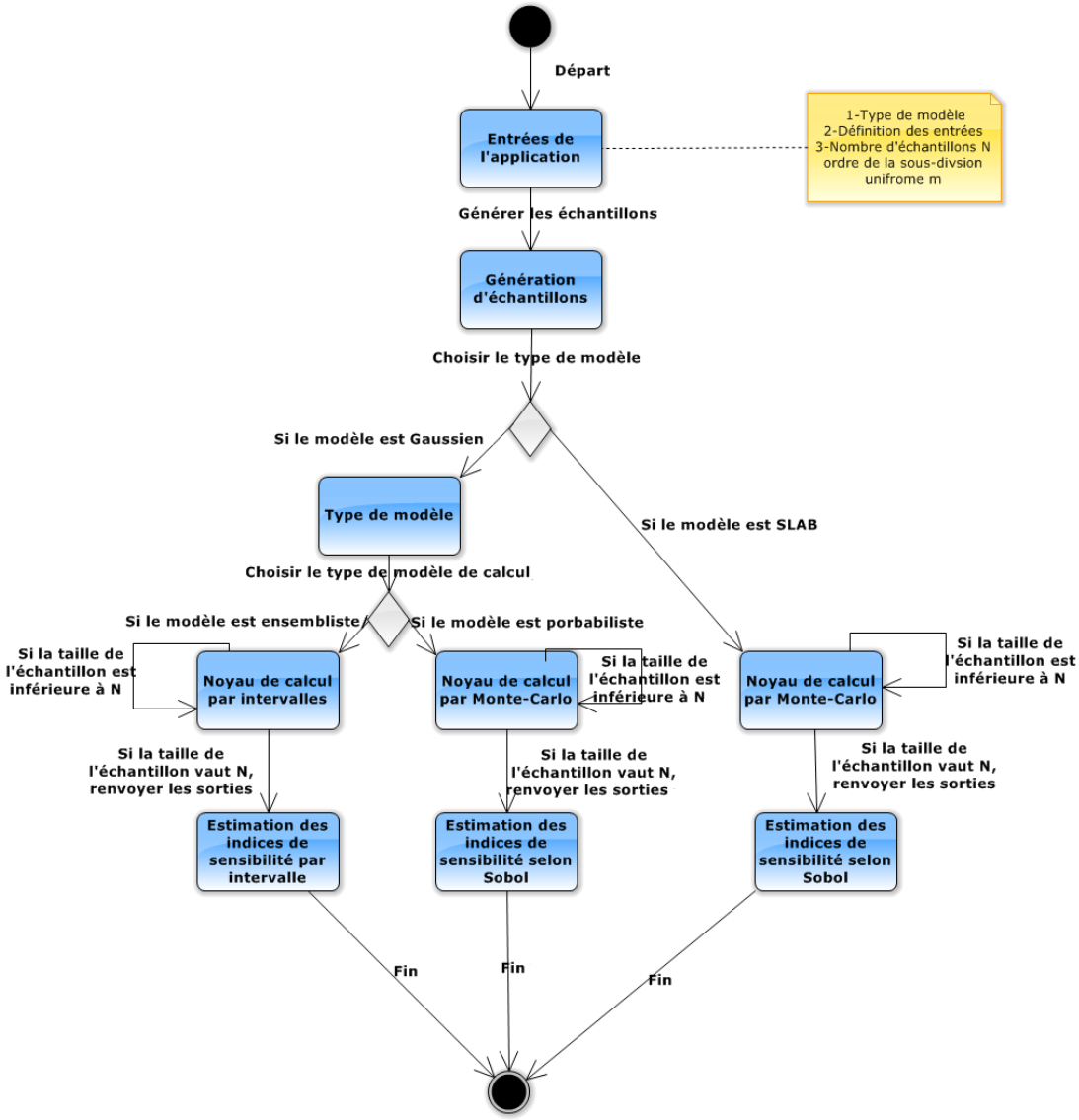

FIGURE 4.4: Fonctionnement de l'outil d'analyse de sensibilité développé

4.1.2.2 Analyse de sensibilité du modèle gaussien

Dans cette partie consacrée au modèle gaussien, on a utilisé les mêmes données que celles présentées dans l'application du chapitre 2 (2.2.1) concernant les valeurs nominales des entrées incertaines du modèle et les incertitudes associées.

Indices de sensibilité de Sobol

Le tableau 4.10 présente les résultats de l'analyse de sensibilité globale pour les six entrées incertaines du modèle gaussien étudié dans sa version probabiliste que sont le débit de fuite q, la vitesse du vent mesurée u_{ref}, et les paramètres de dispersion a_y, a_z, b_y et b_z.

Le but est de réduire le nombre d'entrées à prendre en compte afin de diminuer le temps de calcul.

L'indice de sensibilité du premier ordre S_i et l'indice de sensibilité totale S_{Ti} sont calculés puis moyennés en 5 points géographiques pour $N = 100.000$ échantillons.

Point (x, y, z)	Indice d'ordre 1						Indice total					
	a_y	a_z	b_y	b_z	u_{ref}	q	a_y	a_z	b_y	b_z	u_{ref}	q
$(40, 5, 2)$	0.33	0.02	-0.01	-0.02	0.12	0.46	0.35	0.04	0.01	0.00	0.14	0.48
$(350, 8, 2)$	0.11	0.14	0.01	0.01	0.18	0.59	0.10	0.13	0.00	0.00	0.17	0.58
$(460, 35, 2)$	0.04	0.14	0.00	0.00	0.17	0.61	0.04	0.15	0.00	0.00	0.18	0.62
$(600, 50, 2)$	0.11	0.13	0.00	-0.01	0.16	0.56	0.12	0.14	0.01	0.00	0.17	0.57
$(800, 45, 2)$	0.00	0.16	0.00	0.01	0.19	0.64	0.00	0.16	0.00	0.01	0.19	0.64
Moyennes	**0.12**	**0.12**	**0.00**	**0.00**	**0.16**	**0.57**	**0.12**	**0.13**	**0.00**	**0.00**	**0.17**	**0.58**

TABLE 4.10: Indices de sensibilité du premier ordre et totaux pour le modèle gaussien

L'interprétation des résultats est la suivante :

- Les entrées qui ont une forte influence sur les sorties sont : q, u_{ref}, a_y et a_z.
- Les entrées qui ont une influence moindre sur les sorties et qui seront par la suite fixées à une valeur nominale sont b_y et b_z.

On peut noter que si b_y et b_z sont systématiquement moins influentes que les autres variables quel que soit le point géographique, les tendances entre les entrées restantes changent. Le débit de fuite et la vitesse du vent ont une influence a peu près équivalente partout alors que c'est très fluctuant pour les paramètres de dispersion a_y et a_z. Par exemple, le paramètre de dispersion a_y est très influent sur le premier point et n'a quasiment aucune influence sur le dernier point.

Indice de sensibilité intervalle

Les indices de sensibilité par intervalle S_i pour le modèle gaussien dans sa version intervalle (2.23)-(2.29) sont calculés puis moyennés pour les mêmes points et avec une sous-division d'ordre $m = 1000$:

Point (x, y, z)	Indice de sensibilité intervalle					
	a_y	a_z	b_y	b_z	u_{ref}	q
$(40, 5, 2)$	0.42	0.17	0.06	0.02	0.12	0.23
$(350, 8, 2)$	0.20	0.19	0.04	0.04	0.18	0.36
$(460, 35, 2)$	0.34	0.14	0.08	0.03	0.14	0.27
$(600, 50, 2)$	0.37	0.13	0.09	0.03	0.13	0.26
$(800, 45, 2)$	0.29	0.15	0.07	0.04	0.15	0.30
Moyennes	**0.32**	**0.16**	**0.07**	**0.03**	**0.14**	**0.28**

TABLE 4.11: Indices de sensibilité par intervalle pour le modèle gaussien

Les résultats obtenus montrent que les paramètres b_y et b_z sont clairement les moins

influents comme pour Sobol. Les tendances pour les autres entrées sont similaires : le débit q est devant, puis la vitesse du vent u_{ref} et le paramètres a_z très proche. La grosse différence vient du paramètre a_y qui remonte dans le haut du classement au même niveau que q.
Un autre essai est réalisé pour calculer les indices de sensibilité par intervalle S_i pour les mêmes points, mais maintenant l'ordre de la sous-division uniforme n'est seulement que de $m = 10$ et on ne tire que la moitié $N = 5$ des sous-intervalles obtenus.

Point (x, y, z)	Indice de sensibilité intervalle					
	a_y	a_z	b_y	b_z	u_{ref}	q
Moyennes des 5 points	**0.29**	**0.14**	**0.06**	**0.03**	**0.13**	**0.26**

TABLE 4.12: Indices de sensibilité par intervalle pour le modèle gaussien

Les résultats indiqués dans le tableau 4.12 montrent que les indices changent peu par rapport à ceux calculés pour $m = 1000$ (tableau 4.11). Ceci vient du fait de tirer des intervalles plutôt que des points, et nous pouvons constater qu'avec m beaucoup plus faible, les résultats restent raisonnables, ce qui permet de réduire le temps de calcul nécessaire à l'analyse de sensibilité.
De manière générale, les paramètres a_y et a_z ont une influence relative plus importante avec l'indice intervalle (surtout a_y), ce qui est logique dans le mesure où ce sont des variables multi-occurrentes générant du pessimisme (plus de 30% pour la plupart des points) comme en témoignent les indicateurs de pessimisme calculés dans le tableau 4.13.

Paramètres	Point (x, y, z)					**Moyennes**
	$(40, 5, 2)$	$(350, 8, 2)$	$(460, 35, 2)$	$(600, 50, 2)$	$(800, 45, 2)$	
a_y	1.30	1.05	1.38	1.35	1.37	**1.29**
a_z	1.12	1.01	1.01	1.00	1.00	**1.03**
b_y	1.03	1.00	1.07	1.06	1.07	**1.05**
b_z	1.01	1.00	1.00	1.00	1.00	**1.00**
u_{ref}	1.00	1.00	1.00	1.00	1.00	**1.00**
q	1.00	1.00	1.00	1.00	1.00	**1.00**

TABLE 4.13: Indicateurs de pessimisme associés aux différentes entrées incertaines

Propagation des incertitudes après l'analyse de sensibilité

L'objectif est de refaire l'étude de propagation d'incertitudes pour le modèle de dispersion gaussien effectuée dans la section 2.2.3 du chapitre 2, en prenant en compte cette fois-ci les résultats de l'analyse de sensibilité effectuée précédemment, pour en vérifier la pertinence. A cette fin, l'incertitude sur b_y et b_z a été supprimée, en d'autres termes ces entrées du modèle sont fixées à leurs valeurs nominales.
Toutes les autres entrées du modèle q, a_y, a_z et u_{ref} sont considérées comme incertaines et

peuvent varier respectivement sur leurs supports bornés. La propagation des incertitudes est calculée pour $N = 10000$ échantillons.

Le tableau 4.14 rappelle les valeurs nominales des entrées du modèle étudié avec les incertitudes ajoutées seulement sur q, a_y, a_z et u_{ref}.

$q \pm 5\%$	$a_y \pm 2.5\%$	$a_z \pm 2.5\%$	b_y	b_z	$u_{ref} \pm 2.5\%$	x	y
2216	0.105	0.066	0.915	0.915	4.58	40	5
						350	8
						460	35
						600	50
						800	45

TABLE 4.14: Entrées du modèle

Le tableau 4.15 donne pour les 5 points étudiés la concentration $c_{Nom,k}$ calculée à partir des valeurs nominales des entrées et les intervalles de confiance associés en utilisant les résultats de l'analyse de sensibilité. Sont aussi donnés les indicateurs de la propagation de l'incertitude $(U-MC, U-AI1, U-AI2)$ calculés respectivement par Monte Carlo (2.2), par l'analyse par intervalles (2.20) en utilisant la fonction d'inclusion naturelle (2.23-2.29) et la fonction d'inclusion à test de monotonie (2.30-2.35).

№	$Point(x,y,z)$	$c_{Nom,k}$	$[y_{Min}, y_{Max}]_{MC}$	$[y^-, y^+]_{AI1}$	$[y^-, y^+]_{AI2}$	$U-MC$	$U-AI1$	$U-AI2$
1	(40, 5, 2)	10.46	[9.25,11.75]	[8.52,12.77]	[9.15,11.88]	11.95 %	20.35%	13.02 %
2	(350, 8, 2)	1.22	[1.10,1.36]	[1.08,1.39]	[1.08,1.38]	10.91 %	12.96%	12.16%
3	(460, 35, 2)	0.379	[0.34,0.42]	[0.32,0.45]	[0.33,0.42]	10.20 %	16.34%	11.18%
4	(600, 50, 2)	0.193	[0.17,0.22]	[0.16,0.23]	[0.17,0.22]	11.10 %	17.27%	12.20%
5	(800, 45, 2)	0.187	[0.17,0.21]	[0.16 ,0.22]	[0.16,0.21]	9.55 %	14.80%	10.24%

TABLE 4.15: Les intervalles de confiance et les indicateurs calculés

La figure 4.5 représente la propagation des incertitudes avec la méthode de Monte Carlo et la méthode de l'analyse par intervalles en utilisant la fonction d'inclusion naturelle et la fonction d'inclusion à test de monotonie pour les 5 points de coordonnées (x_k, y_k, z_k) après l'analyse de sensibilité.

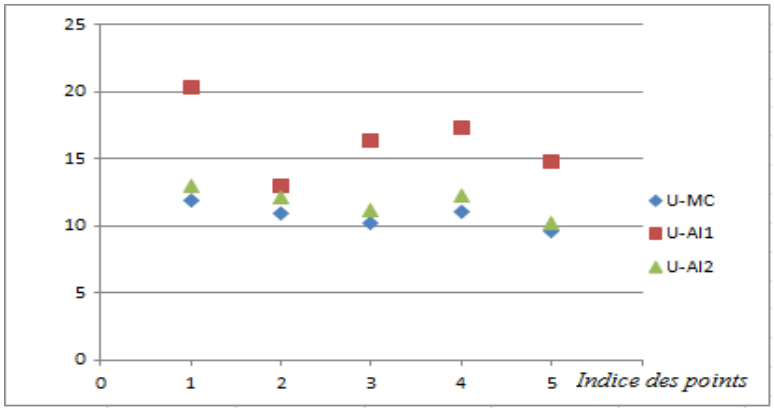

FIGURE 4.5: Propagation de l'incertitude selon les points étudiés

Interprétation des résultats

Avec la simulation de MC, les résultats obtenus montrent que l'incertitude sur la sortie du modèle varie entre 9.55% et 11.95% pour les différents points étudiés. Avec la méthode AI, selon la fonction d'inclusion naturelle l'indicateur varie entre 12.96% et 20.35% et selon la fonction d'inclusion à test de monotonie l'indicateur varie entre 10.24% et 13.02%. Avec AI, le temps de calcul pour estimer la concentration en un point donné est égal à 2.5 ms. Avec la simulation de MC le temps d'exécution est de 8.3 ms.
Le tableau 4.16 compare les temps d'exécution pour calculer la concentration en un point donné avant et après l'analyse de sensibilité pour les 2 méthodes (Monte Carlo et l'analyse par intervalles).

Temps de calcul	MC	AI
Avant l'analyse de sensibilité	85ms	3.5ms
Aprés l'analyse de sensibilité	8.3ms	2.5ms

TABLE 4.16: Temps de calcul avant et après l'analyse de sensibilité

L'analyse de sensibilité aide à fixer les entrées les moins influentes à des valeurs constantes. Cela réduit le temps de calculs qui conduit à un traitement plus rapide. Rappelons pour point de comparaison d'après les tableaux 2.4 et 2.5, que si toutes les entrées incertaines sont prises en compte, les indicateurs varient respectivement entre 10.42% et 11.99% pour $U-MC$, entre 14.34% et 22.10% pour $U-AI1$ et entre 10.91% et 13.76% pour $U-AI2$. Ainsi bien que certaines valeurs solutions de la sortie soient naturellement perdues pour la méthode Monte Carlo, l'analyse de sensibilité réalisée est pertinente parce que cette perte est raisonnable. Pour la méthode de l'analyse par intervalles selon la fonction d'inclusion naturelle, le nombre de solutions perdues est supérieur à celui de la méthode de Monte Carlo, en effet la réduction de l'incertitude sur la sortie du modèle est plus importante. Ces résultats sont attendus parce que certaines des entrées multi-occurrentes (dans notre cas b_y et b_z) sont fixées à des valeurs constantes et nominales. En conséquence, cela conduit à une réduction du phénomène de dépendance entre les grandeurs incertaines de modèle, ce qui réduit le pessimisme de la méthode de l'analyse par intervalles et conduit à des résultats plus précis. Pour terminer, la fonction d'inclusion à test de monotonie conduit avant et après l'analyse de sensibilité à l'intervalle minimal contenant la sortie. Le support obtenu après analyse de sensibilité est systématiquement plus petit que celui estimé avant analyse de sensibilité puisque certaines entrées incertaines ont été fixées.

Réalisation des cartographies après l'analyse de sensibilité par Sobol

L'objectif est de refaire la réalisation des cartographies pour le modèle de dispersion gaussien effectuée dans la section 3.4.1.1 du chapitre 3 par l'approche probabiliste, en prenant en compte cette fois-ci les résultats de l'analyse de sensibilité effectuée précédemment, pour en vérifier la pertinence. A cette fin, comme dans la section précédente, l'incertitude sur b_y et b_z a été supprimée, en d'autres termes ces entrées du modèle sont fixées à leurs valeurs nominales.
Toutes les autres entrées du modèle q, a_y, a_z et u_{ref} sont considérées comme incertaines et peuvent varier respectivement sur leurs supports bornés. La propagation des incertitudes est calculée pour $N = 10000$ échantillons.

La figures 4.6 illustre la fusion des résultats de la méthode de Monte Carlo avant et après l'analyse de sensibilité.

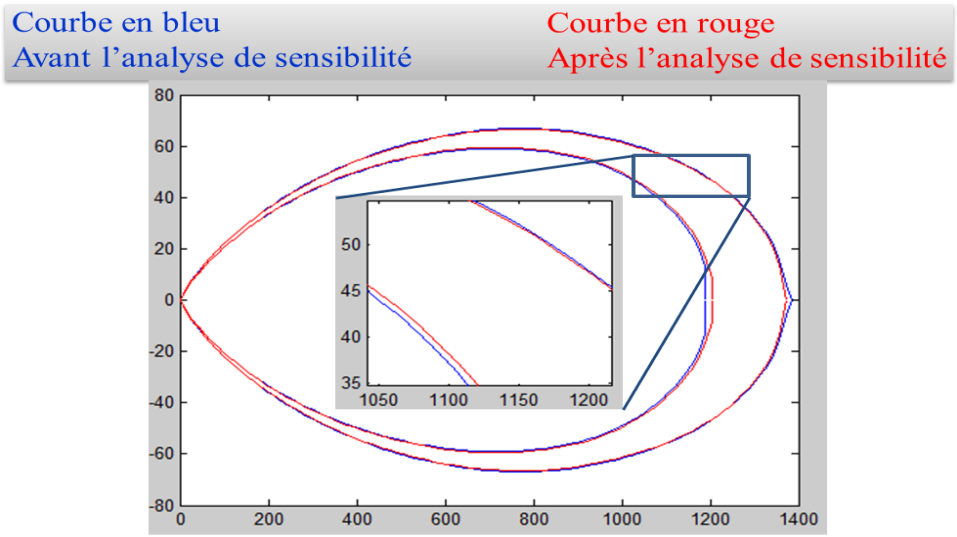

FIGURE 4.6: Panache de NO simulé par l'inversion probabiliste avant et après l'analyse de sensibilité

De cette comparaison, nous pouvons conclure que les résultats sont très similaires avant et après l'analyse de sensibilité. Avant l'analyse de sensibilité avec 100.000 échantillons le temps de calculs est de 513.52s, après l'analyse de sensibilité avec 10.000 échantillons celui-ci n'est plus que de 56.52s, soit une réduction de 90%. Nous pouvons conclure que l'analyse de sensibilité donne des résultats pertinents pour la méthode probabiliste, car après avoir fixé les entrées les moins influentes, la perte de précision sur les cartographies reste raisonnable avec un temps de calcul bien plus faible.

Réalisation des cartographies après l'analyse de sensibilité par intervalles

Les figures 4.7 et 4.8 illustrent respectivement la fusion des cartographies réalisées par l'approche ensembliste (détaillée dans la section 3.4.1.2 du chapitre 3) avant et après l'analyse de sensibilité pour l'approximation extérieure et l'approximation intérieure. La comparaison montre que l'approximation extérieure S_{ext} calculée avant l'analyse de sensibilité est un peu plus grande que celle obtenue après la l'analyse de sensibilité. L'approximation intérieure S_{int} est elle légèrement plus petite avant qu'après. Ces résultats sont tout à fait logiques. En effet S_{ext} est une approximation de l'ensemble $S_{u,\exists p}$ qui augmente en taille lorsque les incertitudes sur $\boldsymbol{p}$ sont plus importantes (plus de valeurs de $\boldsymbol{p}$ conduisent à plus de valeurs possibles pour la sortie). A l'inverse, S_{int} est une approximation de $S_{u,\forall p}$, plus de valeurs de $\boldsymbol{p}$ signifie des contraintes plus difficiles à satisfaire (pour tous les $\boldsymbol{p}$) et donc un ensemble dont la taille diminue lorsque les incertitudes sur $\boldsymbol{p}$ sont plus importantes.

Nous pouvons en déduire que l'analyse de sensibilité effectuée est pertinente pour l'analyse par intervalles parce que les zones de danger et d'incertitude évaluées pour les deux cas sont très proches. En outre, elle permet de réduire, dans une certaine mesure, le temps de calcul : avant l'analyse de sensibilité le nombre de boîtes traitées par la méthode d'inversion ensembliste est de 6709 en 53.36 s, alors qu' après l'analyse de sensibilité le nombre de boîtes traitées est de 6305 en 50.23 s, conduisant à une réduction de presque 6%.

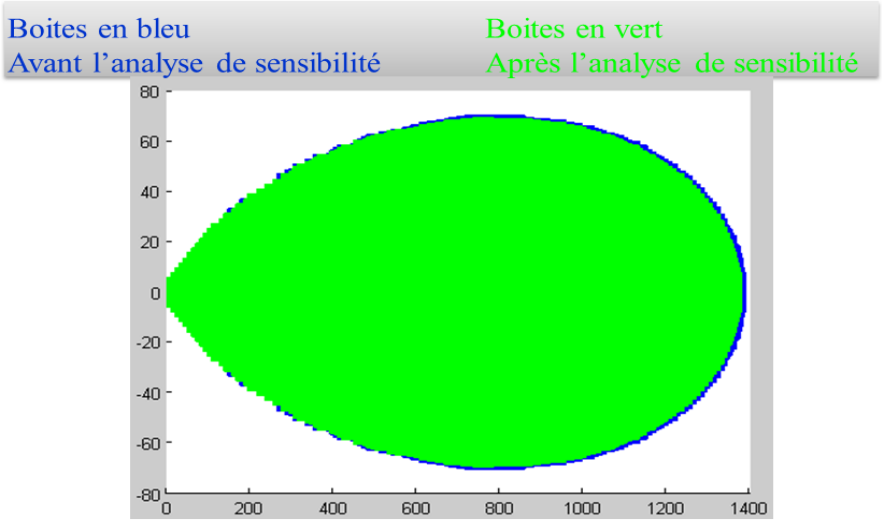

FIGURE 4.7: Panache de NO simulé par l'inversion ensembliste S_{ext}

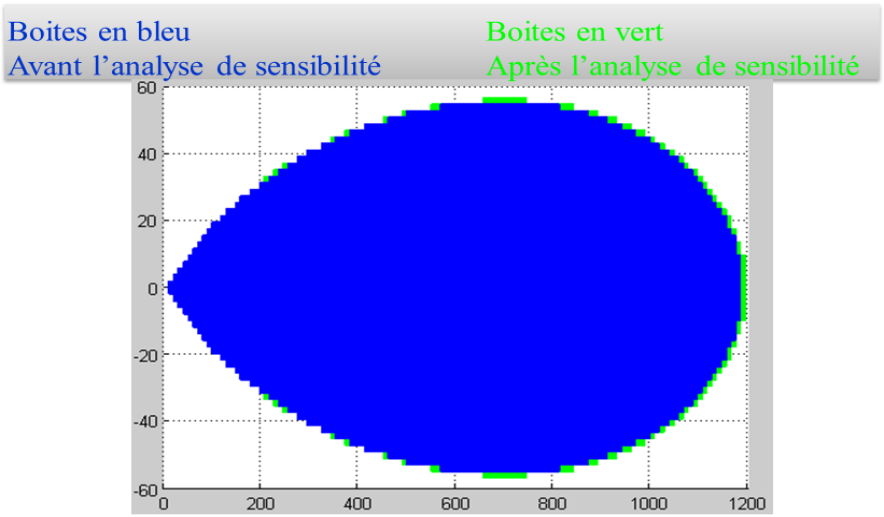

FIGURE 4.8: Panache de NO simulé par l'inversion ensembliste S_{int}

4.1.2.3 Analyse de sensibilité du modèle complexe SLAB

Les travaux portant sur l'analyse de sensibilité d'outils de modélisation de la dispersion atmosphérique de rejets accidentels restent peu nombreux. Parmi ces travaux, une analyse de sensibilité d'un modèle de dispersion atmosphérique néerlandais, appelé NPK-PUFF, pour des scénarios de rejets de contaminants radioactifs a été développée par (Kok et al. (2004)[19]).

Une autre analyse de sensibilité dédiée à la dispersion d'un polluant radioactif en utilisant le code RIMPUFF qui permet de simuler en temps réel des rejets de gaz sous forme de bouffées ou de panaches en prenant en compte les conditions météorologiques a été effectuée par Ferenczi [127] selon la méthode OAT présentée brièvement en section 1.4.4.1.

Bubbico et Mazzarotta (2008)[128] ont appliqué la méthode OAT pour simuler des scénarios de rejets accidentels de produits toxiques (acide chlorhydrique, ammoniac, triméthylamine et brome) pour une durée du rejet égale à 15 minutes.

Notre objectif est ici de réaliser une analyse de sensibilité sur le modèle par code SLAB. Durant les travaux de thèse de S. Pagnon [3] sur la stratégie de modélisation des conséquences d'une dispersion atmosphérique de gaz toxique ou inflammable en situation d'urgence au regard de l'incertitude sur les données d'entrée, une analyse de sensibilité a été menée sur le modèle de dispersion atmosphérique SLAB. Cette analyse est basée sur la méthode de screening dite de Morris couplée avec une analyse de sensibilité locale (présentée en section 1.4.4.2) afin de confirmer (ou infirmer) les résultats obtenus au moyen de la méthode de Morris. Les résultats obtenus par cette étude montrent que :

- certaines entrées comme la vitesse du vent ont une influence décroissante sur les distances d'effets,
- la température et le débit de fuite ont une influence croissante monotone sur l'estimation des distances d'effets,
- d'autres entrées incertaines influent très faiblement sur l'évaluation des distances d'effets, c'est le cas de la hauteur de rejet,
- finalement l'étude de sensibilité locale a confirmé l'existence d'entrées ayant une influence complexe sur le résultat final comme la rugosité.

Globalement, l'analyse de sensibilité d'un modèle est effectuée essentiellement dans le but d'identifier les entrées du modèle les plus influentes sur la sortie. L'analyse de sensibilité locale est relativement simple et rapide, mais elle peut s'avérer insuffisante pour caractériser la sensibilité de modèles complexes tels que SLAB. Nous avons donc retenu pour notre étude les méthodes d'analyse de sensibilité globale basées sur l'étude de la variance en calculant les indices de sensibilité de Sobol.

Démarche adoptée

Une démarche a été mise en œuvre pour effectuer l'analyse de sensibilité globale de SLAB. Comme nous l'avons dit dans la description des modèles de dispersion atmosphérique, SLAB est un modèle de type intégral, qui est capable de traiter différents types de produits pour modéliser les phénomènes liés à la dispersion du gaz. Comme nous avons choisi pour l'analyse de sensibilité du modèle gaussien l'émission d'oxyde nitrique (NO), nous avons choisi pour SLAB un gaz différent, liquéfié tel que l'ammoniac (NH3) qui est plus lourd que l'air près du point de fuite. Les caractéristiques de ce produit et les entrées utilisées par SLAB ont été présentées dans la section 3.4.3.

Compte tenu du nombre important d'entrées, une trentaine initialement, dans un premier temps, nous avons écarté les entrées parfaitement connues (comme par exemple celles liées aux propriétés chimiques du produit), puis parmi les entrées incertaines restantes, un premier tri a été effectué par expertise pour éliminer les entrées connues comme étant peu influentes sur la variation de la sortie. En se basant sur les résultats obtenus par S. Pagnon [3] et d'autres études sur SLAB, nous n'avons donc conservé que les entrées jugées comme influentes et c'est sur ces dernières, au nombre de 5 (débit de fuite, vitesse du vent, humidité, rugosité, température) qu'est réalisée l'analyse de sensibilité globale.

La figure 4.9 illustre la démarche utilisée pour faire l'analyse de sensibilité globale de SLAB.

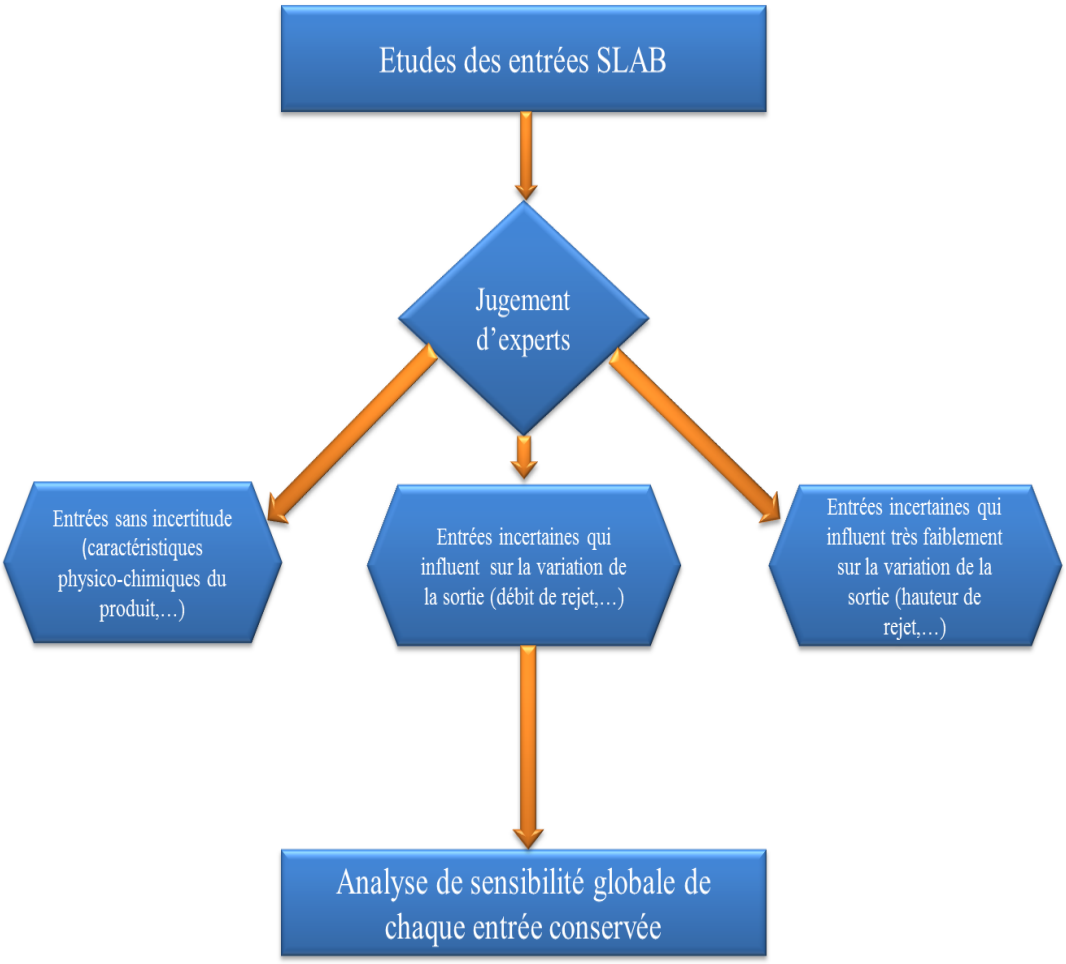

FIGURE 4.9: Démarche utilisée pour l'analyse de sensibilité de SLAB

Résultats de l'étude de l'analyse de sensibilité globale

Les sorties du modèle SLAB utilisées pour réaliser l'analyse de sensibilité correspondent à des concentrations en gaz estimées le long de la ligne de vent. Les sorties relatives à l'ammoniac (NH3) sont calculées à une hauteur $Z = 0$. En ce qui concerne la taille de l'échantillon, on note que les indices de sensibilité diffèrent considérablement pour des échantillons de taille inférieure à $N = 100.000$. Nous avons donc retenu cette valeur pour obtenir un niveau de confiance acceptable, ce qui représente un ordre de grandeur classique avec 5 entrées incertaines à étudier.

Pour le produit NH3, nous avons choisi trois types de sortie :

- $C280$: la concentration évaluée à 280 m dans la direction du vent en ppm,
- $C1415$: la concentration évaluée à 1415 m dans la direction du vent en ppm,
- $C3588$: la concentration évaluée à 3588 m dans la direction du vent en ppm.

La figure 4.10 donne les indices de sensibilité de Sobol du premier ordre et totaux pour la concentration $C280$.

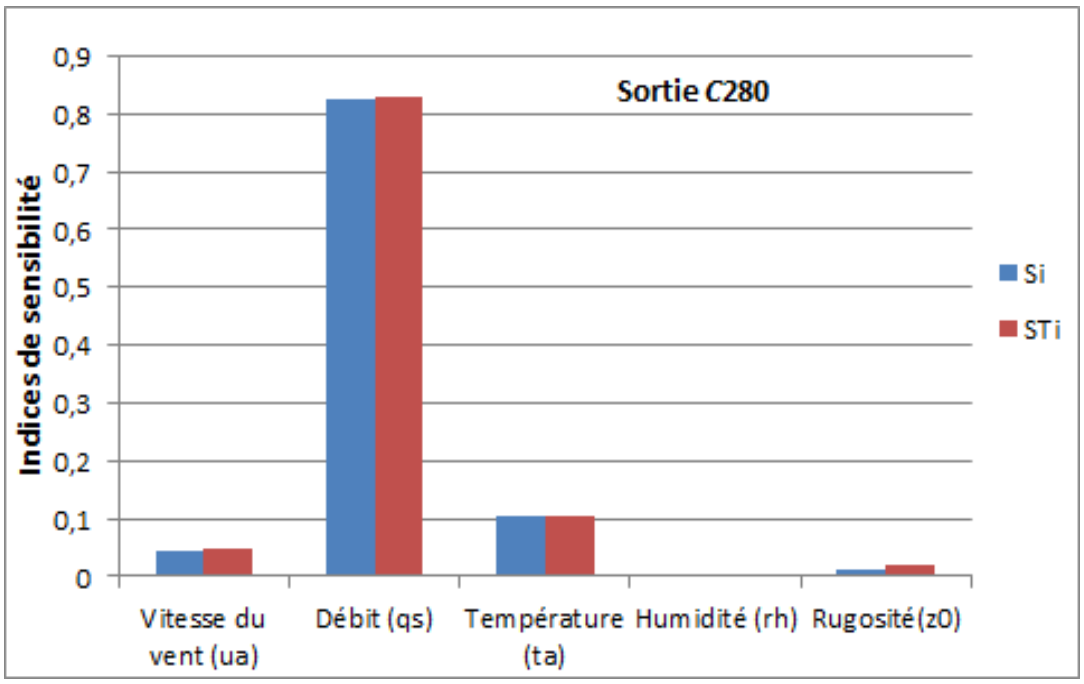

FIGURE 4.10: Indices de sensibilité pour la concentration $C280$

Les figures 4.11 et 4.12 illustrent les indices de sensibilité pour les concentrations en gaz $C1415$ et $C3588$.

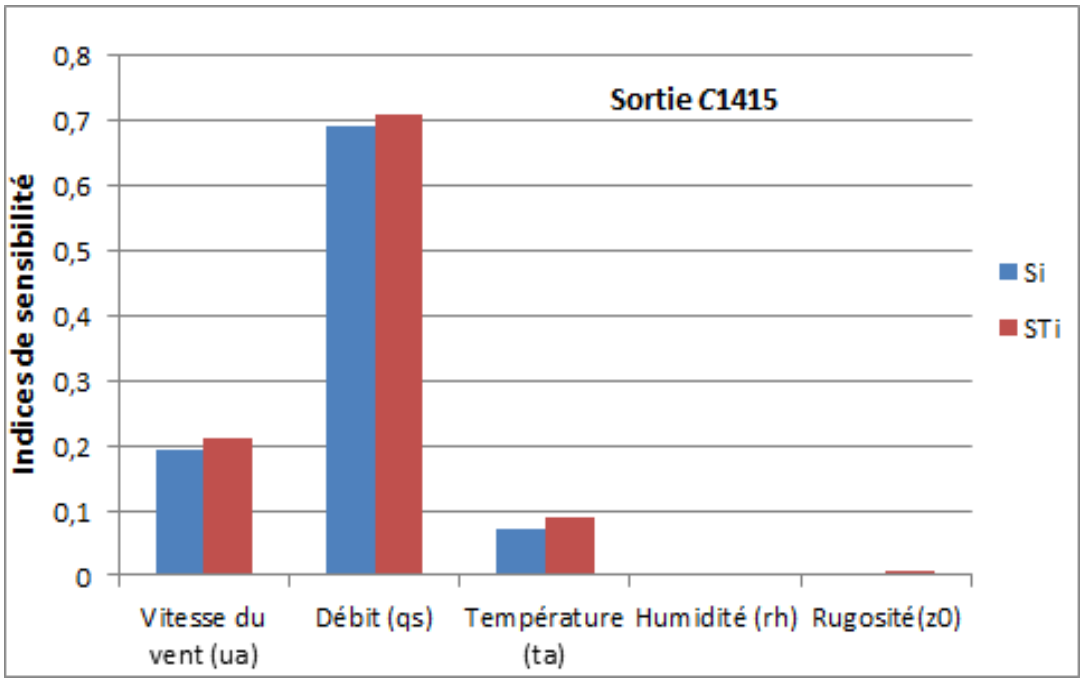

FIGURE 4.11: Indices de sensibilité pour la concentration $C1415$

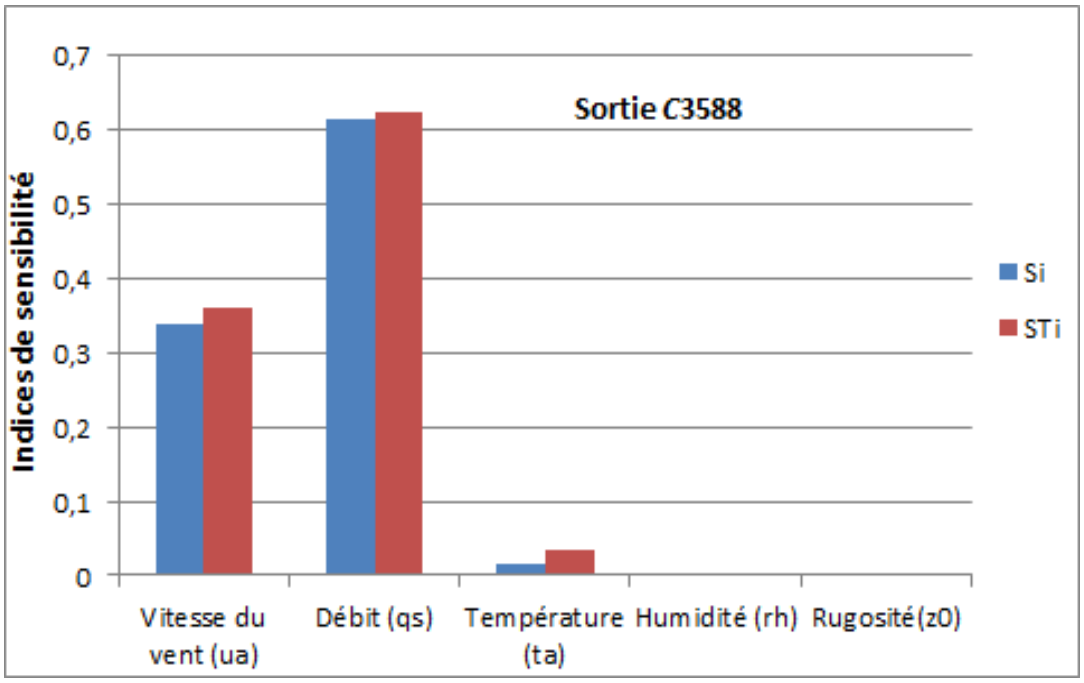

FIGURE 4.12: Indices de sensibilité pour la concentration $C3588$

En observant les résultats obtenus sur les 3 figures (4.10, 4.11, 4.12), on constate que les paramètres les plus influents sur les sorties dans l'ordre décroissant sont :
– le débit de rejet (qs),
– la vitesse du vent (ua),
– la température (ta).
On peut noter que la rugosité ($z0$) et l'humidité (rh) sont systématiquement moins influentes que les autres variables quel que soit le point géographique étudié.

Propagation des incertitudes avant et après l'analyse de sensibilité

L'objectif est de faire l'étude de la propagation d'incertitudes pour le modèle complexe par code SLAB avant et après l'analyse de sensibilité, pour en vérifier la pertinence.

Dans cette partie applicative, la concentration en gaz est estimée pour les 3 points placés sous le vent et déjà utilisés pour l'analyse sensibilité ($C280$, $C1415$ et $C3588$).

Pour ces 3 points, les intervalles de confiance des concentrations en gaz calculés, ainsi que les indicateurs représentatifs de l'amplitude des incertitudes sur ces concentrations sont donnés en utilisant la méthode de Monte Carlo. Le nombre d'échantillons utilisés pour la simulation de Monte-Carlo avant l'analyse de sensibilité est fixé, comme justifié dans la partie précédente, à $N = 100.000$.

Les caractéristiques des entrées utilisées par SLAB et les incertitudes associées ont été présentées dans la section 3.4.3.
Le tableau 4.17 donne pour les 3 points étudiés la concentration $c_{Nom,k}$ calculée à partir des valeurs nominales des entrées, l'intervalle de confiance et l'indicateur de la propagation de l'incertitude $U-MC$ calculés par Monte Carlo (2.2) avant l'analyse de sensibilité.

№	*Point*	$c_{Nom,k}$	$[y_{Min}, y_{Max}]$	$U-MC$
1	$C280$	21917.77	[19600.84,24346.46]	10.82%
2	$C1415$	2942.55	[2604.53,3275.84]	11.40%
3	$C3588$	980.51	[840.86,1135.71]	15.03%

TABLE 4.17: Les intervalles de confiance et les indicateurs calculés

En se basant sur les résultats obtenus par l'analyse de sensibilité, l'incertitude sur $z0$ et rh a été supprimée, en d'autres termes ces entrées du modèle sont fixées à leurs valeurs nominales.
Toutes les autres entrées du modèle qs, ua et ta sont considérées comme incertaines et peuvent varier respectivement sur leurs supports bornés. La propagation des incertitudes est calculée pour $N = 10000$ échantillons.

Le tableau 4.18 donne sur les 3 points étudiés la concentration $c_{Nom,k}$ calculée à partir des valeurs nominales des entrées, l'intervalle de confiance et l'indicateur de la propagation de l'incertitude $U-MC$ calculés par Monte Carlo (2.2) après l'analyse de sensibilité.

№	*Point*	$c_{Nom,k}$	$[y_{Min}, y_{Max}]$	$U-MC$
1	$C280$	21917.77	[19841.96,24081.22]	9.67%
2	$C1415$	2942.55	[2620.14,3279.84]	11.20 %
3	$C3588$	980.51	[848.52, 1137.80]	14.75%

TABLE 4.18: Les intervalles de confiance et les indicateurs calculés

La comparaison des tableaux 4.17 et 4.18 montre que les résultats sont très proches avant et après l'analyse de sensibilité. Avant l'analyse de sensibilité avec 100.000 échantillons le temps de calculs pour un point est de 241s, après l'analyse de sensibilité avec 10.000 échantillons le temps de calculs est de 24s, soit une réduction de 90%. Nous pouvons conclure que l'analyse de sensibilité est pertinente pour la méthode probabiliste, puisqu'après avoir fixé les entrées les moins influentes, la précision des intervalles de confiance estimés en ces points reste raisonnable avec un temps de calcul moindre.

Réalisation des cartographies après l'analyse de sensibilité

En se basant sur les résultats de l'analyse de sensibilité, l'humidité (rh) et la rugosité ($z0$) ont été fixées à leurs valeurs nominales. Les autres entrées de modèle (ua, qs, ta) sont considérées comme incertaines et peuvent varier respectivement sur leurs supports bornées. La cartographie est réalisée en utilisant l'approche probabiliste pour un modèle complexe par code détaillée dans la section 3.2.2 avec $N = 1000$ échantillons.
La figure 4.13 illustre la fusion des résultats de la méthode Monte Carlo avant et après l'analyse de sensibilité. Avant l'analyse de sensibilité avec 100.000 échantillons le temps de calculs est de 250s, après l'analyse de sensibilité avec 1000 échantillons le temps de calculs est de 8s, soit une réduction de 96.8%.

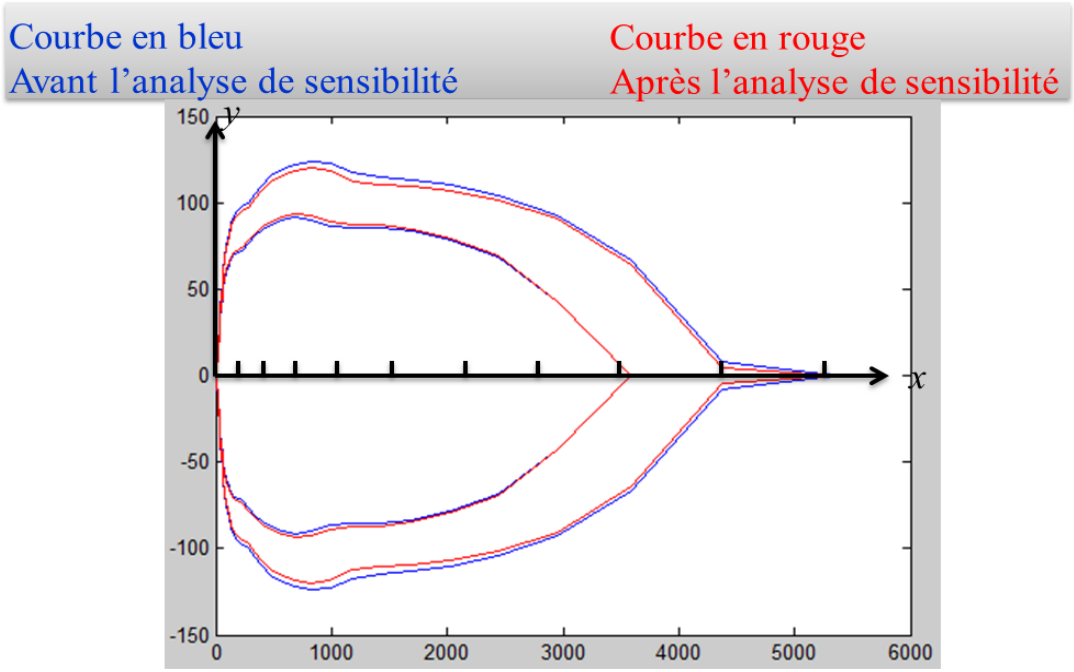

FIGURE 4.13: Panache de NH3 simulé par l'inversion probabiliste

De cette comparaison, nous pouvons conclure que la précision des résultats reste raisonnable avec un temps de calcul bien moindre après avoir fixé les entrées les moins influentes.

4.1.3 Test de monotonie

4.1.3.1 Objectif de l'analyse de monotonie

Dans la section 2.1.2.2 du chapitre 2 présentant les règles du calcul par intervalles, des fonctions d'inclusion particulières permettant de limiter le pessimisme de l'intervalle calculé en sortie du modèle ont été présentées. L'une d'entre elle appelée fonction d'inclusion

à test de monotonie a été détaillée (2.17)-(2.19). Il s'agissait d'exploiter la monotonie par morceaux du modèle par intervalles par rapport à chacune des entrées pour remplacer, quand cela est possible, le calcul classique par intervalles par un calcul de type inf/sup (c'est-à-dire une double évaluation du modèle pour des arguments réels bien choisis), qui lui ne génère pas de pessimisme.
Même si l'objectif est ici différent, le principe reste similaire. Le test de monotonie utilisé dans la phase 1 (figure 4.1) de la démarche proposée est une technique visant à évaluer le support intervalle de la sortie d'un modèle, dont il est difficile (voire impossible) de construire une fonction d'inclusion. C'est typiquement le cas d'un modèle par code tel que SLAB dont on ne connait pas précisément l'expression analytique. Cette technique consiste à vérifier si la fonction ou bien le modèle étudié est monotone par rapport à une entrée incertaine donnée x_i. Si le modèle est monotone sur tout le domaine de variation $[x_i]$ de x_i alors on sait que les bornes inférieure et supérieure de la sortie seront obtenues lorsque x_i prendra comme valeur l'une de ses propres bornes sur $[x_i]$. Si le modèle est monotone par rapport à chacune des entrées, il est alors possible de calculer le support intervalle de la sortie, et donc de faire de manière indirecte une évaluation de type ensembliste, sans utiliser l'analyse par intervalle et donc sans pessimisme. Il y a peu de chance pour des modèles complexes que la propriété de monotonie soit atteinte pour toutes les entrées sur l'intégralité de leurs supports respectifs. La méthode mixte proposée dans la phase 2 ajoutera de la flexibilité sur ce point :

- en travaillant sur de petits sous-intervalles de $[x_i]$ où le modèle peut être monotone par morceaux,
- en permettant de tirer des points plutôt que des supports intervalles pour les entrées où on ne peut pas prouver la monotonie (l'idée étant de privilégier la manipulation des ensembles plutôt que des points, quand cela est possible, pour appréhender un maximum de situations).

4.1.3.2 Algorithme de vérification

Cet algorithme est basé sur le balayage à pas constant de l'intervalle d'étude et par la comparaison systématique de deux images « successives » ou plutôt du signe de leur différence. Si ce signe ne change jamais, la fonction est peut-être monotone, sinon elle ne l'est certainement pas.

Considérons la fonction vectorielle f de la forme :

$$y = f(x_1, ..., x_i, ..., x_p)$$

Entrées

Soient :

- l'intervalle $[x_i] = [x_i^-, x_i^+]$ l'ensemble des valeurs possibles pour une entrée incertaine x_i. Les autres entrées incertaines x_j, avec j différent de i sont fixées à leurs valeurs nominales,
- f, le modèle à étudier,
- N, le nombre de sous-divisions de l'intervalle $[x_i]$,
- u, les valeurs successives de la variable x_i.

Initialisation

- le pas **prend la valeur** $(x_i^+ - x_i^-)/N$
- le sens **prend le signe de la différence** $f(x_1, ..., x_i^+, ..., x_p) - f(x_1, ..., x_i^-, ..., x_p)$
- u **prend la valeur** x_i^-

Traitement

Pour K **variant de 1 à** N

Si $(f(x_1, ..., u+pas, ..., x_p)$-$f(x_1, ..., u, ..., x_p))\times$ signe<0 alors f n'est pas monotone, fin de traitement

Sinon $u = u + pas$;

Si $k == N$ alors f est monotone, fin de traitement

4.1.4 Étude de la multi-occurrence des entrées incertaines

L'existence d'une multi-occurrence des entrées incertaines dans un modèle analytique par intervalles peut conduire (mais pas forcement) à du pessimisme dans l'évaluation du support intervalle de la sortie comme on a pu l'expliquer dans la section 2.1.2.6. Il est donc légitime d'étudier la multi-occurrence des entrées incertaines pour réduire la surestimation des intervalles calculés, en utilisant des fonctions d'inclusion judicieusement choisies en essayant par exemple de réduire au maximum les occurrences multiples, en sélectionnant des fonctions d'inclusion comme les formes centrées bénéficiant de bonnes propriétés pour réduire le pessimisme, ou comme le proposera la méthode mixte développée dans la phase 2 en travaillant sur de plus petits supports intervalles des entrées incertaines.

La première possibilité pour réaliser cette étude, et qui est aussi la plus simple, est qualitative et consiste à repérer l'occurrence des différentes variables intervalles dans l'expression formelle de la fonction d'inclusion choisie.

Une autre possibilité, quantitative et plus précise celle-ci, pour savoir si une entrée donnée génère du pessimisme, est d'utiliser l'indicateur de pessimisme (4.10) qui permet justement d'évaluer le niveau de pessimisme engendré par la multi-occurrence d'une variable intervalle.

4.2 Deuxième phase : proposition d'une nouvelle approche - Méthode mixte

Plusieurs approches ont été développées pour réaliser la propagation des incertitudes sur la sortie d'un modèle. Parmi ces approches, on a présenté et utilisé l'analyse par intervalles et la simulation de Monte Carlo au cours du chapitre 2. La première méthode utilisable dès que l'on a à disposition une fonction d'inclusion intervalle du modèle, permet de calculer un intervalle surestimé contenant l'ensemble des valeurs possibles de la sortie compte tenu des incertitudes sur les entrées. Cette notion de garantie provient du pessimisme généré par le calcul par intervalles qui induit la propriété d'inclusion (2.6). La seconde méthode permet de calculer un intervalle sous-estimé contenant une partie de toutes les valeurs possibles de la sortie du modèle, qui lui peut se présenter aussi bien sous la forme d'une expression analytique que par code informatique. Ceci provient du nombre fini N d'échantillons aléatoires tirés. Toutes les valeurs d'une entrée donnée x_i ne peuvent pas être prises en compte. En outre, le N-échantillon ne peut pas représenter toutes les combinaisons possibles des valeurs d'entrée du modèle. Ainsi, selon le nombre d'incertitudes et la largeur de leurs supports respectifs, une simulation de Monte Carlo pourrait impliquer des milliers ou des dizaines de milliers de simulations de modèles.

En conclusion, l'analyse par intervalles et la simulation de Monte carlo sont des méthodes de propagation d'incertitudes qui ont chacune leurs avantages et inconvénients, apportent même des informations complémentaires dans le sens où elle vont permettre d'encadrer le support exact de la sortie, reposent à la base sur des mécanismes d'évaluation différents et induisent des contraintes spécifiques sur la nature du modèle qui peut être traité.
L'idée, très simple et naturelle à la base, est d'essayer de développer une nouvelle méthode de propagation d'incertitudes, partant de la méthode de Monte-Carlo, plus classique et surtout permettant de traiter tout type de modèle, et de la combiner avec de l'analyse par intervalles dans l'objectif d'appréhender plus de situations. L'objectif visé est ainsi de bénéficier des avantages cumulés des deux méthodes, et à précision équivalente d'utiliser des échantillons de taille réduite pour limiter les temps de calculs par rapport à une simulation de Monte-Carlo classique, ou inversement pour des échantillons de même taille de gagner en précision sur le support de sortie évalué.

4.2.1 Principes de l'approche proposée

Dans cette section, les principes de la méthode proposée mixant simulation de Monte-Carlo et analyse par intervalles sont développés. Cette méthode s'appelle plus précisément Advanced Interval based Monte Carlo Method (AIMCM), mais nous utiliserons

aussi dans la suite le terme de méthode mixte pour abréger. Le principe de la méthode proposée, basée sur une stratégie d'échantillonnage, est de tirer des supports intervalles aléatoires pour les entrées du modèle à la place de valeurs aléatoires afin d'appréhender pour un tirage donné un ensemble de valeurs possibles de la sortie en lieu et place d'une simple valeur estimée.

Cette méthode va être déclinée en deux versions, en fonction du type de modèle utilisé. La première version concerne le cas d'un modèle analytique (par exemple le modèle de dispersion gaussien) dont on a à disposition une fonction d'inclusion, condition préalable pour pouvoir faire du calcul par intervalles. Le principe de fonctionnement de l'AIMCM est alors le suivant :

1. Définir la sortie et les éléments d'entrée du modèle étudié.
2. Associer un support intervalle pour chaque entrée du modèle $x_i \in [x_i] = [x_i^-, x_i^+]_{,i=1,...,p}$.
3. Réaliser une sous-division uniforme de chaque intervalle $[x_i]$ en m sous-intervalles disjoints $[x_i]_j$ dont l'union couvre $[x_i]$, c'est-à-dire :
   $[x_i] = \cup[x_i]_{j,j=1,..,m}$ avec $[x_i]_j = [x_i^- + (j-1)\frac{w([x_i])}{m}, x_i^- + j\frac{w([x_i])}{m}]$,
   où $w([x_i])$ désigne la largeur de l'intervalle $[x_i]$.
4. Soit N le nombre d'échantillons. Le k^{eme} échantillon est créé en choisissant aléatoirement pour chaque entrée du modèle x_i un sous-intervalle $[x_i]_j$ parmi les supports intervalles associés à $[x_i]$. Ensuite l'intervalle $[y]_k$ est calculé en sortie du modèle pour cet échantillon en utilisant l'analyse par intervalles.

   Générer une matrice contenant les résultats de tous les échantillons étudiés :

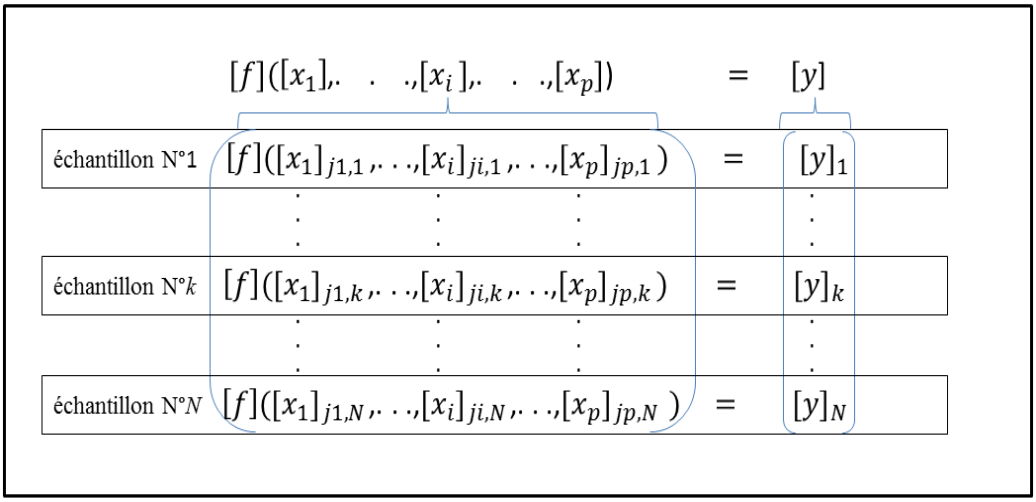

FIGURE 4.14: Matrice d'échantillons

   avec $k = 1, ..., N$ et $j1, k, ..., jp, k = 1, ..., m$.
5. Calculer l'intervalle de confiance $[y]$ comme l'union $\bigcup[y]_{k,k=1,...,N}$.

Implémentation de l'approche proposée

La figure 4.15 présente la phase de calculs de l'évaluation de l'incertitude en utilisant la méthode mixte (AIMCM) pour mettre en oeuvre la propagation d'incertitudes.

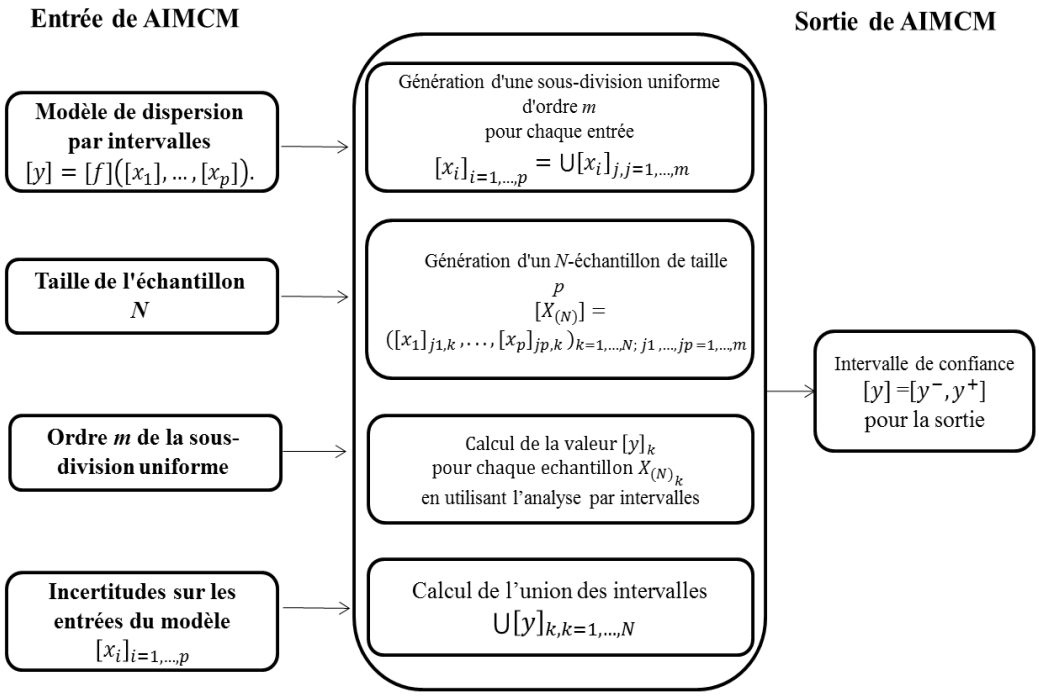

FIGURE 4.15: Phase de calculs de la propagation d'incertitudes

À partir des résultats calculés par la méthode mixte (AIMCM), l'indicateur de la propagation de l'incertitude u est défini par la même relation que l'équation (2.20). Les termes y^- et y^+ sont les bornes inférieure et supérieure de l'intervalle calculé $[y^-, y^+]$.

Traiter le problème de propagation des incertitudes par la méthode de l'analyse par intervalles développée dans le chapitre 2 (section 2.1.2.7) est un cas particulier de la méthode mixte consistant à prendre une sous-division uniforme d'ordre $m = 1$. L'avantage de prendre un ordre plus élevé, comme discuté dans la section 2.1.2.6, est de travailler sur des supports intervalles $[x_i]_j$ plus petits que les supports $[x_i]$ complets, ce qui limite le pessimisme sur les intervalles $[y]_k$ (et donc leur union) calculés en sortie. L'étude de la multi-occurrence des variables lors de la phase 1 peut justement permettre d'orienter le choix de m. L'inconvénient est en contre partie un temps de calcul plus lourd puisque N évaluations du modèle par intervalles sont nécessaires et qu'il n'est maintenant plus garanti (à moins de tirer toutes les combinaisons possibles d'intervalles $[x_i]_j$) que toutes les valeurs possibles de la sortie appartiendront bien à $[y^-, y^+]$.

Nous allons maintenant nous focaliser sur la seconde déclinaison de la méthode mixte qui s'adresse plus particulièrement aux modèles dont on n'a pas de fonction d'inclusion, comme le logiciel SLAB, et pour lesquels on ne peut pas faire de calculs par intervalles. Notons qu'un modèle analytique tel que le modèle de dispersion gaussien peut, dans sa version probabiliste, s'inscrire aussi dans cette catégorie si on ne souhaite pas utiliser sa fonction d'inclusion. Pour résoudre cette difficulté, il est nécessaire de tester la monotonie du modèle utilisé par rapport à chacune de ses entrées incertaines comme proposé dans la phase 1. Si le modèle est monotone sur tout le domaine de variation de x_i alors on sait que les bornes inférieure et supérieure de la sortie seront obtenues lorsque x_i prendra comme valeur l'une de ses propres bornes sur $[x_i]$. Si le modèle est monotone par rapport à chacune des entrées, il est alors possible de calculer le support intervalle de la sortie, et donc de faire de manière détournée une évaluation de type ensembliste, sans utiliser l'analyse par intervalles et donc sans pessimisme.
Si le modèle n'est pas monotone par rapport à une entrée x_i, la solution retenue est de tirer des points plutôt que des supports intervalles pour cette grandeur sur le même principe que la méthode de Monte-Carlo. L'idée consiste toutefois à privilégier la manipulation des ensembles plutôt que des points pour appréhender un maximum de situations lors d'une simulation.

Dans ce qui suit, pour un modèle de dispersion $y = f(x_1, ..., x_p)$ ayant p entrées (f représente ici le modèle par code), on définit :
- I_{MC} et I_{MD} deux sous-ensembles regroupant les indices $i = 1, ..., p$ des entrées par rapport auxquelles le modèle est monotone croissant, respectivement décroissant,
- I_{NM} l'ensemble des indices des entrées x_i par rapport auxquelles la propriété de monotonie ne peut être retenue.

Le principe de fonctionnement de l'AIMCM est alors le suivant :

1. Définir la sortie et les éléments d'entrée du modèle étudié.
2. Associer un support intervalle pour chaque entrée x_i, i élément de I_{MC} ou de I_{MD} : $x_i \in [x_i] = [x_i^-, x_i^+]_{i \in I_{MC} \vee i \in I_{MD}}$.
3. Associer une loi de distribution à chaque entrée incertaine x_i, i élément de I_{NM}.
4. Préciser le nombre d'échantillons N qui peut être choisi arbitrairement en fonction du nombre d'entrées incertaines considérées et du temps de calcul associé à une simulation du modèle (dans notre étude, ce nombre sera souvent de l'ordre de 10^q, où q désigne le nombre d'éléments de l'ensemble I_{NM}).
5. Le $k$ième échantillon est crée de la manière suivante :
   - pour chaque entrée du modèle x_i, i élément de I_{NM}, est choisie une valeur aléatoire $x_{i,k}$ en fonction de la loi de distribution associée,

– pour chaque entrée x_i, i élément de I_{MC} ou I_{MD}, est tiré aléatoirement un support noté $[x_i]_{ji,k} = [(x_i^-)_{ji,k}, (x_i^+)_{ji,k}]$ correspondant à l'un des sous-intervalles $[x_i]_{j=1,...,m}$ de la sous-division uniforme d'ordre m de $[x_i]$, où $(x_i^-)_{ji,k}$ et $(x_i^+)_{ji,k}$ désignent respectivement les bornes inférieure et supérieure de $[x_i]_{ji,k}$.

6. Pour ce kième échantillon, une double évaluation du modèle f est réalisée afin de déterminer le support de la sortie $[y]_k$ de la façon suivante :
$[y]_k = [f(u_{1,k}, ..., u_{p,k}), f(v_{1,k}, ..., v_{p,k})]$
où

$$(u_{i,k}, v_{i,k}) = \begin{cases} ((x_i^-)_{ji,k}, (x_i^+)_{ji,k}) & \text{si} \quad i \in I_{MC} \\ ((x_i^+)_{ji,k}, (x_i^-)_{ji,k}) & \text{si} \quad i \in I_{MD} \\ (x_{i,k}, x_{i,k}) & \text{sinon} \end{cases}$$

Générer une matrice contenant les résultats de tous les échantillons étudiés :

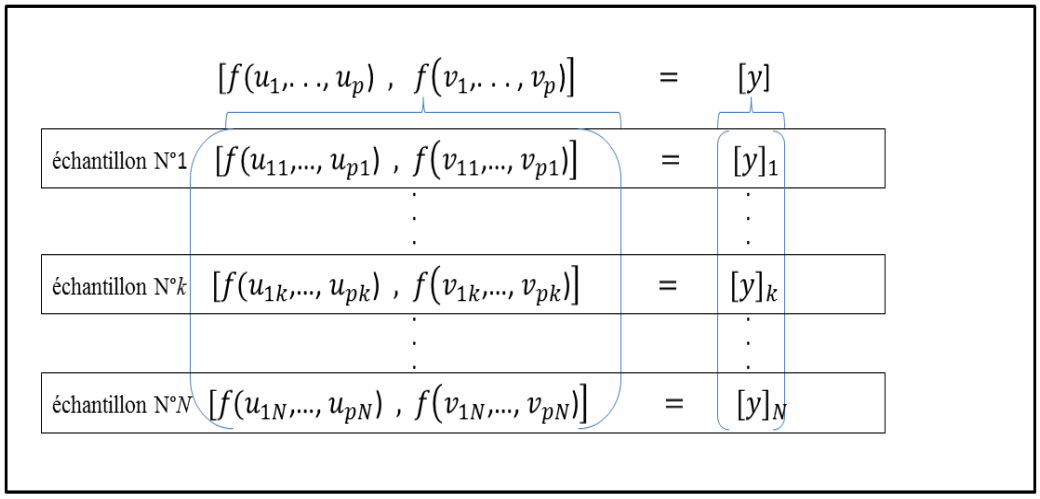

FIGURE 4.16: Matrice d'échantillons

avec $k = 1, ..., N$.

7. Calculer l'intervalle de confiance $[y]$ comme l'union $\bigcup [y]_{k,k=1,...,N}$.

Remarque. Une amélioration pourrait être envisagée pour traiter les entrées par rapport auxquels le modèle n'est pas monotone. Puisqu'on travaille sur de petits sous-intervalles $[x_i]_j$ de $[x_i]$, il peut être intéressant de tester la monotonie du modèle par morceaux sur chaque $[x_i]_j$ tiré plutôt que sur tout le support $[x_i]$, pour se retrouver le plus souvent possible dans la capacité de travailler sur des ensembles de valeurs plutôt que des valeurs.

4.2.2 Application de la méthode mixte au problème de dispersion

4.2.2.1 Modèle gaussien

Dans cette section, on a utilisé les mêmes données que celles présentées dans l'application du chapitre 2 (2.2.1) concernant les valeurs nominales des entrées du modèle de dispersion gaussien et les incertitudes associées.

Propagation des incertitudes

Dans cette partie applicative, la concentration en gaz est estimée en 5 points placés sous le vent de la source d'émission du gaz d'oxyde nitrique. Pour ces 5 points, les supports intervalles des concentrations en gaz calculés, ainsi que les indicateurs représentatifs de l'amplitude des incertitudes sur ces concentrations sont donnés pour la méthode mixte proposée, puis comparés avec ceux obtenus par la simulation de Monte-Carlo et par l'analyse par intervalles.
Compte tenu du faible temps de calculs associé au modèle de dispersion gaussien, et pour permettre une comparaison directe avec les résultats du chapitre 2, nous avons opté pour conserver des incertitudes sur les 6 entrées du modèle. Nous avons donc conservé le nombre d'échantillon qui avait été fixé à $N = 100.000$ pour la méthode classique de Monte-Carlo. Concernant AIMCM, le nombre d'échantillons utilisés est $N = 1000$ et l'ordre de la sous-division uniforme est $m = 10$. Connaissant une fonction d'inclusion du modèle de dispersion gaussien (2.23-2.29), c'est naturellement la première déclinaison de la méthode mixte utilisant le calcul par intervalles qui sera utilisée. Les incertitudes sur les entrées du modèle pour la simulation MC sont représentées par des distributions de probabilités uniformes pour permettre une comparaison avec les méthodes AI et AIMCM qui manipulent des supports bornées.
Le tableau 4.19 donne, pour les 5 points étudiés, les concentrations en gaz $c_{Nom,k}$ calculées à partir des valeurs nominales des entrées étudiées, ainsi que les indicateurs calculés $U - MC$, $U - AI$ et $U - AIMCM$ reflétant le niveau d'incertitude sur la sortie évaluée par la simulation de Monte-Carlo, l'analyse par intervalles et la méthode mixte.

№	$Point(x_k, y_k, z_k)$	$c_{Nom,k}$	$U - MC$	$U - AI$	$U - AIMCM$
1	(40, 5, 2)	10.46	11.99%	22.10%	13.03%
2	(350, 8, 2)	1.22	11.47%	14.34%	12.03%
3	(460, 35, 2)	0.379	10.81%	18.33%	11.25%
4	(600, 50, 2)	0.193	11.39%	19.43%	12.25%
5	(800, 45, 2)	0.187	10.42%	16.57%	10.9%

TABLE 4.19: Les indicateurs calculés

La figure 4.17 représente la propagation des incertitudes avec les méthodes MC, AI et AIMCM pour les 5 points de coordonnées (x_k, y_k, z_k).

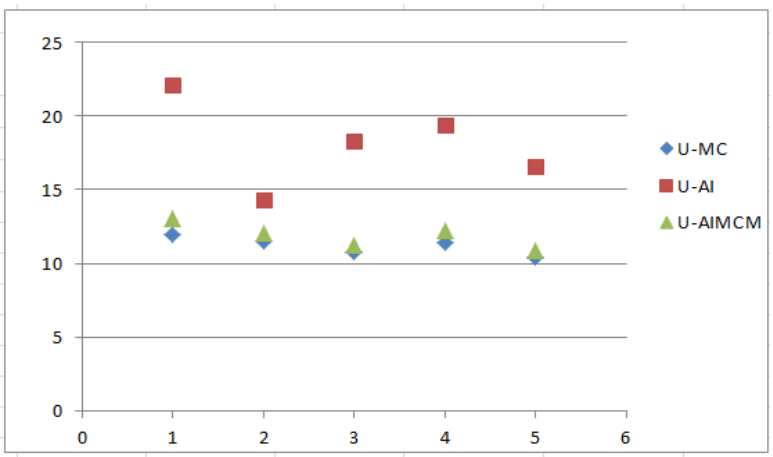

FIGURE 4.17: La propagation de l'incertitude selon les points étudiés

Interprétation des résultats

Les résultats obtenus avec la méthode AIMCM conduisent à un indicateur reflétant le niveau d'incertitude sur la sortie variant entre 10.9% et 13.03%. Pour chacun des points, cet indicateur se retrouve situé entre celui obtenu par la méthode MC et celui obtenu par la méthode AI. Plus précisément, $U-AIMCM$ se retrouve très proche de $U-MC$ alors que le nombre d'échantillons utilisés est beaucoup plus faible. Les résultats se retrouvent même être plus précis au final puisque $U-AIMCM$ se rapproche de $U-AI2$ (tableau 2.5) qui se révèle être le niveau d'incertitude théorique à déterminer puisqu'associé à une fonction d'inclusion minimale. En d'autres termes, la méthode mixte a amélioré les résultats obtenus par la simulation MC classique avec un nombre d'échantillons beaucoup plus faible. Elle conduit dans le même temps à considérablement réduire la surestimation de l'intervalle de sortie résultant du calcul par intervalles. Plusieurs raisons expliquent les différents résultats obtenus par ces approches. La première est due au problème du pessimisme de la méthode AI en raison de multiples occurrences de certaines entrées incertaines du modèle comme a_y, a_z, b_y, bz. Ainsi la sous-division uniforme effectuée par AIMCM conduit à réduire le pessimisme en comparaison de la méthode AI parce que l'utilisation de petits supports pendant le calcul par intervalles réduit la surestimation des intervalles obtenus en sortie. Une autre raison qui se cumule avec la précédente est que l'échantillonnage réalisé ne permet plus de prendre en compte toutes les situations possibles, et on perd donc la garantie des résultats qu'offrait la méthode AI (au prix justement d'une surestimation importante). La deuxième raison est que la simulation MC est basée sur le tirage de valeurs aléatoires pour chaque entrée du modèle alors que

AIMCM manipule des ensembles de valeurs. Ainsi, AIMCM permet de traiter une plus grande proportion des valeurs possibles des entrées et conduit à réduire la sous-estimation induite par la stratégie d'échantillonnage de la simulation MC classique. Finalement, AI et MC conduisent respectivement à des approximations extérieures et intérieures de l'intervalle de confiance exact de la concentration en gaz, alors que AIMCM exprime un compromis entre les deux méthodes et propose une alternative intéressante à priori pour la propagation des incertitudes. En outre, avec AIMCM le temps de calcul est inférieur à celui de MC. La principale raison réside dans la taille de l'échantillon beaucoup plus faible avec AIMCM qui nécessite donc moins de temps de calculs.

4.2.2.2 Modèle SLAB

Propagation des incertitudes

Cette partie applicative a pour but de traiter le cas d'un accident impliquant un gaz toxique et liquéfié comme l'ammoniac (NH3). Le scénario envisagé est le même que pour le cas d'étude présenté dans la section 3.4.3.1 et le modèle de dispersion utilisé est le logiciel SLAB. Ne connaissant pas de fonction d'inclusion pour ce modèle par code, c'est donc la seconde déclinaison de la méthode mixte reposant sur la monotonie du modèle qui sera utilisée.

La concentration en gaz est estimée en 3 points placés sous le vent et qui sont ceux utilisés dans la section 4.1.2.3 ($C280$, $C1415$ et $C3588$). Pour ces 3 points, les supports intervalles des concentrations en gaz, ainsi que les indicateurs représentatifs de l'amplitude des incertitudes sur ces concentrations sont calculés pour la méthode mixte proposée.

En raison du faible nombre d'entrées incertaines, nous avons choisi de ne pas supprimer l'incertitude sur l'humidité et la rugosité ; les incertitudes du modèle affectent donc les 5 grandeurs d'entrée suivantes :

Paramètres	Incertitudes
Température (ta)	$\pm 1\%$
Vitesse du vent (ua)	$\pm 8\%$
Débit de fuite (qs)	$\pm 10\%$
L'humidité (rh)	$\pm 5\%$
Rugosité ($z0$)	$\pm 5\%$

TABLE 4.20: Les incertitudes sur les entrées du modèle

Un test de monotonie préalable (phase 1) a été réalisé selon l'algorithme proposé dans la section 4.1.3, pour étudier la monotonie du modèle SLAB par rapport à chacune de ces 5 entrées incertaines. Les résultats ont montré que la température (ta) et le débit de fuite

(qs) ont une influence de type croissance monotone sur l'estimation de la concentration en gaz sur les différents points choisis et pour les incertitudes considérées, alors que les autres entrées incertaines, c'est-à-dire la vitesse du vent (ua), la rugosité ($z0$) et l'humidité (rh) ont un effet non-monotone sur l'estimation de la concentration en gaz. Ces résultats sont compatibles que ceux obtenus dans les travaux de S. Pagnon [3].

En se basant sur les résultats obtenus par le test de monotonie, les entrées qui sont tirés sous forme d'intervalles sont donc la température (ta) et le débit de fuite (qs), alors que les autres entrées (ua, rh et $z0$) sont tirées sous forme de points.

Les incertitudes sur les entrées du modèle qui sont tirées sous forme de points sont représentées par des distributions de probabilités uniformes. Le nombre d'échantillons utilisés pour la méthode mixte est $N = 1000$ et l'ordre de la sous division uniforme est $m = 10$, nous avons conservé le nombre d'échantillon qui avait été fixé à $N = 100.000$ pour la méthode classique de Monte-Carlo.

Avec la méthode mixte, pour chaque échantillon, le logiciel SLAB est lancé 2 fois, la première pour calculer les bornes inférieures de la concentration en gaz aux différents points en utilisant les bornes inférieures des supports intervalles de la température et du débit de fuite, et de la même manière pour la deuxième en relançant le calcul sur les bornes supérieures.

Pour chaque échantillon, SLAB retourne ainsi les deux tableaux suivants :

x_k	BBC_k	$y_{k,1Max}$	$y_{k,2Max}$	$y_{k,3Max}$	$y_{k,4Max}$	$y_{k,5Max}$	$y_{k,6Max}$
x_1	BBC_1	c_{11Min}	c_{12Min}	c_{13Min}	c_{14Min}	c_{15Min}	c_{16Min}
x_2	BBC_2	c_{21Min}	c_{22Min}	c_{23Min}	c_{24Min}	c_{25Min}	c_{26Min}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
x_n	BBC_n	c_{n1Min}	c_{n2Min}	c_{n3Min}	c_{n4Min}	c_{n5Min}	c_{n6Min}

TABLE 4.21: Tableau des concentrations minimales en chaque point

x_k	BBC_k	$y_{k,1Min}$	$y_{k,2Min}$	$y_{k,3Min}$	$y_{k,4Min}$	$y_{k,5Min}$	$y_{k,6Min}$
x_1	BBC_1	c_{11Max}	c_{12Max}	c_{13Max}	c_{14Max}	c_{15Max}	c_{16Max}
x_2	BBC_2	c_{21Max}	c_{22Max}	c_{23Max}	c_{24Max}	c_{25Max}	c_{26Max}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
x_n	BBC_n	c_{n1Max}	c_{n2Max}	c_{n3Max}	c_{n4Max}	c_{n5Max}	c_{n6Max}

TABLE 4.22: Tableau des concentrations maximales en chaque point

Il faut noter que le fait de lancer une double simulation signifie des entrées différentes à chaque fois et donc un paramétrage différent pour SLAB. Ceci conduit à évaluer la concentration en des points différents du panache pour un même échantillon. Concrètement, le paramètre interne BBC va changer entre les deux simulations et il en sera donc de même pour les ordonnées $y_{k,j(.)}$. A noter que les symboles Min et Max des ordonnées dans les tableaux (4.21) et (4.22) sont inversés par rapport à ceux de la concentration car plus l'ordonnée d'un point est importante, plus la concentration en gaz en ce point est faible.

Le tableau 4.23 donne pour les 3 points étudiés la concentration $c_{Nom,k}$ calculée à partir des valeurs nominales des entrées et l'indicateur de la propagation de l'incertitude $U-MC$ et $U-AIMCM$ réflétant le niveau d'incertitude sur la sortie évalué par la simulation Monte Carlo et la méthode mixte proposée.

№	$Point$	$c_{Nom,k}$	$U-MC$	$U-AIMCM$
1	$C280$	21917.77	10.82%	10.39%
2	$C1415$	2942.55	11.40%	11.11%
3	$C3588$	980.51	15.03%	14.50%

TABLE 4.23: Les indicateurs calculés

Interprétation des résultats

Les résultats obtenus avec la méthode AIMCM conduisent à un indicateur reflétant le niveau d'incertitude sur la sortie variant entre 10.39% et 14.50%. Pour chacun des points, cet indicateur se retrouve très proche de $U-MC$ alors que le nombre d'échantillons utilisés est beaucoup plus faible. En d'autres termes, la méthode mixte conduit à des résultats en propagation d'incertitudes très proches de ceux obtenus par la simulation MC classique avec un nombre d'échantillons beaucoup plus faible. La raison principale est que la simulation MC est basée sur le tirage de valeurs aléatoires pour chaque entrée du modèle alors que AIMCM manipule des ensembles de valeurs pour les entrées par rapport auxquelles le modèle est monotone. Ainsi, AIMCM permet de traiter une plus grande proportion des valeurs possibles des entrées et conduit à réduire la sous-estimation induite par la stratégie d'échantillonnage de la simulation MC classique. En outre, avec AIMCM le temps de calcul est inférieur à celui de MC. Cela s'explique par la taille de l'échantillon beaucoup plus faible avec AIMCM qui nécessite donc moins de temps de calculs.

4.3 Troisième phase : réalisation des cartographies

Pour rappel, les méthodes d'inversion sont spécifiques aux modèles utilisés et sont donc présentées au niveau application. Celles proposées dans la phase 3 en lien avec la méthode mixte reposent sur les méthodes vues dans le chapitre 3 (sections 3.2.1 et 3.2.2) dans le cadre de l'approche probabiliste (simulation de Monte-Carlo) mais adaptées ici, car pour chaque échantillon, nous avons maintenant des supports intervalles en sortie et non plus de simples valeurs.

4.3.1 Modèle gaussien

4.3.1.1 Réalisation des cartographies avec la méthode mixte

La réalisation des cartographies pour le modèle gaussien consiste à inverser le modèle de dispersion en tenant compte des incertitudes associées aux entrées.
Cette partie traite le problème d'inversion via AIMCM. Au début, la variable x_k est discrétisée sur tout le support de l'étude : $x_{k,k=1,...,n}$ de sorte que $x_{k+1} = x_k + \triangle x$ avec $\triangle x$ une constante choisie comme dans la section 3.2.1 et $x_1 = 0$. L'objectif est de trouver pour une valeur x_k donnée la valeur de y_k telle que la concentration en gaz au point (x_k, y_k) soit égale à un seuil choisi c_s. Mais la différence avec AIMCM provient du calcul en sortie des intervalles $[y_k]_{i,i=1,...N}$ à partir d'une fonction d'inclusion de la relation (3.4) au lieu de calculer des valeurs $y_{k,i,i=1,...N}$ de y_k. Evaluer la zone d'incertitudes revenant toujours à déterminer, pour chaque x_k et les N échantillons, les valeurs extrêmes de la sortie y_k où la concentration vaut c_s, l'union d'intervalles $\cup[y_k]_{i,i=1,...N}$ est évaluée. Le maximum $y_{k,max}$ et le minimum $y_{k,min}$ sont obtenus respectivement en déterminant les bornes supérieure et inférieure de cette union d'intervalles. Pour toutes les abscisses x_k de la zone d'étude, nous devons répéter cette opération et nous obtenons au final deux courbes frontières y_{Max} et y_{Min} (figure 4.18) qui délimitent :

- la zone de sécurité où la concentration en gaz doit être inférieure à c_s pour toutes les valeurs possibles des entrées incertaines du modèle,
- la zone de danger où la concentration en gaz doit être supérieure à c_s pour toutes les valeurs possibles des entrées incertaines du modèle,
- la zone d'incertitude où la concentration en gaz peut être supérieure ou inférieure à c_s en fonction des valeurs possibles des entrées incertaines du modèle.

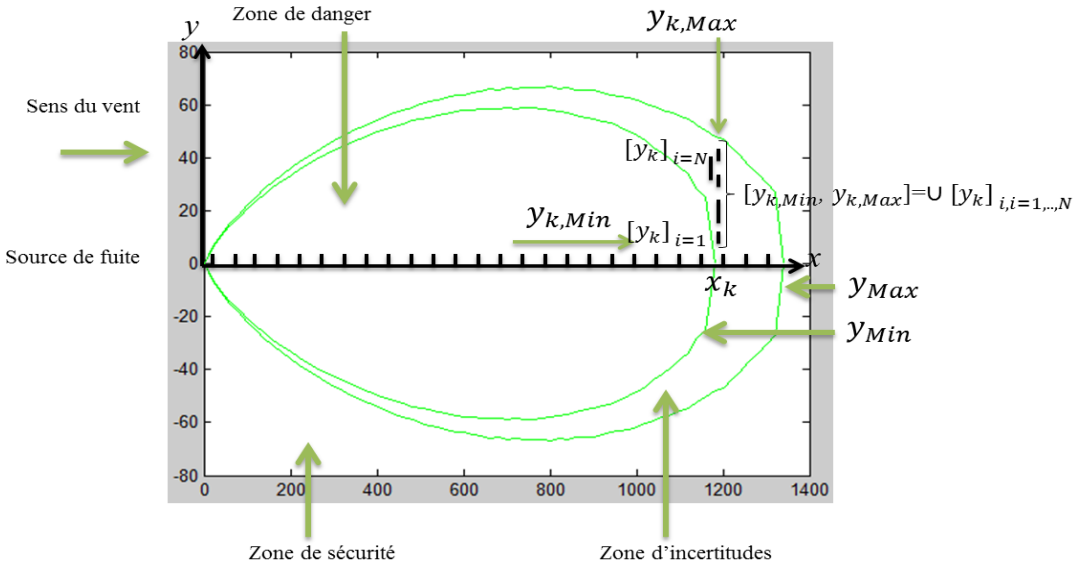

FIGURE 4.18: Panache de NO simulé en phase 3 par AIMCM.

4.3.1.2 Comparaison des cartographies obtenues par MC et AIMCM

Nous pouvons facilement comparer les résultats obtenus en fusionnant les cartographies de MC avec celles de AIMCM (figure 4.19). La zone de danger à l'intérieur de la courbe verte y_{Min} obtenue par AIMCM et celle à l'intérieur de la courbe bleue y_{Min} obtenue par MC sont presque confondues. De la même manière, la zone d'incertitudes (délimitée par les deux courbes bleues) calculée par MC et la zone d'incertitudes (délimitée par les deux courbes vertes) calculées par AIMCM sont aussi pratiquement identiques. Ainsi AIMCM conduit à une précision similaire des zones calculées par rapport à MC avec un nombre d'échantillons beaucoup plus faible ($N = 1000$ échantillons pour AIMCM quand $N = 100.000$ pour MC), ce qui conduit à réduire le temps de calculs. Le temps d'exécution pour obtenir les résultats avec AIMCM n'est que de 36.08 secondes tandis que le temps de calculs pour obtenir les résultats avec MC vaut 513.52 secondes. Donc la méthode mixte proposée permet d'obtenir quasiment la même cartographie que par MC mais avec une réduction du temps de calculs de 93%.

Cette réduction du temps de calculs dans la génération de la cartographie conduit à une réponse plus rapide. Remarquons que le croisement des courbes y_{Max} pour les dernières valeurs de x_k provient d'un nombre N d'échantillons différent pour AIMCM et MC et donc de tirages différents pour les échantillons. Si les sous-intervalles utilisés dans les échantillons de AIMCM contenaient toutes les valeurs d'entrées utilisées dans les échantillons de MC, la zone incertitudes calculée par AIMCM encadrerait nécessairement la zone correspondante obtenue par MC.

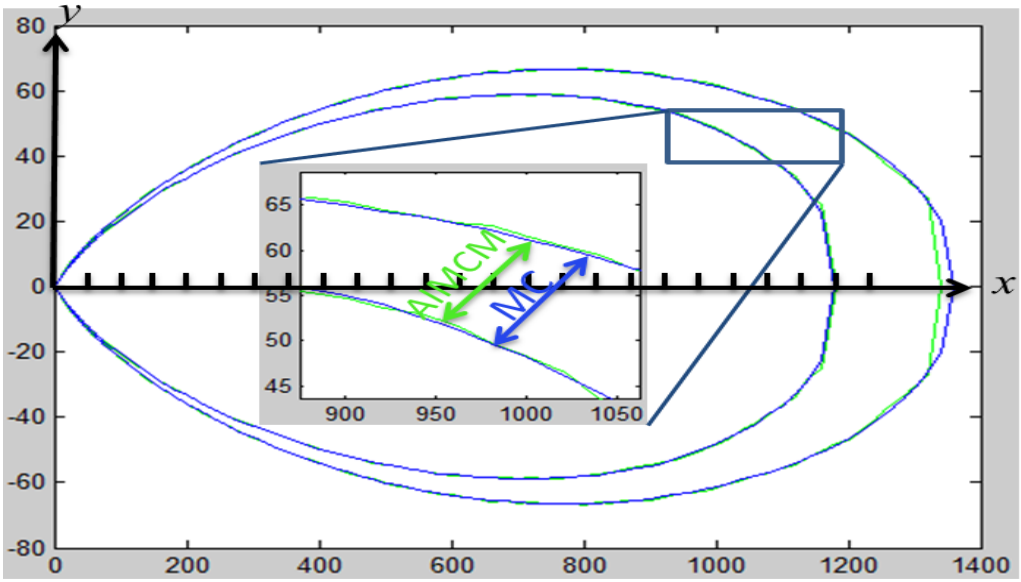

FIGURE 4.19: Panache de NO simulé par AIMCM et MC.

4.3.2 Modèle SLAB

4.3.2.1 Réalisation des cartographies avec la méthode mixte

L'objectif est de déterminer les zones géographiques dans lesquelles la concentration en gaz est inférieure (zone de sécurité) ou supérieure (zone de danger) à un seuil réglementaire donné en utilisant le modèle SLAB et la méthode mixte (AIMCM).
Les données utilisées dans cette partie sont les mêmes que celles utilisées dans la partie 3.4.3.
Pour rappel, comme pour la méthode d'inversion probabiliste présentée dans la section 3.2.2, l'axe des abscisses est discrétisé avec un pas variable (plus faible à proximité de la source) et pour chaque valeur x_k obtenue, la concentration est évaluée par SLAB en 6 ordonnées différentes $y_{k,j,j=1,\ldots,6}$.
Pour un échantillon en entrée (une simulation), l'objectif est toujours de trouver, pour une valeur x_k donnée, la valeur de y_k telle que la concentration au point (x_k, y_k) soit égale à c_s. Mais la différence avec AIMCM est que pour chaque échantillon d'indice i, une double évaluation (simulation) est effectuée afin de déterminer l'intervalle de concentration en gaz puisque certaines entrées (température et débit de fuite) sont tirées sous-formes d'intervalles. Le logiciel SLAB renvoie donc, suivant la méthode détaillée dans la section propagation d'incertitudes 4.2.2.2, deux tableaux 4.21 et 4.22 contenant les valeurs minimales et maximales de concentration en chaque point considéré.

Pour chaque x_k, on recherche les valeurs des bornes des concentrations en gaz qui permettent d'encadrer au plus proche le seuil c_s, ce qui conduit à deux cas possibles :

– le cas 1 où le seuil c_s se situe entre deux supports de concentration $[c_{k,j-1Min}, c_{k,j-1Max}]$ et $[c_{k,jMin}, c_{k,jMax}]$ consécutifs : $c_{k,jMax} < c_s < c_{k,j-1Min}$,

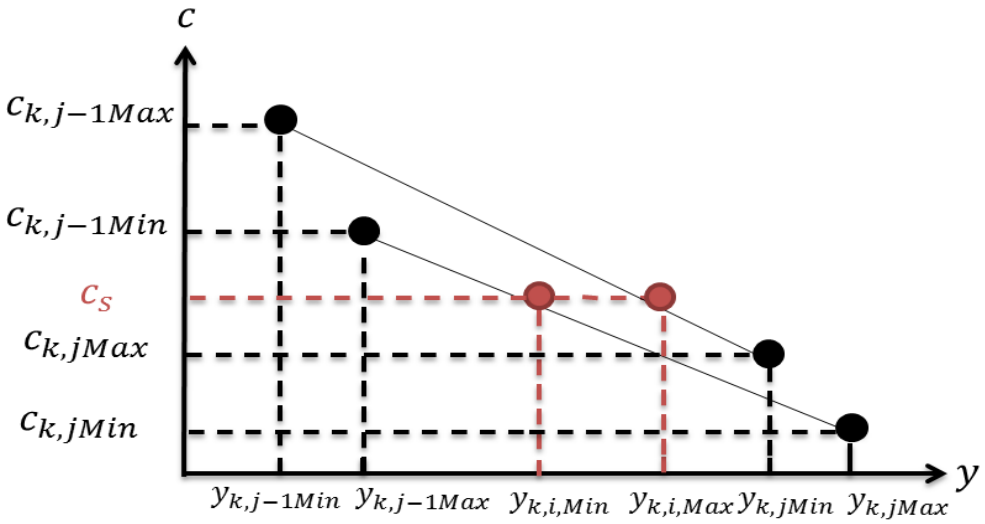

FIGURE 4.20: Cas 1

– le cas 2 où le seuil c_s se situe directement dans le support de concentration $[c_{k,jMin}, c_{k,jMax}]$: $c_{k,jMin} <= c_s <= c_{k,jMax}$.

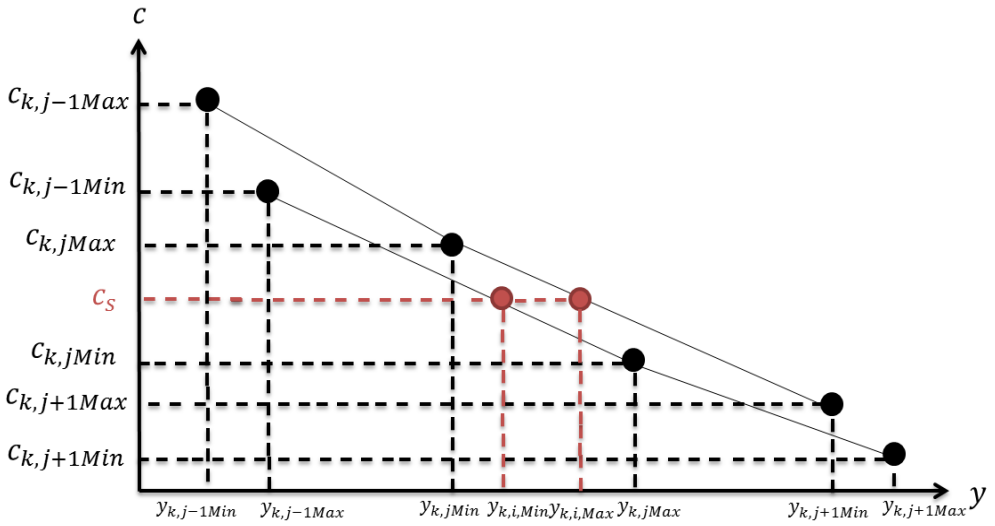

FIGURE 4.21: Cas 2

Selon le cas où se trouve le seuil c_s et à partir des données des tableaux 4.21 et 4.22, une interpolation est réalisée afin d'en déduire les valeurs $y_{k,i,Min}$ et $y_{k,i,Max}$ délimitant la zone d'incertitudes pour une valeur donnée x_k et l'échantillon d'indice i utilisé.
Pour le premier cas, une double interpolation linéaire (figure 4.20) est utilisée pour estimer respectivement $y_{k,i,Min}$ et $y_{k,i,Max}$, sur le même principe que dans la section 3.2.2 :

$$a_{Min} = \frac{c_{k,j-1Min} - c_{k,jMin}}{y_{k,j-1Max} - y_{k,jMax}} \text{ et } b_{Min} = c_{k,jMin} - (a_{Min} \times y_{k,jMax})$$

ce qui conduit à : $y_{k,i,Min} = \dfrac{c_s - b_{Min}}{a_{Min}}$

$$a_{Max} = \frac{c_{k,j-1Max} - c_{k,jMax}}{y_{k,j-1Min} - y_{k,jMin}} \text{ et } b_{Max} = c_{k,jMax} - (a_{Max} \times y_{k,jMin})$$

ce qui conduit à : $y_{k,i,Max} = \dfrac{c_s - b_{Max}}{a_{Max}}$

Pour le deuxième cas où le seuil c_s est élément de $[c_{k,jMin}, c_{k,jMax}]$, il est nécessaire d'utiliser les concentrations estimées aux points $(x_k, y_{k,j-1})$ et $(x_k, y_{k,j+1})$ encadrant le point $(x_k, y_{k,j})$, c'est-à-dire réaliser une double interpolation parabolique sur 3 points (figure 4.21). Dans un premier temps, les coefficients de chaque parabole utilisée sont identifiés afin de passer par les données renvoyées par SLAB. Ensuite, deux équations du second degré sont résolues pour trouver la valeur de y_k menant à $c_k = c_s$. Pour l'abscisse x_k choisie et le $i$ème échantillon sélectionné, les valeurs $y_{k,i,Max}$ et $y_{k,i,Min}$ interpolées sont détaillées ci-après.

Identification des coefficients de la parabole :

$$c_k = a_{0Min} + a_{1Min} \times (y_k - y_{k,jMax}) + a_{2Min} \times (y_k - y_{k,jMax})^2$$

$$a_{0Min} = c_{k,jMin}$$

$$a_{1Min} = \frac{c_{k,j-1Min} - c_{k,jMin}}{y_{k,j-1Max} - y_{k,jMax}} + \frac{c_{k,j-1Min} - c_{k,jMin}}{y_{k,j+1Max} - y_{k,j-1Max}} - \frac{(y_{k,j-1Max} - y_{k,jMax}) \times (c_{k,j+1Min} - c_{k,jMin})}{(y_{k,j+1Miax} - y_{k,j-1Max}) \times (y_{k,j+1Max} - y_{k,jMax})}$$

$$a_{2Min} = \frac{c_{k,j+1Min} - c_{k,jMin}}{(y_{k,j+1Max} - y_{k,jMax}) \times (y_{k,j+1Max} - y_{k,j-1Max})} - \frac{c_{k,j-1Min} - c_{k,jMin}}{(y_{k,j+1Max} - y_{k,j-1Max}) \times (y_{k,j-1Max} - y_{k,jMiax})}$$

$$\triangle = a_{1Min}^2 - 4a_{2Min}(a_{0Min} - c_s)$$

Racines solution :

$$y_{k,i,Min} = \frac{-a_{1Min} \pm \sqrt{\triangle}}{2a_{2Min}} + y_{k,jMax}$$

Il faut garder la racine comprise entre $y_{k,j-1Max}$ et $y_{k,j+1Max}$.

Identification des coefficients de la parabole :

$$c_k = a_{0Max} + a_{1Max} \times (y_k - y_{k,jMin}) + a_{2Max} \times (y_k - y_{k,jMin})^2$$

$$a_{0Max} = c_{k,jMax}$$

$$a_{1Max} = \frac{c_{k,j-1Max} - c_{k,jMax}}{y_{k,j-1Min} - y_{k,jMin}} + \frac{c_{k,j-1Max} - c_{k,jMax}}{y_{k,j+1Min} - y_{k,j-1Min}} - \frac{(y_{k,j-1Min} - y_{k,jMin}) \times (c_{k,j+1Max} - c_{k,jMax})}{(y_{k,j+1Min} - y_{k,j-1Min}) \times (y_{k,j+1Min} - y_{k,jMin})}$$

$$a_{2Max} = \frac{c_{k,j+1Max} - c_{k,jMax}}{(y_{k,j+1Min} - y_{k,jMin}) \times (y_{k,j+1Min} - y_{k,j-1Min})} - \frac{c_{k,j-1Max} - c_{k,jMax}}{(y_{k,j+1Min} - y_{k,j-1Min}) \times (y_{k,j-1Min} - y_{k,jMin})}$$

$$\triangle = a_{1Max}^2 - 4a_{2Max}(a_{0Max} - c_s)$$

Racines solution :

$$y_{k,i,Max} = \frac{-a_{1Max} \pm \sqrt{\triangle}}{2a_{2Max}} + y_{k,jMin}$$

Il faut garder la racine comprise entre $y_{k,j-1Min}$ et $y_{k,j+1Min}$.
On a donc réussi, pour l'échantillon d'indice i, à calculer pour chaque abscisse discrétisée x_k, les valeurs limites $y_{k,i,Min}$ et $y_{k,i,Max}$ délimitant les zones de danger et d'incertitudes. Le principe est ensuite de réitérer la même opération pour chaque échantillon, c'est-à-dire pour les N combinaisons aléatoires de supports intervalles et de valeurs des entrées incertaines du modèle SLAB. On obtient alors pour chaque valeur x_k, N supports intervalles $[y_k]_i = [y_{k,i,Min}, y_{k,i,Max}], i = 1, ..., N$.
Evaluer la zone d'incertitudes revenant toujours à déterminer, pour chaque x_k et les N échantillons, les valeurs extrêmes de la sortie y_k où la concentration vaut c_s, l'union d'intervalles $\bigcup[y_k]_{i,i=1,...,N}$ est évaluée. Pour chaque x_k, les bornes de cette union $y_{k,Max}$ et $y_{k,Min}$ expriment les frontières (courbes y_{Max} et y_{Min}) entre les zones de sécurité, de danger et d'incertitudes. Une localisation en un point (x_k, y_k) avec $y_k < y_{k,Min}$ correspond à une concentration en gaz supérieure à c_s. Si $y_k > y_{k,Max}$, la concentration en gaz est inférieure à c_s. Si $y_{k,Min} < y_k < y_{k,Max}$, la concentration en gaz peut être inférieure ou supérieure à c_s en fonction des incertitudes (figure 4.22). Le nombre d'échantillons est ajusté jusqu'à ce qu'aucune modification significative dans les cartographies ne soit observée. Cela conduit à choisir $N = 10000$ échantillons.

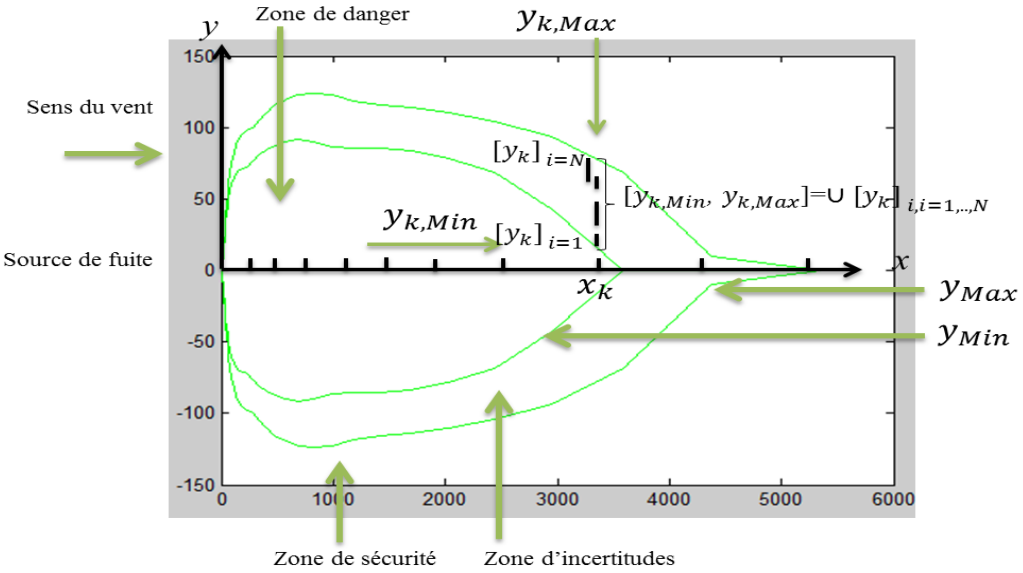

FIGURE 4.22: Panache de NH3 simulé par AIMCM

4.3.2.2 Comparaison des cartographies obtenues par MC et AIMCM

La figure 4.23 illustre la fusion des cartographies obtenues par la méthode de Monte Carlo et la méthode AIMCM. De cette comparaison, nous pouvons conclure que les résultats obtenus par la méthode mixte sont tout aussi précis (voire même ici sous-estiment légèrement moins la zone d'incertitudes) que ceux obtenus par Monte-Carlo, tout en reposant

sur un nombre d'échantillons plus faible. Le temps de calculs pour la simulation de Monte Carlo avec 100.000 échantillons est de 250s, alors qu'il n'est que de 58s avec AIMCM qui utilise ici 10000 échantillons, soit une réduction de 76.8% du temps de calculs.

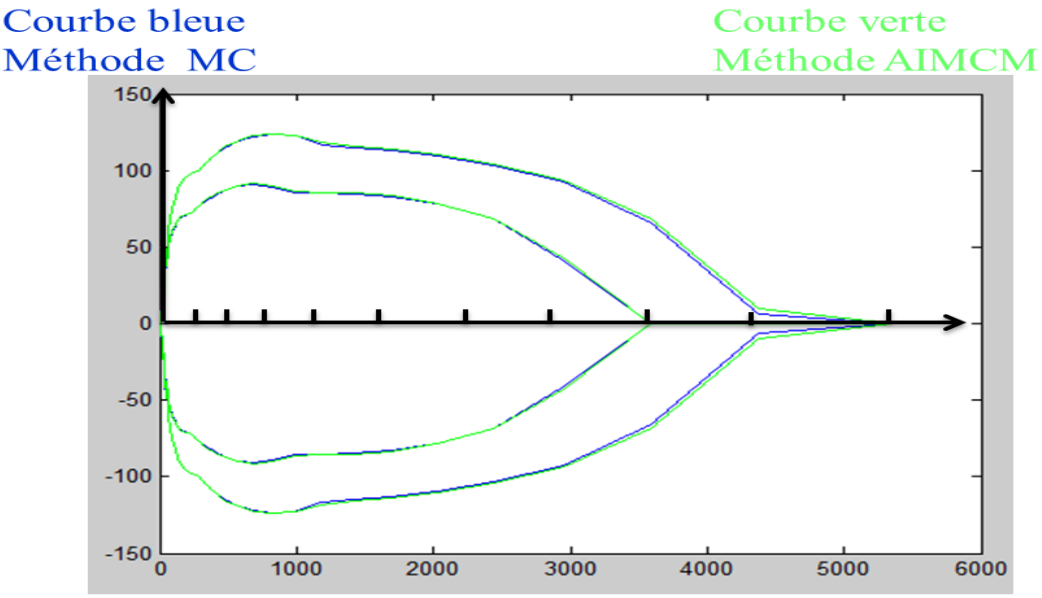

FIGURE 4.23: Panache de NH3 simulé par AIMCM et MC

La figure 4.24 illustre la fusion des résultats de la méthode Monte Carlo avec toujours 100.000 échantillons et de la méthode AIMCM avec cette fois-ci seulement 1000 échantillons. La précision de la méthode mixte proposée reste tout à fait acceptable avec un temps de calculs de seulement 5s, soit une réduction de 98% par rapport à la simulation de Monte-Carlo, ce qui, d'un point de vue opérationnel pour le terrain, est très intéressant.

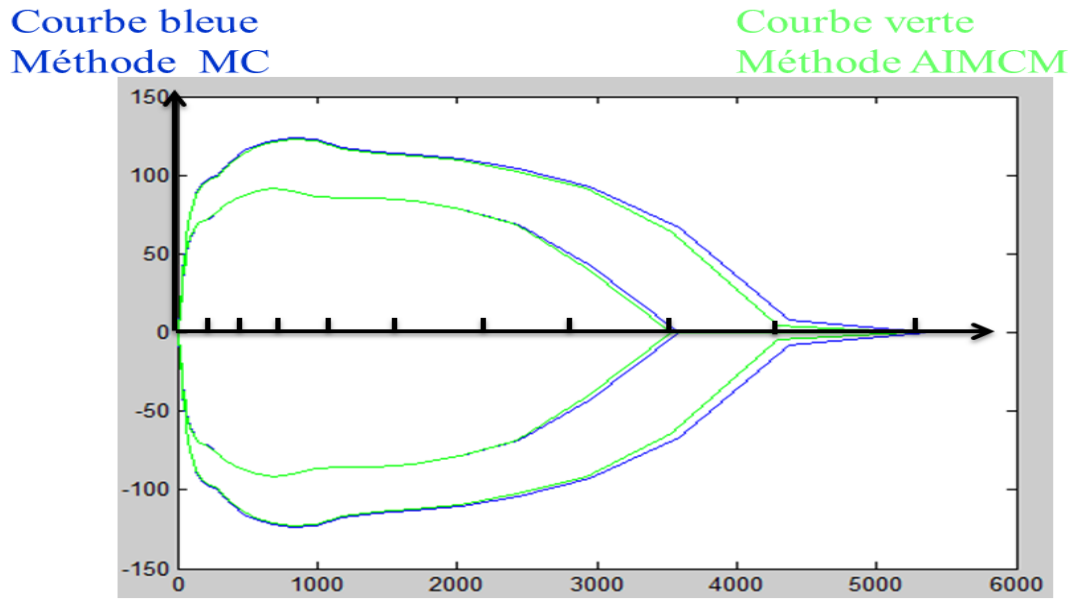

FIGURE 4.24: Panache de NH3 simulé par AIMCM et MC

4.4 Conclusion

Dans ce chapitre, nous avons proposé une **méthodologie générale** pour réaliser les cartographies en améliorant les performances en termes de temps du calculs et de précision ; et qui s'adresse à des modèles de dispersion dont on connait l'expression analytique (ce qui en sous-entend d'avoir à disposition une fonction d'inclusion) ou au contraire dont on a une description que sous-la forme d'un code informatique (boite noire).

Cette méthodologie, détaillée tout au long du chapitre, est divisée en 3 phases.

La première phase consiste à analyser préalablement les modèles d'effets en vue de faciliter leur traitement lors des phases suivantes. Il est ainsi proposé de réaliser une analyse de sensibilité globale dans le but de ne plus prendre en compte les incertitudes sur les entrées les moins influentes sur la variation de la sortie du modèle. Pour ce faire, peuvent être utilisés les indices de Sobol dédiés à des modèle probabilistes, ou un tout nouvel indice de sensibilité intervalle dédié spécifiquement au traitement de modèles par intervalles. Il est aussi possible d'étudier en fonction du type de modèle, soit sa monotonie par rapport aux entrées incertaines, soit le pessimisme généré par la multi-occurrence de ces mêmes entrées. Notons qu'un indicateur a été développé afin de déterminer le niveau de pessimisme engendré par une entrée donnée dans le cas d'un modèle par intervalles.

La deuxième phase s'appuie sur une nouvelle approche développée pour la propagation des incertitudes mixant les approches probabiliste et ensembliste. Reposant sur le principe de la simulation de Monte-Carlo, l'idée principale de la méthode mixte (AIMCM) est, au niveau de l'échantillonnage des entrées, d'essayer de tirer des sous-ensembles de valeurs plutôt que de simples valeurs afin d'appréhender plus de situations lors de chaque simulation et ainsi réduire la taille de l'échantillon. Le principe de la méthode est général, puisqu'elle permet de traiter aussi bien :

- des modèles (analytiques) intervalles où sont tirés des supports intervalles plus ou moins grands en fonction du pessimisme à maîtriser,
- des modèles probabilistes (analytiques ou par code) en tirant des supports intervalles sur les entrées incertaines par rapport auxquelles le modèle est monotone, des valeurs sinon.

Dans le cas d'un modèle par intervalles générant peu de pessimisme, cette méthode se limite à manipuler les supports complets des entrées sans les découper, ce qui correspond à l'approche ensembliste détaillée dans le chapitre 2 pour la propagation des incertitudes. Dans le cas d'un modèle probabiliste sans propriété de monotonie, elle revient à faire une simulation de Monte-Carlo classique en ne tirant que des valeurs. Il faut juste garder à l'esprit que, si en sortie ce sont bien des supports intervalles qui sont obtenus avec la

méthode mixte proposée, la manière de les calculer dépend de la nature du modèle. Cet aspect nécessite donc de la décliner en deux versions :

- une première reposant sur l'analyse par intervalles si un modèle par intervalles (fonction d'inclusion) est utilisé,
- sinon une seconde reposant sur une évaluation inf/sup du modèle.

La dernière phase repose sur une méthode d'inversion utilisant les résultats de la phase précédente pour réaliser les cartographies en inversant les modèles de dispersion. Si le principe de résolution lors de cette phase reste général, la méthode d'inversion doit être adaptée à la fois au type de modèle, mais aussi aux contraintes imposées par exemple par le logiciel codant le modèle qui renvoie les résultats de simulation sous un format imposé.

La démarche générale en trois phases, incluant la méthode mixte (AIMCM), a été appliquée au modèle de dispersion gaussien et au logiciel SLAB. L'analyse de sensibilité a été détaillée pour permettre d'évaluer la pertinence des indices de sensibilité utilisés, notamment celui intervalle proposé. Par la suite, nous avons choisi de conserver toutes les entrées incertaines pour mieux mettre en lumière les améliorations apportées en termes de performances par la méthode mixte, mais il est clair que les temps de calculs mentionnés auraient été encore plus faibles si l'on avait exploité les résultats de l'analyse de sensibilité. Les résultats obtenus en propagation d'incertitudes et réalisation des cartographies ont été comparés à ceux obtenus à l'aide des approches probabiliste (MC) et ensembliste (AI) présentés dans les chapitres 2 et 3. Il en ressort que la méthode mixte conduit à des résultats similaires avec un temps de calculs bien inférieur à l'approche MC grâce à une taille d'échantillon plus faible, tout en permettant de traiter une plus grande variété de modèles que l'approche AI. A l'inverse, pour une taille d'échantillon similaire, AIMCM aura tendance à diminuer la sous-estimation induite par l'approche MC en traitant une plus grande diversité des valeurs possibles des entrées (et donc de la sortie). De même, AIMCM aura tendance à réduire la surestimation des résultats obtenus par l'approche AI (quand le modèle est compatible) en maîtrisant mieux le pessimisme généré par le calcul par intervalles en travaillant sur de petits supports sur les entrées.

En termes de perspectives, faire une étude de monotonie par morceaux apportera une plus grande flexibilité à l'approche proposée en permettant de tirer plus facilement des ensembles de valeurs, ce qui aura pour conséquence de réduire encore plus la taille des échantillons manipulés. De plus l'objectif de ce travail était essentiellement de comparer des approches ou des méthodes et non les modèles d'effets utilisés. Il pourrait être intéressant de tester d'autres scénarios avec des produits de caractéristiques différentes, des plages d'incertitudes différentes,... voire de tester un scénario identique pour les deux modèles afin de comparer les cartographies obtenues.

Conclusion et perspectives

Dans ce document, nous avons considéré le problème de l'évaluation du niveau de risque des zones soumises au Transport des Matières Dangereuses (TMD), notamment le problème de dispersion atmosphérique de produits toxiques. Cette évaluation des risques repose sur la quantification de l'intensité des phénomènes qui se produisent, à l'aide de modèles d'effets (analytiques ou codes informatiques) en prenant en compte les incertitudes affectant certaines entrées. Les travaux menés dans le cadre de cette thèse avaient pour objectifs :

- d'évaluer les incertitudes sur la concentration en gaz émis en propageant celles induites par les grandeurs d'entrée incertaines lors de l'évaluation des modèles de dispersion,
- de réaliser des cartographies en évaluant les zones de danger et d'incertitudes utilisées pour évaluer l'intensité des risques dans la zone impactée en procédant à l'inversion des modèles de dispersion.

Ces objectifs se sont concrétisés par la proposition d'une méthodologie originale pour réaliser ces cartographies, à la fois générale dans sa capacité à traiter des modèles de dispersion aussi bien analytiques que par code informatique, mais aussi dans les outils mis en œuvre pour permettre d'améliorer les performances en termes de temps de calculs ou de précision par rapport aux approches probabiliste ou ensembliste relativement plus classiques.

Une première partie de la thèse a été consacrée à propager les incertitudes issues des grandeurs d'entrée incertaines dans les modèles de dispersion. Deux approches ont été présentées pour ce faire : l'approche ensembliste qui repose sur l'Analyse par Intervalles (AI) pour les modèles analytiques et l'approche probabiliste qui repose sur la simulation de Monte-Carlo (MC), plus classique et utilisable que les modèles de dispersion soient analytiques ou définis par du code informatique. L'objectif était de détailler les principes de fonctionnement de chacune de ces approches et d'en poser les mécanismes (échantillonnage, calculs par intervalles, pessimisme,...) dans un contexte académique (propagation d'incertitudes pour un modèle gaussien) dans un souci de pédagogie avant de passer à un niveau de complexité supérieur (inversion d'un modèle incertain). Cela a permis de comparer les deux approches afin d'en connaitre les avantages et inconvénients en termes

de précision et temps de calculs pour résoudre le problème proposé. Les résultats obtenus ont montré que l'analyse par intervalles conduit à une surestimation de l'intervalle de confiance sur la concentration en gaz émis. Pour la méthode de Monte Carlo, le processus utilisé pour la propagation des incertitudes repose sur un échantillonnage aléatoire qui conduit à une sous-estimation de l'intervalle de confiance recherché. En termes de temps de calculs, l'analyse par intervalle s'avère plus rapide que la simulation de Monte Carlo.

Dans une seconde partie, nous nous sommes intéressés au problème de génération des cartographies des zones de danger et de sécurité, qui tout en reposant sur les résultats de propagation d'incertitudes, correspond au réel objectif de ce travail. Deux modèles ont été utilisés pour illustrer les approches proposées. Le modèle de dispersion gaussien, plus simple et analytique, a permis de réaliser une comparaison des approches utilisées sur un plan académique. Le second modèle SLAB, plus sophistiqué et par code, avait pour but de se confronter à une étude plus réaliste. La réalisation des cartographies a été abordée avec la méthode probabiliste (Monte Carlo) qui consiste à inverser « à la main » le modèle de dispersion (manipulation formelle du modèle analytique, interpolation des concentrations renvoyées par SLAB), puis avec une méthode ensembliste générique qui consiste à formuler ce problème sous la forme d'un ensemble de contraintes à satisfaire (CSP) et à le résoudre ensuite par une méthode d'inversion ensembliste. Notons que la génération des cartographies en contexte incertain se traduit par l'estimation d'une troisième zone, appelée zone d'incertitudes, située entre les zones de danger et de sécurité, où il est impossible de dire, compte tenu des incertitudes sur les entrées, si la concentration en gaz dépasse réellement ou non le seuil préconisé. Concernant le modèle gaussien, la réalisation des cartographies avec les deux approches a montré que l'inversion ensembliste génère un majorant de la zone (cumulée de danger et d'incertitudes) où le risque est potentiellement inacceptable ; alors qu'à l'inverse, l'approche probabiliste en calcule un minorant. L'intérêt d'utiliser les deux approches réside dans l'encadrement des zones de danger et d'incertitudes exactes. Dans le cadre de l'évaluation des risques TMD, il est a priori plus intéressant d'avoir une surestimation de la zone où les risques sont élevés, avec une précision maîtrisée, au lieu d'en obtenir seulement une partie. Une sous-estimation peut en effet conduire à une évacuation insuffisante ou la mise en place de barrières inadéquates.

Tous ces éléments nous ont finalement conduits à développer une méthodologie générale pour réaliser les cartographies en cherchant d'une part à s'adapter à la nature du modèle de dispersion utilisé, et d'autre part à améliorer les performances en termes de temps du calculs et de précision par rapport à l'approche probabiliste.
En ce qui concerne le premier point, cette approche est utilisable pour propager les incertitudes et résoudre le problème d'inversion associé à tout type de modèles de dispersion spatialisés et statiques, qu'ils soient analytiques et par code informatique. Dans

le premier cas, le calcul par intervalles sera privilégié en travaillant sur des fonctions d'inclusion. Dans le second cas, il est proposé d'étudier la monotonie des modèles par rapport aux entrées incertaines pour être capable d'évaluer l'intervalle de sortie même pour un modèle de type code informatique.
Pour ce qui est du compromis entre réduire les temps de calculs et conserver une précision raisonnable dans les zones estimées, différents outils sont proposés. Une analyse de sensibilité globale est réalisée pour déterminer quelles sont les entrées incertaines les moins influentes qu'il est raisonnable de fixer à une valeur nominale, et ainsi limiter la taille des échantillons à manipuler. Nous avons notamment proposé un nouvel indicateur de sensibilité par intervalles adapté aux modèles intervalles. Il permet de plus de déterminer le niveau de pessimisme engendré par une entrée donnée dans le cas d'un modèle par intervalles pour mieux le maitriser. Enfin une nouvelle méthode pour la prise en compte et la propagation des incertitudes mixant les approches probabiliste et ensembliste a été développée. Appelée méthode mixte, elle consiste à tirer des ensembles de valeurs (intervalles) plutôt que de simple valeurs pour appréhender plus de situations possibles, tout en utilisant un nombre de tirages réduit. Elle permet aussi de travailler sur des intervalles plus petits pour limiter le pessimisme du calcul par intervalles lorsque celui-ci est utilisé.

Enfin la mise en application de cette méthodologie a été réalisée afin de comparer les résultats entre la simulation classique de Monte-Carlo et l'approche mixte proposée. Il en ressort que la méthode mixte conduit à des résultats similaires en termes de précision, mais avec des temps de calculs bien inférieurs grâce à la manipulation de tailles d'échantillons plus faibles, tout en permettant de traiter une plus grande variété de modèles que l'approche purement intervalles. Cela conduit à des durées de quelques secondes pour estimer une zone de danger parfaitement compatibles avec une utilisation opérationnelle. A l'inverse, pour une taille d'échantillon similaire, la méthode mixte aura tendance à diminuer la sous-estimation induite par la simulation de Monte-Carlo en traitant une plus grande diversité des valeurs possibles des entrées (et donc de la sortie). De même, la méthode mixte aura tendance à réduire la surestimation des résultats obtenus par l'approche par intervalles (quand le modèle est compatible) en maîtrisant mieux le pessimisme généré par le calcul par intervalles en travaillant sur de petits supports sur les entrées.

Ces travaux des recherches offrent de nombreuses perspectives :

• L'approche proposée pourrait être étendue en réalisant une étude de monotonie par morceaux qui apporterait une plus grande flexibilité en permettant de tirer plus facilement des ensembles de valeurs, ce qui aurait pour conséquence de réduire encore plus la taille des échantillons manipulés.

• L'objectif de ce travail était essentiellement de comparer des approches ou des méthodes et non les modèles d'effets utilisés. Il pourrait être intéressant de tester d'autres scénarios avec des produits ayant des caractéristiques différentes, des plages d'incertitudes différentes sur les entrées,... voire de tester un scénario identique pour les deux modèles afin de comparer les cartographies obtenues.

• La méthode mixte pourrait être appliquée à d'autres types des problèmes comme la détection des défauts de capteurs mesurant la concentration en gaz toxique ou la localisation de la source de fuite.

• Il serait intéressant de réfléchir à un nouvel indicateur pour l'analyse de sensibilité spécifique à la méthode mixte, tenant compte des spécificités des modèles probabiliste et ensembliste.

• Les cartographies générées dans le cadre de ce travail sont en 2D, il serait donc intéressant de générer des cartographies tridimensionnelles.

• Bien que nous n'ayons pour l'instant appliqué la démarche méthodologique qu'aux modèles SLAB et gaussien, cette démarche est générale et pourrait être appliquée à d'autres modèles de dispersion.

• Il serait intéressant d'améliorer la méthode proposée pour l'adapter aux modèles de dispersion spatialisés et dynamiques.

• Au niveau applicatif, ces travaux pourraient être développés pour d'autres modes de transport comme par canalisation, le ferroviaire ou le maritime.

• L'étude pourrait également être prolongée et transposée à d'autres phénomènes dangereux comme l'explosion et l'incendie.

Annexe A

Fonctionnalités du service web d'évaluation du niveau de risque

L'outil développé fournit un service web qui permet à l'utilisateur de déterminer les zones de danger avec les niveaux d'intensité des phénomènes dangereux pour un TMD (Transport des Matières Dangereuses).
Il fournit le diamètre ou les zones sous forme de listes de boites de la zone d'intensité SEI (effets irréversibles) et SEL (effets létaux), la zone SELS (effets létaux significatifs) étant moins utilisée. Il permet de décrire la répartition de la population sous forme de densités ou de zones à enjeux donnés.

Le logiciel expose une fonction principale comme service web. Elle prend comme paramètres d'entrée :
- les coordonnées GPS de la zone explorée,
- la liste des camions qui sont à l'intérieur de cette zone et leurs caractéristiques (contenu,...),
- l'information liée à l'état de la météo.

Au niveau de la sortie cette fonction retourne :
- une liste de boites avec le niveau de danger associé et des informations liées à cette zone.

Le calcul est réalisé en utilisant, soit un modèle de dispersion comme celui présenté dans la section 1.4.1.1, soit en utilisant une zone majorante, pré-calculée et indépendante des conditions météo, les conditions défavorables étant prises en compte, pour les produits les moins utilisés.
Cette approche a été choisie de façon à fournir un résultat dans tous les cas, bien que dans le cadre d'une thèse il ne soit pas possible de fournir des modèles pour toutes les situations.

Architecture système global-Geofencig MD

Ce service web est déployé sur le serveur G-SCOP sous l'adresse `http://project.g-scop.grenoble-inp.fr/TMD_web_test/services`. Le serveur G-SCOP développé fait partie de l'architecture technique et fonctionnelle du système global-Geofencig-MD comme présenté dans la figure A.1 :

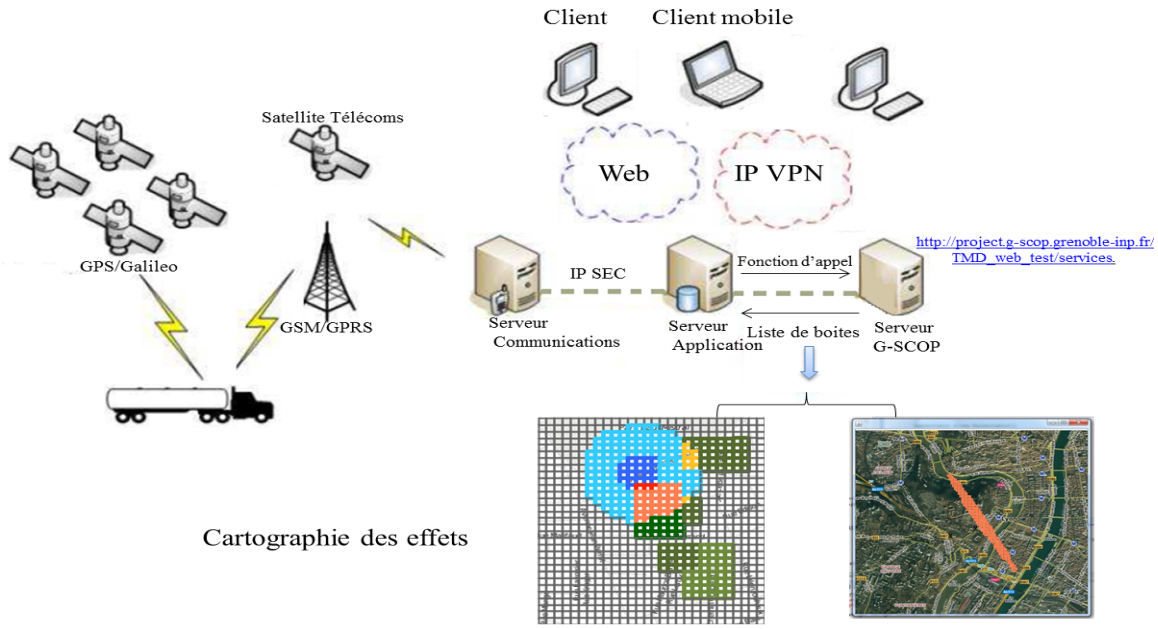

FIGURE A.1: Architecture système global-Geofencig MD

La fonction principale a le format suivant :
public Boite [ ] calculerNiveauRisque(ZoneExploree zone, DonneesMeteo meteo, DonneesCamion [ ] Camions).
Entrées :
- coordonnées GPS de la zone de travail « Coins haut gauche/bas droit ».
- liste des camions (Coord GPS/Code ONU/Quantité) qui sont à l'intérieur de la zone de travail.
- données météorologiques (état de météo/température/sens de vent/vitesse de vent/-temps).

Sorties :
- liste de boites "rectangulaires" avec les informations nécessaires (Coord GPS/impact/-vulnérabilité/ niveau de risque) pour chaque boite.

Le niveau de risque est calculé en fonction de deux paramètres :

1. l'intensité du phénomène dangereux, obtenu à partir de la base de données pré-calculée ou du modèle de dispersion.
2. l'importance de la population touchée obtenue en utilisant la densité de la zone touchée.

Pour chaque camion on prend en compte 3 informations principales :

1. le code ONU important pour préciser le produit émis,
2. la quantité contenue
   - pour déterminer si on a une large ou petite dispersion dans le cas d'une zone précalculée,
   - ou pour alimenter le modèle de dispersion.
3. les données météo (état de météo/température/sens de vent/vitesse de vent/-temps) :
   - on prend en compte l'heure pour préciser si la dispersion a lieu le jour ou la nuit (temps>= 21600s && temps <=64800s),
   - et le sens et la vitesse du vent pour le modèle de dispersion.

Les couleurs en fonction des enjeux et de l'intensité du phénomène dangereux sont les suivantes :

Pour les enjeux on a 4 niveaux de vulnérabilité (en fonction de la densité ou du nombre d'habitants de la zone) :
- « 1 » correspond au niveau le plus fort «vulnérabilité forte »
- « 2 » signifie un niveau fort
- « 3 » correspond au niveau le moins fort
- « 4 » pas d'enjeu.

Pour l'intensité de phénomène dangereux, on a 3 niveaux : $1 \rightarrow 3$:
- « 1 » correspond au niveau le plus fort « intensité forte » (SEL)
- « 2 » signifie un niveau moins fort (SEI)
- « 3 » pas d'intensité.

Pour les niveaux des risques on 5 niveaux : $0 \rightarrow 4$:
- « R1 » correspond au niveau de risque le plus fort
- « R2 » signifie un niveau de risque intermédiaire
- « R3 » correspond au niveau le moins fort.

Il existe deux cas où le risque est considéré comme inexistant :
- R0 : pas des risques, ni intensité, ni enjeu
- R4 : le niveau d'intensité est 1 ou 2 mais l'enjeu est 4 (pas d'enjeu), ou bien, le niveau d'enjeu est 1,2 ou 3 et l'intensité est 3 (pas d'intensité)

Ce niveau est défini en fonction de la grille de cotation suivante :

	Intensité fort 1	Intensité moyen 2	Intensité faible
Niveau de vulnérabilité 1	R1	R2	R4
Niveau de vulnérabilité 2	R2	R3	R4
Niveau de vulnérabilité 3	R2	R3	R4
Niveau de vulnérabilité 4	R4	R4	R0

TABLE A.1: Les niveaux des risques

Le code de couleur des boîtes R4 est le suivant :

Niveau	Couleur
Niveau d'intensité =1	
Niveau d'intensité =2	
Niveau de vulnérabilité= 1	
Niveau de vulnérabilité= 2	
Niveau de vulnérabilité= 3	

TABLE A.2: Couleur des boites R4

Application sur le modèle de dispersion gaussien

Les figures A.2 et A.3 montrent les zones de danger calculées, suite à un accident TMD sur une carte représentant une partie de l'agglomération lyonnaise.

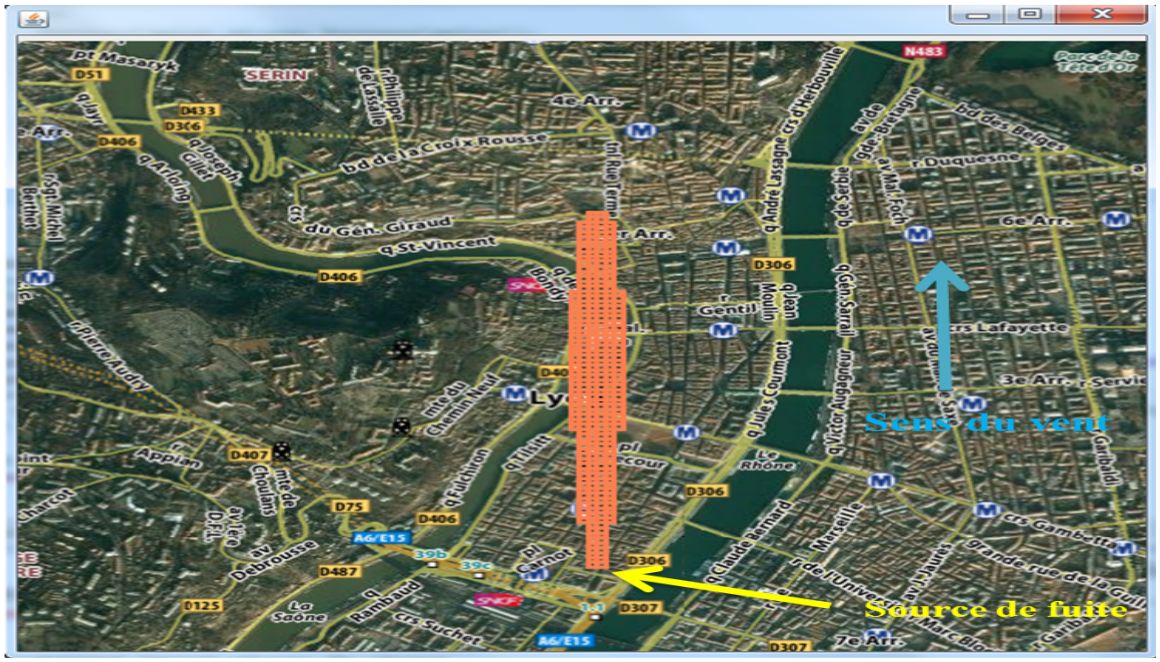

FIGURE A.2: Carte de niveau de risque

FIGURE A.3: Carte de niveau de risque

Application sur le modèle de zone majorante

La figure A.4 montre la zone de danger calculée en majorant, suite à un accident TMD sur une carte représentant une partie de l'agglomération grenobloise.

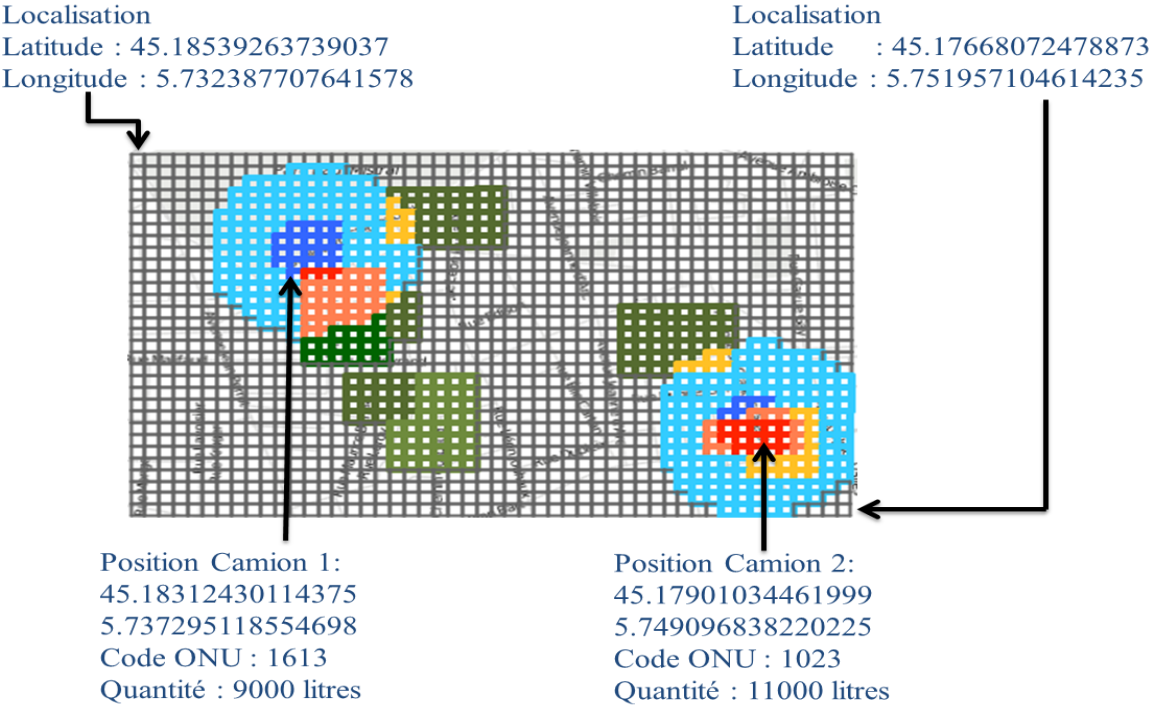

FIGURE A.4: Carte de niveau de risque

Références bibliographiques

[1] Turner D. B. Workbook of atmospheric dispersion estimates. *Public Health Service Publication n°999-Ap-26*, 1970.

[2] D Ermak. User's manual for slab : An atmospheric dispersion model for denser-than-air releases. *Lawrence Livermore National Laboratory*, UCRL-MA-105607, June 1990.

[3] Stéphane PAGNON. *Stratégies de modélisation des conséquences d'une dispersion atmosphérique de gaz toxique ou inflammable en situation d'urgence au regard de l'incertitude sur les données d'entrée.* PhD thesis, École Nationale Supérieure des Mines, Saint-Étienne, France, 2012.

[4] Ministère de l'Écologie et du Développement durable. *Circulaire du 10 mai 2010, récapitulant les règles méthodologiques applicables aux études de dangers, à l'appréciation de la démarche de réduction du risque à la source et aux plans de prévention des risques technologiques (PPRT) dans les installations classées en application de la loi du 30 juillet 2003.* Tech. Rep.

[5] Nishant PANDYA. *Analyse de sensibilité paramétrique d'un outil de modélisation des conséquences de scénarios d'accidents. Application à la dispersion atmosphérique de rejets avec le logiciel Phast.* PhD thesis, L'UNIVERSITÉ DE TOULOUSE, Toulouse, France, 2009.

[6] Emmanuel Demael. *Modélisation de la dispersion atmosphérique en milieu complexe et incertitudes associées.* PhD thesis, Ecole Nationale des Ponts et chaussées, 2007.

[7] ISO. risk management – principles and guidelines,. *ISO 31000*, 2009.

[8] Guide iso/cei 51. Safety aspects - guidelines for their inclusion in standards.

[9] ALE B. risk assessment practices in the netherlands. *Safety Science*, 40 :105–126, 2002.

[10] Jacques Charbonnier. *Le risk management : Méthodologie et pratiques.* Editions L'argus de l'assurance, paris, 2006.

[11] Pierre Périlhon. *La Gestion des risques - Méthode MADS-MOSAR II.* Editions Demos, paris, 2007.

[12] Jean-Marie Flaus. *Risk Analysis Socio-technical and Industrial Systems.* ISTE Ltd 2013, 27-37 St George' s Road London SW19 4EU UK, 2013.

[13] Ministère de l'Écologie et du Développement durable, direction de la Prévention des pollutions et des risques, sous-direction de la Prévention des risques majeurs. *Le transport de matiéres dangereuses.* Tech. Rep, Décembre 2002.

[14] IRMA Institut des risques majeurs. Le risque de transport de matiéres dangereuses. URL http://www.irma-grenoble.com/PDF/05documentation/brochure/risques_majeurs2007/12Risque_Transport.pdf.

[15] Risques majeurs. Le risque de transport de matiéres dangereuses, september 2009. URL http://www.risquesmajeurs.fr/le-risque-de-transport-de-matieres-dangereuses.

[16] Congrès ATEC ITS. Les rencontres de la mobilité intelligente. France, 2014.

[17] Santa Cruz A. S. M. Godoy, S. M. and N. J. Scenna. Strrap system – a software for hazardous materials risk assessment and safe distances calculation. *Reliability Engineering and System Safety*, 92 :847–857, 2007.

[18] Angela Maria Tomasoni. *Modèles et méthodes d'évaluation et de gestion des risques appliqués aux systèmes de transport de marchandises dangereuses(TMD) reposant sur les nouvelles technologies de l'information et de la communication (NTIC).* PhD thesis, l'École nationale supérieure des mines de Paris, Paris, France, 21 avril 2010.

[19] Eleveld H. Kok Y. S. and Twenhöfel C. J. W. Sensitivity and uncertainty analyses of the atmospheric dispersion model npk-puff. *9th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, page 79 – 83, 2004.

[20] Doury.A. Le vademecum des transferts atmosphériques. *Rapport DSN n°440*, 1987.

[21] Hanna S.R. et. al. Handbook on atmospheric diffusion. *Technical Information Center. U.S. Department of Energy*, 1982.

[22] Pasquill F. Atmospheric diffusion. *Ellis Horwood*, 1974.

[23] O. Zeman. The dynamics and modeling of heavier-than-air, cold gas releases. *Atmospheric Environment (1967)*, 1982, 16, 741 - 751.

[24] D Ermak and S. Chan. A study of heavy gas effects on the atmospheric of dense gases. *Lawrence Livermore National Laboratory*, UCRL-92494, Rev. 1, April 1985.

[25] Donald L. Ermak. *USER'S MANUAL FOR SLAB : AN ATMOSPHERIC DISPERSION MODEL FOR DENSER-THAN-AIR RELEASES.* Physics Department, Atmospheric and Geophysical Sciences Division University of California, Lawrence Livermore National Laboratory Livermore, California 94550, june 1990.

[26] G.Morgan and M.Henrion. *Uncertainty. A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Addison-Wesley, Cambridge University Press, Cambridge, 1990.

[27] J. GENTLE. Random number generation and monte carlo methods. *Springer, New York, seconde édition*, 2003.

[28] R. E. Moore. *Methods and Applications of Interval Analysis.* SIAM, Philadelphia, PA, 1979.

[29] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.

[30] D. Dubois and H. Prade. Fussy sets and systems-theory and applications. *Academic Press New York, NY*, (3) :159–176, 1980.

[31] Bourgine B. Guyonnet D. and Chilès J.-P. Prise en compte de l'incertitude dans l'évaluation du risque d'exposition aux polluants du sol. *Programme GESSOL. Rapport final. Rapport BRGM RP-51683-FR*, 2002.

[32] Dubois D. Fargier H.-Côme B. Guyonnet D., Bourgine B. and Chilès J.-P. Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 129, 2003.

[33] Ferson S. andGinzburg L. Different methods are needed to propagate ignorance and variability. *Reliability and System Safety*, 54 :133–144, 1996.

[34] P. GLASSERMAN. Monte carlo methods in financial engineering. *Springer*, 2004.

[35] B. M. AYYUB and G. J. KLIR. *Uncertainty modeling and analysis in engineering and the sciences.* CRC Press, 2006.

[36] I. A. MacDonald. *Quantifying the effects of uncertainty in building simulation.* PhD thesis, Université de Strathclyde, 2002.

[37] R. HEIJUNGS. Representing statistical distributions for uncertain parameters in lca. relationships between mathematical forms, their representation in ecospold, and their representation in cmlca. *The International Journal of Life Cycle Assessment, 10(4), 248–254*, 2005.

[38] Arnold Neumaier. *Interval methods for systems of equations*. Cambridge University press, 1990.

[39] Moore R.E. Interval analysis. *Prentice Hall, Englewood Cliffs, New Jersey*, 1966.

[40] Hansen E. R. Global optimization using interval analysis. *Marcel Dekker, New York*, 1992.

[41] Didrit O. *Analyse par intervalles pour l'automatique ; résolution globale et garantie de problèmes non-linéaires en robotique et en commande robuste*. PhD thesis, Paris XI Orsay, 1997.

[42] Piet-Lahanier H. Milanese M., Norton J. and Walter E. Bounding approches to system identification. *Plenum Press, New York and London*, 1996.

[43] Walter E. and Piet-Lahanier H. Estimation of parameter bounds from bounded-error data. *a survey.Math. and Comput. in Simul*, 32 :449–468, 1990.

[44] Flaus J-M. Adrot O. and Ragot J. Estimation of bounded model uncertainties. *Journal of Robotics and Mechatronics*, 18(5), 2006.

[45] Blesa J. *Robust Identification and Fault Diagnosis using Set-membership Approaches*. PhD thesis, Universitat Politecnica de Catalunya, 2011.

[46] Didrit O. Jaulin L., Kiefer M. and Walter E. Applied interval analysis. *Springer-Verlag, London*, 2001.

[47] Hadj-Sadok M.Z. and Gouzé J.L. Estimation of uncertain models of activated sludge process with interval observers. *Journal of Process Control*, 11(3) :299–310, 2001.

[48] Jaulin L. and Walter E. Guaranteed nonlinear parameter estimation from bounded-error data via interval analysis. *Mathematics and Computers in Simulation*, (35) : 123–137, 1993.

[49] Jaulin L. and Walter E. Guaranteed tuning, with application robust control and motion planning. *In Automatica*, 32 :1217–1221, 1996.

[50] Flaus J-M. and Adrot O. Estimation d'état sûre pour procédés non linéaires par méthodes ensemblistes – application à un procédé biotechnologique. *Journal Européen des Systèmes Automatisés (JESA)*, 37(9/2003, Edition Hermes Lavoisier.), 2003.

[51] Adrot O. and Flaus J-M. Fault detection based on uncertain switching models with bounded parameters. *Safeprocess2009, 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, 30th June – 3rd July 2009.

[52] L. Travé-massuyès J. Vehi Armengol, J. and M. A. Sainz. Semiqualitative simulation using modal interval analysis. *14th World Congress of International Federation of Automatic Control*, Beijing, China, 1999.

[53] J. Quevedo-T. Escobet Puig, V. and S. de las Heras. Passive robust fault detection approaches using interval models. *15th Triennal IFAC World Congress (IFAC, Ed.)*, Barcelona, Spain, 2002.

[54] O. Adrot Ploix S. and J. Ragot. Bounding approach to the diagnosis of uncertain static systems. *IFAC Safeprocess*, 2000.

[55] Ploix S. and Adrot O. Parity relations for uncertain dynamic systems. *Automatica*, 42 :1553–1562, 2006.

[56] Adrot O. and Flaus J.M. Fault detection based on uncertain models with bounded parameters and bounded parameter variations. *17th IFAC World Congress 2008*, Seoul Korea, 2008.

[57] Combastel C. and Raka S.A. A set-membership fault detection test with guaranteed robustness to parametric uncertainties in continous time linear dynamical systems. *7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, 2009.

[58] Sarrate R. Ocampo-Martinez C. Puig V., Escobet T. and Tornil-Sin S. Robust fault diagnosis of non-linear systems using constraints satisfaction. *Safeprocess'09, 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, Spain, 2009.

[59] Adrot O. and Flaus J-M. Guaranteed fault detection based on interval constraint satisfaction problem. *Conference on Control and Fault-Tolerant Systems, Systol'10*, pages 708–713, France, 2010.

[60] Shahriari K. *Analyse de Sûreté de Procédés Multi-Modes par des Méthodes à base d'intervalles.* PhD thesis, Université Joseph Fourier, 2007.

[61] Raka S.A. *Méthodes et outils ensemblistes pour le pré-dimensionnement de systèmes mécatroniques.* PhD thesis, Université de Cergy-Pontoise, 2011.

[62] Saltelli and Al. *Sensitivity Analysis in Practice a Guide to Assessing Scientific Models.* Wiley, 2004.

[63] B. Bettonvil and J. P. Kleijnen. Searching for important factors in simulation models with many factors : Sequential bifurcation. *European Journal of Operational Research, 1997, 96, 180-194.*

[64] F. Campolongo and R. Braddock. The use of graph theory in the sensitivity analysis of the model output : a second order screening method. *Reliability Engineering and System Safety*, 64 :1–12, 1999.

[65] A. M. Deana and S. M. Lewis. Comparison of group screening strategies for factorial experiments. *Computational Statistics and Data Analysis, 2002, 39, 287 – 297.*

[66] M. D. Morris. Input screening : Finding the important model inputs on a budget. *Reliability Engineering and System Safety, 2006, 91, 1252-1256.*

[67] J. Campolongo, F. Cariboni and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modeling and Software*, 22 : 1509 – 1518, 1999.

[68] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33 :161–174, 1991.

[69] T. Turányi and K. Chan-E. S. (ed.) Rabitz, H. A. ; Saltelli. Local methods in sensitivity analysis. *Wiley, Chichester, 2000, p79-99.*

[70] Julien Jacques. *« Pratique de l'analyse de sensibilité : comment évaluer l'impact desentrées aléatoires sur la sortie d'un modèle mathématique », extrait de la thèse de l'auteur intitulé « Contributions à l'analyse de sensibilité et à l'analyse discriminante généralisée ».* PhD thesis, Université Joseph Fourier, Grenoble, France, 2011.

[71] I.M. Sobol. Global sensitivity for non linear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulations*, 55 :271–280, 2001.

[72] H. Niederreiter. Random number generation and quasi-monte carlo methods. *SIAM*, 1992,247 pages.

[73] A. Owen. Monte carlo extension of quasi-monte carlo winter. *Simulation Conference,*, Washington (DC, USA), 1988.

[74] R. McKay, M. Beckman and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21 :239–245, 1979.

[75] M.D.McKay. Evaluating prediciton uncertainty. *Technical Report NUREG/CR-6311, US Nuclear Regulatory Commission and Los Alamos National Laboratory*, 21, 1995.

[76] K.E. Shuler A.G. Petschek R.I. Cukier, C.M. Fortuin and J.H. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients - theory. *Journal Chemical Physics*, 59 :3873–3878, 1973.

[77] R.I. Levine R.I. Cukier and K.E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal Computational Physics*, 26 :1–42, 1978.

[78] K.E. Shuler R.I. Cukier and J.H. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients - analysis of the approximations. *Journal Chemical Physics*, 63 :1140– 1149, 1975.

[79] J.H. Schaibly and K.E. Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficeints. *Journal Chemical Physics*, 59 :3879–3888, 1973.

[80] Metropolis N. and Ulam S.M. The monte carlo method. *Journal of the American Statistical Association*, 44 :335–341, 1949.

[81] Hammersley J.M.and Morton K.W. A new monte carlo technique : Antithetic variates. *Proceedings of the Cambridge Philosophical Society*, 52 :449–475, 1956.

[82] Hammersley J. M. and Handscomb D.C. Monte carlo methods. Methuen, London, 1964.

[83] Haber S. A modified monte-carlo quadrature. *Mathematics of Computation*, 20 : 361–368, 1966.

[84] Kuipers L. and Niederreiter H. Uniform distribution of sequences. *John Wiley and Sons*, 1974.

[85] Boyle P.P. Option : a monte carlo approach. *Journal of Financial Economics*, 4 : 323–338, 1977.

[86] Pierre-Alain Patard. *Essais sur les méthodes de simulation numérique et sur la modélisation des données de marché.* PhD thesis, Université claude bernard Lyon 1 institue de science financiére et d'assurences, Lyon, 2008.

[87] Richtmyer R.D. On the evaluation of definite integrals and a quasimonte carlo method based on properties of algebraic numbers. *Report LA- 1342, Los Alamos Scientific Laboratories*, 1951.

[88] Halton J.H. On the effciency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik 2*, pages 84–90, 1960.

[89] Haber S. Numerical evaluation of multiple integrals. *SIAM Review*, 12 :481–526, 1970.

[90] Niederreiter H. On a number-theoretical integration method. *Aequationes Mathematicae 8*, pages 304–311, 1972.

[91] Cranley R. and Patterson T.N.L. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13 :904–914, 1976.

[92] Faure H. Discrépance de suites associées à un système de numération (en dimension 1). *Bulletin de la S.M.F.*, 109 :143–182, 1981.

[93] Faure H. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arithmetica*, 41 :337–351, 1982.

[94] Zinterhof P. Gratis lattice points for multidimensional integration. *Computing*, 38 :347–353, 1987.

[95] Finschi L. Quasi-monte carlo : An empirical study on low-discrepancy sequences. *Working Paper, Eidgenossische Technische Hochschule Zürich*, 1996.

[96] Pagès G. and Xiao Y.-J. Sequences with low discrepancy and pseudorandom numbers : theoretical results and numerical tests. *Journal of Statistical Computation and Simulation*, 56 :163–188, 1997.

[97] Tuffin B. *Simulation accélérée par les méthodes de Monte Carlo et quasi-Monte Carlo : théorie et applications.* PhD thesis, Université Rennes 1, Rennes, 1997.

[98] Nacim MESLEM. *Atteignabilité hybride des systèmes dynamiques continus par analyse par intervalles. Application à l'estimation ensembliste.* PhD thesis, L'Université Paris Est, Paris, France, 2008.

[99] Rania MERHEB. *FIABILITÉ DES OUTILS DE PRÉVISION DU COMPORTEMENT DES SYSTÈMES THERMIQUES COMPLEXES.* PhD thesis, L'UNIVERSITÉ DE BORDEAUX 1, BORDEAUX, France, 2013.

[100] Massa DAO. *Caractérisation d'ensembles par des méthodes intervalles. Applications en automatique.* PhD thesis, UNIVERSITÉ D'ANGERS ISTIA, Angers, France, 2006.

[101] Tarek Raissi. *Méthodes ensemblistes pour l'estimation d'état et de paramètres.* PhD thesis, L'UNIVERSITE PARIS XII VAL de MARNE, Paris, France, 2004.

[102] R. Krawczyk and A. Neumaier. Interval slopes for rational functions and associated centered forms. *SIAM Journal of Numerical Analysis*, 22 :604–616, 1985.

[103] E. R. Hansen and R. I. Greenberg. An interval newton method. *Applied Mathematical Computing*, 12 :89–98, 1983.

[104] A. Neumaier. Taylor forms - use and limits. *Reliable Computing*, 9 :43–79, 2003.

[105] Committee the Prevention of Disasters. *Methods for the calculation of physical effects due to releases of hazardous materials-yellow book.* Netherlands, 2005.

[106] Documentation INERIS. *Méthodes pour l'évaluation et la prévention des risques accidentels (DRA-006).* Tech. Rep, 2006.

[107] Documentation INERIS. *Emissions accidentelles substances chimiques dangereuses dans l'atmosphére seuils de toxicité aigue.* INERIS–DRC-08-94398-12846A.

[108] Leeming D.G. and Saccomanno F.F. Use of quantified risk assessment in evaluating the risks of transporting chlorine by road and rail. *Transportation Research Record*, n°1430,27-35, 1994.

[109] P.Meseguer. Constraint satisfaction problems : An overview. *AI Commun*, 2(1) :3–17, 1989.

[110] Puig V Tornil-Sin S, Ocampo-Martinez C and Escobet T. Robust fault detection of no-linear systems using set member ship state estimation based on constraint satisfaction. *Engineering Applications of Artificial interlligence*, pages 25(1) :1–10, 2012.

[111] C. Toumazou B. Delaunay P. Herrero, P. Georgiou and L. Jaulin. An efficient implementation of sivia algorithm in a high-level numerical programming language. *Reliable computing*, pages 239–251, 2012.

[112] L. Zadeh. Quantitative fuzzy semantics. *Information Sciences*, (3) :159–176, 1971.

[113] Toby Walsh. Stochastic constraint programming. *In Proceedings of the 15th ECAI, European Conference on Artificial Intelligence. IOS*, page 111–115, Press, 2002.

[114] L. Jaulin and E. Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4) :1053–1064, 1993.

[115] R. E. Moore. Parameter sets for bounded-error data. *Mathematics and Computers in Simulation*, 34(2) :113–119, 1992.

[116] L. Jaulin and B. Desrochers. Introduction to the algebra of separators with application to path planning. *Engineering Applications of Artificial Intelligence*, 33 : 141–147, 2014.

[117] R. Sarrate C. Ocampo-Martinez V. Puig, T. Escobet and S. Tornil-Sin. Robust fault diagnosis of non-linear systems using constraints satisfaction. *Safeprocess'09, 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Spain, 2009.

[118] R. Krzysztof. Constraint programming. *Cambridge University Press*, 2003.

[119] I. Braems. *Méthodes ensemblistes garanties pour l'estimation de grandeurs physiques.* PhD thesis, University Paris XI Orsay, Paris, France, 2002.

[120] Puig V Tornil-Sin S, Ocampo-Martinez C and Escobet T. Robust fault detection of no-linear systems using set member ship state estimation based on constraint satisfaction. *Engineering Applications of Artificial Intelligence*, 25 :1–10, 2012.

[121] Puig V Tornil-Sin S, Ocampo-Martinez C and Escobet T. Robust fault detection of no-linear systems using interval constraint satisfaction and analytical relations. *IEEE Transactions on systems, Man, and cybernetics systems*, pages 1–12, 2013.

[122] O. Adrot and S. Ploix. Fault detection based on set-membership inversion. *Safeprocess'06, 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Beijing, China 2006.

[123] P. Frank. Fault diagnosis in dynamic systems using analytical and knowledge–based redundancy – a survey and some new results. *Automatica J. IFAC*, 26(3) :459–474, 1990.

[124] R. Clark. Instrument fault detection. *IEEE Transactions on Aerospace and Electronic Systems*, 14 :456–465, 1978.

[125] Sobol I.M. « sensitivity estimates for non linear mathematical models ». *Mathematical Modelling and Computational Experiments, vol*, .1 (4), 1993.

[126] K. Chan A. Saltelli and E.M. Scott. *Sensitivity Analysis.* Wiley, 2000.

[127] Z. Ferenczi. Sensitivity analysis of the mesoscale puff model– rimpuff. *Proceedings of 10th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, page 599 – 603, 2005.

[128] R. Bubbico and Mazzarotta B. Accidental release of toxic chemicals : Influence of the main input parameters on consequence calculation. *Journal of Hazardous Materials*, 151 :394 – 406, 2008.