Sensitivity analysis of spatial models: application to cost-benefit analysis of flood risk management plans

Nathalie Saint-Geours

▶ **To cite this version:**

Nathalie Saint-Geours. Sensitivity analysis of spatial models: application to cost-benefit analysis of flood risk management plans. Applications [stat.AP]. Université Montpellier II - Sciences et Techniques du Languedoc, 2012. English. NNT : . tel-00761032v2

HAL Id: tel-00761032

https://theses.hal.science/tel-00761032v2

Submitted on 4 Dec 2012

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

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

THÈSE

pour obtenir le grade de docteur délivré par

Université Montpellier II

École doctorale I2S
Spécialité Biostatistiques

Analyse de sensibilité de modèles spatialisés

Application à l'analyse coût-bénéfice de projets de prévention des inondations

---

Nathalie SAINT-GEOURS

Soutenue le 29/11/2012 devant le jury composé de :

M. Christian LAVERGNE	Professeur, Université Montpellier III	Directeur de thèse
M. Hervé MONOD	Directeur de recherche, INRA	Rapporteur
M. Jeroen AERTS	Professeur, IVM (Pays-Bas)	Rapporteur
M. Stefano TARANTOLA	Chercheur, EC - Joint Research Center (Italie)	Examinateur
Mme. Bénédicte BUCHER	Directrice de recherche, IGN	Examinatrice
M. André MAS	Professeur, Université Montpellier II	Examinateur
M. Jean-Stéphane BAILLY	Enseignant-chercheur, AgroParisTech	Encadrant
M. Frédéric GRELOT	Chercheur, Irstea	Encadrant

Thèse préparée conjointement au sein de l'Unité Mixte de Recherche Territoires, Environnement, Télédection et Information Spatiale
et de l'Institut de Mathématiques et de Modélisation de Montpellier

Analyse de sensibilité de modèles spatialisés : application à l'analyse coût-bénéfice de projets de prévention des inondations

L'analyse de sensibilité globale basée sur la variance permet de hiérarchiser les sources d'incertitude présentes dans un modèle numérique et d'identifier celles qui contribuent le plus à la variabilité de la sortie du modèle. Ce type d'analyse peine à se développer dans les sciences de la Terre et de l'Environnement, en partie à cause de la dimension spatiale de nombreux modèles numériques, dont les variables d'entrée et/ou de sortie peuvent être des données distribuées dans l'espace. Le travail de thèse réalisé a pour ambition de montrer comment l'analyse de sensibilité globale peut être adaptée pour tenir compte des spécificités de ces modèles numériques spatialisés, notamment la dépendance spatiale dans les données d'entrée et les questions liées au changement d'échelle spatiale. Ce travail s'appuie sur une étude de cas approfondie du code NOE, qui est un modèle numérique spatialisé d'analyse coût-bénéfice de projets de prévention du risque d'inondation. On s'intéresse dans un premier temps à l'estimation d'indices de sensibilité associés à des variables d'entrée spatialisées. L'approche retenue du « *map labelling* » permet de rendre compte de l'auto-corrélation spatiale de ces variables et d'étudier son impact sur la sortie du modèle. On explore ensuite le lien entre la notion d'« *échelle* » et l'analyse de sensibilité de modèles spatialisés. On propose de définir les indices de sensibilité « *zonaux* » et « *ponctuels* » pour mettre en évidence l'impact du support spatial de la sortie d'un modèle sur la hiérarchisation des sources d'incertitude. On établit ensuite, sous certaines conditions, des propriétés formelles de ces indices de sensibilité. Ces résultats montrent notamment que l'indice de sensibilité zonal d'une variable d'entrée spatialisée diminue à mesure que s'agrandit le support spatial sur lequel est agrégée la sortie du modèle. L'application au modèle NOE des méthodologies développées se révèle riche en enseignements pour une meilleure prise en compte des incertitudes dans les modèles d'analyse coût-bénéfice des projets de prévention du risque d'inondation.

Mots clés : Modèle Spatialisé ; Analyse de Sensibilité ; Incertitude ; Échelle ; Géostatistique ; ACB ; Inondations ; Dommages.

Sensitivity analysis of spatial models: application to cost-benefit analysis of flood risk management plans

Variance-based global sensitivity analysis is used to study how the variability of the output of a numerical model can be apportioned to different sources of uncertainty in its inputs. It is an essential component of model building as it helps to identify model inputs that account for most of the model output variance. However, this approach is seldom applied in Earth and Environmental Sciences, partly because most of the numerical models developed in this field include spatially distributed inputs or outputs . Our research work aims to show how global sensitivity analysis can be adapted to such spatial models, and more precisely how to cope with the following two issues: i) the presence of spatial auto-correlation in the model inputs, and ii) the scaling issues. We base our research on the detailed study of the numerical code NOE, which is a spatial model for cost-benefit analysis of flood risk management plans. We first investigate how variance-based sensitivity indices can be computed for spatially distributed model inputs. We focus on the "*map labelling*" approach, which allows to handle any complex spatial structure of uncertainty in the model inputs and to assess its effect on the model output. Next, we offer to explore how scaling issues interact with the sensitivity analysis of a spatial model. We define "*block sensitivity indices*" and "*site sensitivity indices*" to account for the role of the spatial support of model output. We establish the properties of these sensitivity indices under some specific conditions. In particular, we show that the relative contribution of an uncertain spatially distributed model input to the variance of the model output increases with its correlation length and decreases with the size of the spatial support considered for model output aggregation. By applying our results to the NOE modelling chain, we also draw a number of lessons to better deal with uncertainties in flood damage modelling and cost-benefit analysis of flood risk management plans.

Key words: Spatially distributed model; Sensitivity analysis; Uncertainty; Scale; Geostatistics; CBA; Flood; Damage.

AgroParisTech - UMR TETIS (AgroParisTech, Cirad, Irstea)
Maison de la télédétection - 500 rue Jean-Francois Breton 34093 Montpellier Cedex 5

« La connaissance est une navigation dans un océan d'incertitudes
à travers des archipels de certitudes. »
Edgar Morin (1921 -)

Les passes du Bassin d'Arcachon

À mon grand-père paternel.

Notes au lecteur

Franglais - Frenchglish

Afin d'en permettre la lecture à l'ensemble des membres du jury, la majeure partie de ce manuscrit de thèse a été rédigée en anglais. L'introduction est présente en deux versions, en français et en anglais. Seuls le préambule et certaines des annexes sont uniquement rédigés en français.

Most of this thesis manuscript is written in English, so that all the thesis committee members can read it. Only the preamble and some of the appendices are written in French.

Publications

Ce manuscrit de thèse est pour partie composé de publications dont une est parue dans « Mathematical Geosciences » (section §4.1) et une autre est en cours d'évaluation par « Journal of Flood Risk Management » (section §3.3). Un papier court publié dans les actes de la conférence « Accuracy 2010 » est aussi intégré au manuscrit (section §3.2.4). D'autres publications, non intégrées au corps du document, ont été en revanche insérées en annexe pour information. Pour vous procurer les versions finales de ces diverses publications, vous pouvez m'envoyer un courriel.

This thesis manuscript contains two publications, one of which has already been published in « Mathematical Geosciences » (section §4.1), and the second one is currently being reviewed by « Journal of Flood Risk Management » (section §3.3). A short paper published in the proceedings of the « Accuracy 2010 » international conference is also included (section §3.2.4). Other publications were inserted in the Appendices for information. If you want a copy of the final versions of these papers, please send me an email and I will keep you updated on the publication progress.

Remerciements

C'est avec grand plaisir que je tiens à remercier tous ceux qui ont permis à ce travail de thèse d'aboutir.

Mes premiers remerciements s'adressent à mes trois encadrants, qui ont su guider l'élaboration et la réalisation de cette thèse depuis le début ; leurs compétences professionnelles et leurs qualités humaines m'ont permis de réaliser ce travail dans d'excellentes conditions. Un grand merci à Jean-Stéphane Bailly pour sa très grande disponibilité, son écoute et son soutien tout au long de ces trois ans. Merci de m'avoir initiée aux mystères de la géostatistique. Un grand merci également à Frédéric Grelot, pour ses conseils avisés et son exigence, et pour m'avoir fait découvrir le vaste monde des inondations. Merci aussi à vous deux pour votre relecture attentive du présent manuscrit et pour m'avoir donné l'occasion de me frotter à la pratique passionnante de l'enseignement. Je remercie enfin Christian Lavergne pour avoir accepté d'encadrer cette thèse, pour son suivi régulier de mon travail et pour m'avoir ouvert les portes de la communauté des statisticiens.

Je suis aussi reconnaissante envers Pierre Bazile, Cyril Kao et la direction scientifique de l'Engref, pour leur soutien et leurs conseils avisés lors du montage de ce projet de Formation Complémentaire Par la Recherche.

Je tiens ensuite à remercier Hervé Monod et Jeroen Aerts d'avoir apporté leur regard critique sur cette thèse en acceptant d'en être les rapporteurs. Mes remerciements vont également à Bénédicte Bucher, Stefano Tarantola et André Mas pour avoir eu l'amabilité de participer au jury de soutenance.

Plus largement, je tiens aussi à adresser mes remerciements à tous ceux qui, rencontrés au détour d'une école-chercheur, d'un séminaire ou d'une conférence, ont contribué à enrichir cette thèse par leur remarques, leurs suggestions et nos discussions. Ainsi, ce travail de recherche s'est largement nourri de ma participation au réseau MEXICO (www.reseau-mexico.fr) ainsi qu'au GDR MASCOT-NUM (www.gdr-mascotnum.fr), tous deux consacrés aux méthodes d'analyse de sensibilité et d'exploration numérique des modèles. Je tiens particulièrement à remercier Bertrand Iooss qui m'a ouvert les portes de ces deux réseaux dès le début de ma thèse. Merci également à Clémentine Prieur, pour l'intérêt qu'elle a porté à mon travail et pour m'avoir invitée à présenter mes travaux à Grenoble et Clermont-Ferrand . Mes remerciements vont aussi à Claude Bruchou qui m'a permis de participer au groupe de lecture en analyse de sensibilité de l'INRA d'Avignon. Merci enfin à Mathieu Ribatet, Gabriel Lang, Bertrand Iooss et Clémentine Prieur pour leur participation à mes comités de thèse.

Je remercie vivement Andrea Saltelli, Stefano Tarantola ainsi que toute l'équipe EAS du Joint Research Center à Ispra pour m'avoir accueillie pendant deux mois à l'automne 2011. Mon séjour transalpin a été marqué par la qualité de nos échanges conviviaux, qui donneront lieu je l'espère à d'autres collaborations dans un futur proche. Un grand merci à Ine, James et William pour m'avoir très vite intégrée au petit monde du JRC. Merci enfin à l'Irstea, au réseau Peer et au GDR MASCOT-NUM qui ont financé ce séjour.

Mes remerciements s'adressent également à la Mission Rhône, qui a financé une partie de mes travaux. Merci en particulier à Mathieu Metral qui a montré un intérêt constant pour la démarche d'analyse de sensibilité des évaluations économiques du risque d'inondation, et qui nous a permis d'appliquer nos développements théoriques à plusieurs terrains d'étude sur la basse vallée du Rhône. Merci également au Syndicat Mixte de la Vallée de l'Orb et du Libron.

C'est grâce au projet mené sur la basse vallée du Rhône que j'ai eu la chance de travailler avec Thibaud Langer, d'abord dans le rôle nouveau pour moi et enrichissant d'encadrant de stage, puis en tant que collègue. Merci Thibaud pour ton dynamisme et ton envie d'apprendre ! Ton arrivée à mi-parcours a constitué une belle bouffée d'air frais et m'a permis de me sortir un temps de mon quotidien de thésarde solitaire.

Merci aussi à Eric Gilbert d'avoir consacré du temps pour échanger sur mes problématiques de recherche et d'avoir partagé un peu de son expérience professionnelle sur l'évaluation économique du risque d'inondation. Ces discussions m'ont notamment permis de mieux situer la portée opérationnelle et les limites de ce travail de thèse.

Un immense merci bien sûr à toute l'équipe de l'UMR TETIS et plus largement à tous les habitants de la Maison de la Télédétection. L'excellente ambiance qui règne en ces lieux a largement contribué à rendre ces trois années très agréables et pleines de bons souvenirs. Merci à tous pour votre accueil, votre soutien à la fois professionnel et amical.

Merci également à l'équipe EPS de l'Institut de Mathématiques et de Modélisation de Montpellier pour leur accueil chaleureux lors de mes passages réguliers. Je remercie notamment Bernadette Lacan, pour son suivi efficace des tracasseries administratives qui sont le lot des doctorants.

Merci aussi aux collègues de G-EAU, notamment Pauline pour ta relecture attentive du Chapitre 2.

Un salut plus personnel enfin à tous ceux avec qui j'ai partagé mon quotidien durant ces trois ans. À la MTD d'abord : Gab asa Mr Poujol, Tristan, Isa, Élo, Valou, Marc, Thibaud, Roberto, Clothilde, Jean, Mickael et j'en oublie bien d'autres ; à défendre les couleurs de la MTD sur les terrains de volley (la victoire cette année ?) : Claire, Silvan, Élodie, Christophe, Thomas, Mathieu, Mohammad, Cédric, Jeff, Jean-Stéphane, Gab, Soizic… à partager les joies de la colocation (à qui je souhaite encore longue vie !) : Joana, Lydie, Nono, Grégoire, le Balthazar Circus péruvien, Fadi, Christophe, Hélène, Fabrice, Erwan, Chloé, Lisa, Olivier, Damien, Carla ; à partager les joies de la colocation de bureau, Kenji ; à partager enfin tous les bons moments de la vie montpelliéraine en dehors du travail : Thomas, Jeanne, Elo, Valou, Constance, Teddy, Claire, Damien, Carla ; merci enfin à Julie et François pour la pause randonnée basque en pleine période de rédaction, j'en avais grand besoin !

Merci aussi à mes parents pour m'avoir inculqué le goût des études, du travail et une certaine curiosité sans lesquels cette thèse n'aurait jamais vu le jour. Merci à ma soeur et à mon frère pour leur sarcastique soutien. Merci enfin à Emma dont l'arrivée et les gazouillis ont égayé ma fin de thèse.

Enfin, un merci spécial à Laure pour m'avoir accompagnée pendant ces quatre dernières années, en toute sécurité bien sûr !

Contents

Table of contents **xi**

List of figures **xiii**

List of tables **xvii**

Acronyms **xix**

Notations **xxi**

Introduction (français) **1**

Introduction (english) **9**

1 Theoretical background **15**

1.1 Spatial models 18

1.2 Sensitivity analysis 24

1.3 Chapter conclusion 33

2 Building a modelling framework for cost-benefit analysis of flood management plans **35**

2.1 Cost-benefit analysis of flood risk management plans 38

2.2 NOE modelling framework for cost-benefit analysis based on avoided damages 40

2.3 Case studies 58

2.4 Chapter conclusion 72

3 Spatially distributed inputs in variance-based global sensitivity analysis 75

3.1 State of the art and comparison of the available methods 79

3.2 Sampling issues 98

3.3 Application to the NOE code on the Orb Delta case study 111

3.4 Chapter conclusion 133

4 Scale issues in variance-based global sensitivity analysis 137

4.1 Influence of support 140

4.2 Influence of spacing 153

4.3 Application to the NOE code on the Orb Delta case study 156

4.4 Chapter conclusion 169

Conclusion and prospects 173

References 183

List of publications 195

Appendices 199

List of Figures

1 Structure générale de la thèse . . . 7
2 Outline of the thesis . . . 14

1.1 Spatial domain Ω . . . 19
1.2 A spatial model . . . 19
1.3 Representation of spatially distributed data . . . 20
1.4 Three alternative definitions of scale in space . . . 21
1.5 Upscaling as a two-steps procedure . . . 22
1.6 A spatially additive model . . . 23
1.7 A point-based model . . . 24
1.8 Flowchart for sensitivity analysis . . . 25

2.1 Flood-prone study area: present and future situation . . . 41
2.2 General flowchart of the NOE modelling framework . . . 43
2.3 Flood-prone study area: reduction of flood damages . . . 44
2.4 Flood-prone study area: a flooding event . . . 46
2.5 Average Annual Damages in unidimensional case . . . 47
2.6 Flood-prone study area: set of flood scenarios . . . 49
2.7 Approximation of the AAD indicator with trapezoidal rule (unidimensional case) . . . 50
2.8 Flowchart of the NOE modelling framework: flood damage estimation . . . 52
2.9 Flood-prone study area: inundation process for a given flood scenario . . . 53
2.10 Flood-prone study area: spatially explicit description of assets . . . 54
2.11 Two assets maps with different spatial support on two study sites . . . 55
2.12 Flood-prone study area: flood exposure . . . 56

2.13 Depth-damage curve: example for wheat crop . . . 57
2.14 Flood-prone study area: damage costs estimation for a given flood scenario . . . 57
2.15 The Orb Delta study site . . . 59
2.16 Orb Delta case study: structural flood-control measures . . . 60
2.17 Orb Delta case study: observed Q-f data and fitted discharge-frequency curve . . . 62
2.18 Orb Delta case study: hazard maps . . . 63
2.19 Orb Delta case study: assets map . . . 64
2.20 Orb Delta case study: damage map $\mathcal{D}(e_2)$ (zoom) . . . 66
2.21 From vector map of damages to raster map of damages . . . 66
2.22 Orb Delta case study: raster maps of the AAD indicator . . . 67
2.23 Orb Delta case study: raster maps of the ΔAAD indicator . . . 68
2.24 Orb Delta case study: nominal results for total flood damage estimates $D(e_j)$ . . . 69

3.1 VB-GSA with spatially distributed inputs: macro-parameter method . . . 80
3.2 VB-GSA with spatially distributed inputs: dimension reduction method . . . 82
3.3 VB-GSA with spatially distributed inputs: trigger input method . . . 84
3.4 VB-GSA with spatially distributed inputs: map labelling method . . . 85
3.5 VB-GSA with spatially distributed inputs: second level method . . . 87
3.6 Three random realisations of the $\phi_{\text{camp.}}$ function . . . 89
3.7 Decision tree for VB-GSA with spatially distributed inputs; dashed boxes indicate that the method can be coupled with meta-modelling . . . 96
3.8 Map labelling approach: convergence of sensitivity indices estimates . . . 102
3.9 Study area represented by a discrete grid $\mathcal{G}$ . . . 104
3.10 Three simulations of Gaussian random field by Simple Random Sampling . . . 106
3.11 First and total-order sensitivity indices of input factors Z and $\mathbf{U}$ . . . 108
3.12 Average semivariograms for exponential model, SRS or LHS sets of maps . . . 109
3.13 VB-GSA of NOE model: simplified flowchart . . . 114
3.14 VB-GSA of NOE model: computation of the ΔAAD indicator . . . 117
3.15 VB-GSA of NOE model: flowchart for uncertainty and sensitivity analysis . . . 119
3.16 VB-GSA of NOE model: annual maximum flow series AMFS 1967-2009 . . . 120
3.17 VB-GSA of NOE model: empirical distributions of return intervals T . . . 120

3.18 VB-GSA of NOE model: distribution of the elevation of ground floor of buildings 123
3.19 VB-GSA of NOE model: depth-damage curve for private housing 123
3.20 VB-GSA of NOE model: empirical distribution of the NPV indicator 126
3.21 VB-GSA of NOE model: avoided damages against exceedance frequency 127
3.22 ZEC case study: avoided damages against exceedance frequency 132

4.1 Point-based and spatially additive model $\mathcal{F}$. 142
4.2 VB-GSA results depending on the size of disc-shaped support ν 148
4.3 VB-GSA results depending on correlation length a 149
4.4 VB-GSA results depending on covariance nugget parameter η 149
4.5 Grid of points $\mathbf{x}_i$ and spacing parameter s 154
4.6 Change of support in NOE model: three spatial supports 158
4.7 Change of support in NOE model: nominal maps of the ΔAAD indicator 159
4.8 Change of support in NOE model: sensitivity indices on three spatial supports 161
4.9 Change of support in NOE model: uncertainty on ΔAAD raster data 163
4.10 Change of support in NOE model: average values of the maps of sensitivity indices 164
4.11 Change of support in NOE model: maps of sensitivity indices 165
4.12 Change of support in NOE model: ratio π 166

$\mathcal{A}$.1 Creating sample $M_{1,2}^{(j)}$ 202
$\mathcal{A}$.2 Estimation des indices de sensibilité par approche brute ou par modules 205
$\mathcal{A}$.3 Overlay analysis for flood exposure assessment 213
$\mathcal{A}$.4 Plaine d'inondation de la Vilaine à Redon 217
$\mathcal{A}$.5 Le Rhône entre Fourques et Beaucaire 217
$\mathcal{A}$.6 Zone d'étude du projet ZEC 218
$\mathcal{A}$.7 Test case $\mathcal{F}_1$: distribution of model output and sensitivity indices 221
$\mathcal{A}$.8 Test case $\mathcal{F}_2$: distribution of model output and sensitivity indices 222
$\mathcal{A}$.9 Test case $\mathcal{F}_3$: distribution of model output and sensitivity indices 223
$\mathcal{A}$.10 Test case $\mathcal{F}_1$: scatterplots of model inputs against model output 224
$\mathcal{A}$.11 Test case $\mathcal{F}_2$: scatterplots of model inputs against model output 224
$\mathcal{A}$.12 Test case $\mathcal{F}_3$: scatterplots of model inputs against model output 224

$\mathcal{A}$.13 Test case $\mathcal{F}_1$: plots of contribution to the sample mean 225

$\mathcal{A}$.14 Test case $\mathcal{F}_2$: plots of contribution to the sample mean 225

$\mathcal{A}$.15 Test case $\mathcal{F}_3$: plots of contribution to the sample mean 225

$\mathcal{A}$.16 Spacing s of the grid of points $\mathbf{x}_i$ 242

$\mathcal{A}$.17 Two cells c_i and c_j centered on points $\mathbf{x}_i$ and $\mathbf{x}_j$, with two points $\mathbf{x} \in c_i$ and $\mathbf{y} \in c_j$ 245

List of Tables

1.1 Main methods of sensitivity analysis. 28

2.1 Orb Delta case study: flood scenarios 61
2.2 Orb Delta case study: content of the assets map 62
2.3 Orb Delta case study: damage functions 65
2.4 Orb Delta case study: model inputs of the NOE code 65
2.5 Orb Delta case study: nominal values of the output indicators 67
2.6 Orb Delta case study: spatial scales in the NOE code 69
2.7 Main characteristics of the extra case studies 71

3.1 Sensitivity indices estimates for each test case and each method 91
3.2 Methods for VB-GSA with spatially distributed inputs 97
3.3 Analytical test case: exact values of sensitivity indices 105
3.4 Exactness and precision of sensitivity indices for different numbers n of generated maps . 107
3.5 Exactness of the estimates of E_Λ and σ^2_Λ using LHS or SRS for increasing n 109
3.6 VB-GSA of NOE model: sources of uncertainty 118
3.7 VB-GSA of NOE model: confusion matrix of LU-GDB dataset 121
3.8 VB-GSA of NOE model: descriptive statistics for each output of interest 126
3.9 VB-GSA of NOE model: total-order sensitivity indices 128

4.1 Sensitivity analysis results with respect to the outputs of interest $Y(\mathbf{x}^*)$ and $|v| \cdot Y_v$ 144
4.2 Descriptive statistics of total ΔAAD over various supports 161

Acronyms

AAD/AAD'	Average Annual Damages in the present (resp. future) situation
ΔAAD	Average Annual Avoided Damages
ACB	Analyse Coût-Bénéfice (see CBA)
IPCC	Intergovernmental Panel on Climate Change
CBA	Cost-Benefit Analysis
CBA-AD	Cost-Benefit Analysis based on Avoided Damages
CEPRI	Centre Européen de Prévention des Risques d'Inondation
DOE	Design Of Experiments
DTM	Digital Terrain Model
GIS	Geographic Information Systems
GRF	Gaussian Random Field
GSA	Global Sensitivity Analysis
HDMR	High-Dimensional Model Representation
LHS	Latin Hypercube Sampling
NOE	Modelling chain for cost-benefit analysis based on avoided damages
NPV	Net Present Value
OAT	One-At-a-Time sensitivity analysis
PAPI	Programme d'Actions de Prévention des Inondations
SA	Sensitivity Analysis
SIG	Système d'Information Géographique (voir GIS)
SRF	Stationary Random Field
SRS	Simple Random Sampling
VB-GSA	Variance-Based Global Sensitivity Analysis

Notations

Notations related to the NOE modelling framework (§2.2 on page 40)

m (resp. m')	number of flood scenarios in the present (resp. future) situation
e_j	flood scenario
$q(e_j)$ or q_j	peak discharge associated with flood scenario e_j
$T(e_j)$ or T_j	return interval associated with flood scenario e_j
$f(e_j)$ or f_j or $F_Q(q_j)$	annual exceedance probability associated with flood scenario e_j
$\omega(e_j)$	weight of flood scenario e_j in the computation of the AAD indicator
$\mathcal{H}(e_j)$	set of hazard maps associated with flood scenario e_j (maps of water depth, water velocity, flood duration)
$D(e_j)$	damages associated with flood scenario e_j
$\mathcal{D}(e_j)$	map of spatially distributed damages associated with flood scenario e_j
B_i	benefits associated with a flood risk management plan at time step i
C_i	costs associated with a flood risk management plan at time step i
CI	investment costs associated with a flood risk management plan
CE	annual maintenance costs associated with a flood risk management plan
R	number of years over which a flood risk management plan is assessed
τ_i	discount rate at time step i

Notations related to numerical models (§1.1 on page 18)

$\mathcal{F}$	numerical model
$\mathcal{F}_{\text{loc}}$	local numerical model (§1.1.5 on page 23)
f	mathematical function implemented by a numerical code
U_i	scalar model input

$\mathbf{U}$	vector of scalar model inputs U_i
Z	spatially distributed model input
Y	model output
K	number of model inputs
k	number of scalar model inputs
α or β	subset of $[\![1; K]\!]$

Notations related to random variables

p_X	probability distribution function (pdf) of random variable $\mathbf{X}$
$\mathbb{E}(\mathbf{X})$ or μ_X	expected value of random variable $\mathbf{X}$
$\mathrm{var}(\mathbf{X})$ or σ_X^2	variance of random variable $\mathbf{X}$
$\mathbb{E}_Z Y$	conditional expectation of random variable Y given Z
$\mathcal{U}[a, b]$	uniform pdf on the interval $[a; b]$
$\mathcal{N}(\mu, \sigma)$	normal pdf of expected value μ and variance σ^2

Notations related to spatially distributed variables (§1.1.1 on page 18)

$\Omega \in \mathbb{R}^2$	spatial domain
$\mathbf{x} \in \Omega$	a point of spatial domain (synomyms: site, location)
$\mathbf{h} = \mathbf{x} - \mathbf{x}'$	vector between two points (synonym: lag vector)
$v \subset \Omega$ or $\mathcal{V} \subset \Omega$	subset of spatial domain (synonyms: block, support, zone)
$\lvert v \rvert$ or $\lvert \mathcal{V} \rvert$	surface area of a block
$\mathcal{G}$	regular grid (discrete set of points $\mathbf{x}_i$)
c_i	cell of a regular grid $\mathcal{G}$, centered on point $\mathbf{x}_i$
$\lvert c \rvert$	individual surface area of each cell in a regular grid $\mathcal{G}$
G	number of cells in a regular grid $\mathcal{G}$

Notations related to random fields

$\mathcal{P}$	stochastic process used to generate random realisations of a spatially distributed input Z
$\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$	random field

$Z(\mathbf{x})$	value of a random field on the site $\mathbf{x} \in \Omega$
Z_v	average value of a random field over a block v
V	variance of a random field
V_v	block variance of a random field over block v
$C(\cdot)$	covariance function of a random field
$\bar{C}(v, v)$	average value of a covariance function $C(\cdot)$ over block v
$\rho(\cdot)$	correlogram associated with a covariance function $C(\cdot)$
a	range associated with a covariance function $C(\cdot)$
A	integral range associated with a covariance function (see §4.1.2.1 on page 141)
η	nugget associated with a covariance function $C(\cdot)$

Notations related to sensitivity analysis (§1.2.2 on page 29)

S_X	first-order sensitivity index of model input X
S_i	first-order sensitivity index of the i^{th} model input
ST_X	total-order sensitivity index of model input X
ST_i	total-order sensitivity index of the i^{th} model input
$S_\alpha^{\text{gr.}}$	first-order sensitivity index of the group of model inputs $(U_i)_{i \in \alpha}$ with α a subset of $[\![1; K]\!]$
$ST_\alpha^{\text{gr.}}$	total-order sensitivity index of the group of model inputs $(U_i)_{i \in \alpha}$ with α a subset of $[\![1; K]\!]$
$S_X(\mathbf{x})$	site sensitivity index of model input X at site $\mathbf{x}$
$S_X(v)$	block sensitivity index of model input X over block v
M_1 and M_2	base samples in the space of model inputs used to estimate sensitivity indices (§1.2.2.5 on page 32)
M_{tot}	total sample in the space of model inputs used to estimate sensitivity indices (§1.2.2.5 on page 32)
N	number of lines of base samples M_1 and M_2 (§1.2.2.5 on page 32)
N_{tot}	number of lines of total sample M_{tot} (§1.2.2.5 on page 32)

Notations related to sensitivity analysis with spatially dsitributed inputs (§3.1 on page 79)

ξ	"*trigger*" input (§3.1.2.3 on page 83)

$\mathbf{L}$	"*random label*" used in the "*map labelling*" approach (§3.1.2.4 on page 84)
λ	real-valued random variable used in the "*macro-parameter*" and "*dimension reduction*" methods (§3.1.2 on page 79)
θ	real-valued random variable used in the "*second level*" approach (§3.1.2.6 on page 86)
n	number of random realisations of a spatially distributed input Z in the "*map labelling*" approach (§3.1.2.4 on page 84)

Miscellaneous

$\tilde{X}$	proxy for any quantity of interest X

Introduction

Modèles numériques spatialisés

Les termes « *modèle* » et « *modélisation* » peuvent recouvrir des pratiques variées qui selon Armatte and Dalmedico (2004) visent à *« représenter un système concret ou réel par un objet formel qui permette de penser ce réel mais aussi d'agir sur lui »*. Dans ce mémoire, nous utiliserons le terme « *modèle* » de manière très restrictive pour désigner un code numérique, considéré comme une boîte noire, qui calcule des variables de sortie à partir de plusieurs variables d'entrée. Nous nous limiterons plus précisément aux modèles qui s'appuient sur une description mécaniste des processus étudiés, par opposition aux modèles empiriques ou statistiques, et parmi eux aux modèles déterministes uniquement (non stochastiques). Avec l'essor de l'informatique et l'explosion des capacités de calcul, la modélisation numérique s'est peu à peu imposée comme une activité incontournable dans les sciences de la nature et les sciences de l'ingénieur. Pour répondre aux grands défis environnementaux de notre époque, une part significative de l'effort de recherche est ainsi dédiée à la construction de modèles capables de décrire (modèles diagnostiques) des systèmes physiques, biologiques, environnementaux, économiques ou sociaux complexes, simuler ou prédire leur comportement (modèles pronostiques), proposer des stratégies d'action aux décideurs publics (modèles d'aide à la décision), voire même défendre des positions lors de négociations internationales (sur le climat ou le devenir de l'agriculture mondiale, par exemple). Ainsi que l'expliquent Bouleau et al. (2004), la modélisation est une activité qui est aujourd'hui devenue le vecteur principal des passages entre Science et Société.

Parmi les modèles numériques utilisés pour l'exploration des problématiques environnementales, nombreux sont ceux qui s'appuient sur des données distribuées spatialement, telles que des Modèles Numériques de Terrain, des cartes de sol, des cartes d'occupation du sol, etc. (Ostendorf 2011). Ces modèles que nous qualifierons de « spatialisés » tirent parti du développement rapide des outils et méthodes permettant d'acquérir, de structurer, d'exploiter et de diffuser l'information géographique. Le scientifique comme le citoyen ou le décideur ont aujourd'hui à leur disposition un ensemble de données environnementales spatialisées toujours plus grand, ainsi que des outils toujours plus nombreux et plus performants pour utiliser ces données : Systèmes d'Information Géographique (SIG), imagerie satellitaire, cartographie interactive sur le Web, technologies de géolocalisation embarquées, etc. Les modélisateurs se sont appropriés ces nouvelles données et construisent aujourd'hui des modèles spatialisés qui permettent une description explicite des structures, des dépendances et des dynamiques spatiales intervenant dans les processus physiques, biologiques ou anthropiques qu'ils étudient.

Analyses d'incertitude et de sensibilité

Cependant, à mesure que la modélisation numérique prend une place prépondérante dans de nombreux champs de la Science, un discours critique se construit et souligne les défauts et limites de cette démarche (Oreskes et al. 1994). Parmi les points soulevés, ces critiques soulignent l'importance des incertitudes qui interviennent dans tout processus de modélisation. Ces incertitudes peuvent être liées à un manque de connaissance sur certains des phénomènes étudiés, à la variabilité naturelle des grandeurs représentées, à des erreurs de mesure, à des choix de modélisation simplificateurs ou encore à des approximations numériques (Walker et al. 2003). Elles se combinent et se propagent à travers le modèle et entraînent une incertitude sur les résultats et indicateurs fournis par celui-ci. Lorsqu'un modèle est utilisé comme appui à des décisions de nature opérationnelle ou stratégique, il faut alors s'interroger, comme le soulignent Vasseure et al. (2005), sur « *la valeur d'une décision basée sur des données dont la qualité est mal connue ou mal comprise par le décideur* ». Les modèles spatialisés n'échappent pas à ce questionnement et « *quiconque utilise une information incertaine (c'est-à-dire l'écrasante majorité des utilisateurs de données cartographiées) doit réfléchir avec attention aux sources possibles de l'incertitude et à la manière de s'en occuper* » (Fisher et al. 2005).

Afin de répondre au moins partiellement à ces difficultés, la communauté scientifique a développé des méthodes qualitatives et quantitatives qui permettent d'étudier comment réagissent les sorties d'un modèle à des perturbations sur ses variables d'entrée : ce sont des méthodes regroupées sous les termes d'« *analyse de sensibilité* » et d'« *analyse d'incertitude* ». L'analyse d'incertitude se concentre sur la propagation des incertitudes à travers le modèle, et vise à quantifier l'incertitude résultante qui existe sur la sortie. Elle permet typiquement d'associer un intervalle de confiance aux résultats fournis par un modèle. L'analyse de sensibilité va plus loin : elle cherche à mesurer l'influence de l'incertitude de chacune des variables d'entrée sur la précision du résultat du modèle. Elle permet de hiérarchiser les variables d'entrée en fonction de leur contribution à la variabilité de la sortie du modèle. Elle vise ainsi à identifier les variables d'entrée critiques, celles qui conditionnent la décision finale de l'utilisateur du modèle, et sur lesquelles il faut orienter les efforts de recherche futurs. Les méthodes d'analyse d'incertitude et d'analyse de sensibilité ont peu à peu été adoptées par les modélisateurs dans différents champs disciplinaires, notamment dans la recherche environnementale (Cariboni et al. 2007; Tarantola et al. 2002), et sont aujourd'hui reconnues comme des étapes essentielles dans la construction d'un modèle numérique (European Commission 2009a; CREM 2009).

Pourtant, ces méthodes d'analyse de sensibilité sont peu souvent appliquées à l'étude de modèles numériques spatialisés. Parmi les raisons qui freinent leur utilisation dans ce domaine, on peut citer l'explosion des problèmes de dimensionnalité, ainsi que le manque de maturité de certains modèles. Nous nous intéresserons plus particulièrement aux deux autres limites que voici : i) les données d'entrée des modèles spatialisés présentent généralement une auto-corrélation spatiale, alors que les méthodes classiques d'analyse de sensibilité ne considèrent que des variables scalaires indépendantes ; ii) les notions d'échelle, de support ou de résolution, qui jouent un rôle prépondérant dans les modèles spatialisés, sont ignorées dans les cadres formels des méthodes d'analyse de sensibilité classiques. Le besoin d'adaptation des méthodes d'analyse de sensibilité au contexte spécifique des modèles spatialisés apparaît donc important. Naturellement, des éléments de réponse ont déjà été apportés à ce problème. Ainsi on trouve dans la littérature dédiée à l'analyse de sensibilité des travaux portant sur les variables d'entrées corrélées, mais ces études ne s'intéressent que rarement au cas particulier de la dépendance spatiale. Par ailleurs les statistiques spatiales, et plus particulièrement la géostatistique, proposent des cadres théoriques pour décrire l'incertitude pesant sur des variables spatialisées et pour appréhender les notions d'échelle, de support ou de résolution. Elles fournissent aussi des outils pour simuler ces incertitudes. Cependant, ce corpus théorique n'a jamais été rapproché de celui de l'analyse de sensibilité de modèles numériques.

Un exemple : l'analyse coût-bénéfice des projets de prévention du risque d'inondation

Nous nous sommes intéressés dans cette thèse à un exemple particulier de situation où il est fait appel à des modèles numériques spatialisés : ce contexte est celui de l'évaluation économique des projets de prévention du risque d'inondation.[a] Plus précisément, notre travail a porté sur les démarches d'analyse coût-bénéfice (ACB) qui visent à caractériser, au moyen d'un indicateur synthétique ou spatialisé, la pertinence économique de projets d'aménagement de protection contre les crues (barrages, digues, etc.). La mise en œuvre ce type d'analyse passe généralement par le développement d'un modèle numérique, qui fait intervenir divers modules à même de décrire l'ensemble de la chaîne menant au risque d'inondation (modules hydrologiques, hydrauliques, d'occupation des sols, fonctions d'endommagement). Ces modèles numériques ont généralement une forte composante spatiale, de part la nature de leurs données d'entrée (topographie du terrain, réseau hydrographique, carte d'occupation du sol, etc.), la nature de leurs sorties (indicateurs de risque spatialisés), et les traitements qu'ils mobilisent (notamment des opérations d'analyse spatiale à l'aide de logiciels SIG).

La nécessité de mieux prendre en compte les incertitudes dans les modèles d'analyse économique du risque d'inondation, et plus particulièrement dans les études ACB appliquées aux projets de prévention des crues, fait consensus au sein de la communauté scientifique qui développe ces modèles (Apel et al. 2004). Les motivations pour une étude approfondie de ces incertitudes sont multiples. Dans une première phase de développement d'un modèle, une telle étude permet au modélisateur de mieux comprendre le comportement de l'outil qu'il construit, de faire émerger de nouvelles questions ou des pistes d'amélioration. Dans une phase d'utilisation du modèle, l'analyse d'incertitude et de sensibilité peut permettre d'améliorer la robustesse des indicateurs économiques qui sont produits, notamment en identifiant les variables d'entrée clés dont il faut au mieux préciser la valeur. Enfin, dans une phase opérationnelle, la prise en compte explicite des incertitudes permet de fournir aux utilisateurs finaux du modèle (pouvoirs publics, gestionnaires de territoire) une information plus complète pour les aider dans leur prise de décision (Ascough et al. 2008).

On compte ainsi de nombreux travaux récents qui visent à appliquer des analyses d'incertitude et/ou de sensibilité à tout ou partie de modèles numériques d'évaluation économique des crues. La plupart de ces travaux se limitent à une analyse d'incertitude, dont le périmètre peut varier d'un unique « *module* » du modèle complet—par exemple, l'occupation du sol (Te Linde et al. 2011), la simulation hydraulique (Bales and Wagner 2009), l'estimation des dommages (Merz et al. 2004)— jusqu'à la chaîne de modélisation dans son intégralité (Apel et al. 2008). Quelques publications plus rares abordent aussi le problème de la hiérarchisation des différentes sources d'incertitude dans ces modèles, en les soumettant à une analyse de sensibilité (Koivumäki et al. 2010; de Moel and Aerts 2011; de Moel et al. 2012; Pappenberger et al. 2008).

Cependant, trois remarques générales peuvent être formulées à l'encontre de ces études. Tout d'abord, la majorité d'entre elles s'intéressent à la précision de l'estimation des dommages dus aux crues sur un territoire, mais peu examinent l'incertitude qui en résulte sur les indicateurs de performance économique des projets de prévention des crues produits dans le cadre d'une ACB. De plus, ces études ne s'appuient que rarement sur les dernières avancées faites dans le domaine de l'analyse de sensibilité des modèles

[a]Pour comprendre ce choix, il est nécessaire d'expliquer que ce travail de recherche a gravité autour de trois unités de recherche : l'UMR TETIS, dont les compétences portent sur la maîtrise de la chaîne de l'information géographique ; l'UMR G-EAU, où sont notamment développés des modèles spatialisés sur des problématiques liées à la gestion de l'eau à l'échelle des territoires ; l'Institut de Mathématiques et de Modélisation de Montpellier (I3M), au sein duquel plusieurs travaux ont déjà été menés sur les méthodes d'analyse de sensibilité.

numériques, et ont le plus souvent recours à des approches naïves dont les limites sont connues—ainsi, à notre connaissance, seuls de Moel et al. (2012) utilisent dans leurs travaux l'analyse de sensibilité globale basée sur la décomposition de la variance du modèle, qui est pourtant largement utilisée dans d'autres champs thématiques. Enfin, elles font généralement peu de cas du caractère spatial des modèles étudiés, et des problèmes spécifiques que ce caractère pose pour mener une analyse de sensibilité. Deux points sont notamment passés sous silence, à savoir les problèmes liés à la dépendance spatiale dans les données d'entrée, et les questions relatives à l'échelle spatiale à laquelle ces modèles sont construits. Ainsi, il apparaît possible et souhaitable de progresser encore dans la prise en compte des incertitudes dans ces modèles dédiés à l'analyse économique du risque d'inondation, et plus particulièrement dans les études ACB appliquées aux projets de prévention des crues.

Objectifs de la thèse

Le but de ce travail de recherche est double. Un premier objectif est de nature méthodologique : il s'agit de proposer des méthodes d'analyse de sensibilité adaptées à l'étude de modèles numériques dont les entrées et/ou les sorties sont spatialisées. Ces méthodes doivent permettre d'appréhender les spécificités de ces modèles spatialisés, notamment la présence de dépendance spatiale dans les données et les questions liées aux notions d'échelle. Le second objectif est de nature appliquée : il s'agit, après avoir proposé un modèle pour décrire les études ACB appliquées aux projets de prévention du risque d'inondation, d'étudier la propagation des incertitudes dans ce modèle, et plus précisément d'identifier les sources d'incertitude principales, à l'aide de méthodes d'analyse de sensibilité adaptées.

Méthode de recherche Pour tenter d'atteindre simultanément ces deux objectifs méthodologique et appliqué, nous avons opté pour une approche essentiellement inductive. Nous nous sommes appuyés sur l'examen approfondi d'une situation particulière : l'analyse de sensibilité d'une ACB d'un projet de prévention des inondations dans la basse vallée de l'Orb (Hérault, France). Une première étape de notre travail a ainsi consisté à proposer un cadre formel, baptisé NOE, pour modéliser les ACB appliquées aux projets de prévention des crues, puis à implémenter un code numérique pour mettre en œuvre une analyse coût-bénéfice sur le terrain d'étude de l'Orb. À partir de l'étude de ce modèle singulier, nous avons tenté de faire émerger des questionnements, des méthodes et des énoncés plus généraux sur l'analyse de sensibilité de modèles spatialisés. Ainsi, plus qu'un simple cas d'étude, le modèle NOE et son application à la basse vallée de l'Orb ont servi de base à notre réflexion : ils occupent donc une large place dans ce mémoire de thèse.

Limites Nous avons fait le choix de restreindre notre recherche à une unique famille de méthodes d'analyse de sensibilité, celles basées sur la décomposition de la variance de la sortie du modèle (« *variance based global sensitivity analysis* » ou VB-GSA). Ces méthodes produisent des indices de sensibilité qui mesurent la contribution de chaque entrée du modèle à la variabilité de la sortie, en tenant compte à la fois de la structure (ou des équations) du modèle étudié, mais aussi de la plage de variabilité des entrées[b]. Trois raisons principales motivent notre choix : i) ces méthodes ne nécessitent aucune hypothèse préalable sur la nature du modèle étudié (linéarité, régularité), qui est considéré comme une simple « *boîte noire* » ; ii) elles explorent largement l'espace des incertitudes sur les variables d'entrée du modèle (caractère global) ; et iii) elles décrivent non seulement l'impact des variables d'entrée incertaines prises

[b] Elles diffèrent en ce sens des notions de sensibilité usuellement utilisées en physique, qui s'appuient uniquement sur le calcul de dérivées partielles locales.

une à une, mais aussi l'impact des interactions entre ces variables d'entrée. Ensuite, nous nous sommes intéressés uniquement au cas où le modèle étudié présente des temps de calcul faibles, qui permettent notamment de recourir à un très grand nombre de simulations dans des approches de type Monte Carlo. Les méthodes d'analyse de sensibilité s'appuyant sur la construction d'émulateurs ou de méta-modèles ne seront donc pas abordées dans ce mémoire. Enfin, les caractéristiques du modèle NOE nous ont amené à examiner plus particulièrement les modèles spatialisés dit « *ponctuels* » et « *spatialement additifs* » (ces termes seront définis plus loin dans ce document).

Questions de recherche Dans ce cadre restreint, des questions plus précises émergent : comment peut-on définir et estimer des indices de sensibilité basés sur la variance pour des variables d'entrées présentant une auto-corrélation spatiale ? Quelles sont les stratégies d'échantillonnage et de simulation de ces variables spatialisées les plus appropriées pour estimer leurs indices de sensibilité ? L'analyse de sensibilité basée sur la variance permet-elle de rendre compte de manière pertinente des questions d'échelle spatiale pour des modèles ponctuels et spatialement additifs ? Voilà les questions auxquelles nous tenterons de répondre dans ce document.

Structure du document

Ce mémoire est composé de quatre chapitres. Nous avons pris le parti de mêler dans ces chapitres des développements d'ordre théorique ou méthodologique et des résultats numériques obtenus sur le modèle NOE. Deux raisons motivent ce choix : d'une part, cette articulation reflète le déroulement réel de notre travail de recherche, où les développements méthodologiques et appliqués se sont nourris mutuellement ; d'autre part, cette présentation croisée permet d'enrichir la discussion en fin de chaque chapitre sur les questions de recherche abordées. La Figure 1 en page 7 résume la structure du document.

Dans un premier chapitre, nous posons un certain nombre de définitions et de notations pour mieux cerner notre objet de recherche méthodologique que sont les modèles numériques spatialisés. Nous y présentons également une brève revue sur les méthodes d'analyse de sensibilité en général et l'analyse de sensibilité basée sur la variance en particulier. Une fois ces bases posées, nous formulons de manière plus détaillée nos questions de recherche dans la conclusion de ce chapitre.

Dans un second chapitre, nous présentons le travail de modélisation qui a conduit à la constitution de notre cas d'étude. Nous proposons d'abord un cadre de modélisation baptisé NOE pour décrire les études ACB des projets de prévention des crues basée sur l'approche des dommages évités. Nous précisons ensuite comment ce cadre de modélisation a été décliné sur le site d'étude particulier de la basse vallée de l'Orb.

L'objet du troisième chapitre est de proposer des méthodes pour intégrer des variables d'entrée spatialisées dans l'analyse de sensibilité basée sur la variance. Une revue des approches existantes y est complétée par une étude numérique de ces approches, par des développements sur les problèmes d'échantillonnage des variables spatialisées dans ces approches et par l'application d'une des méthodes présentées au modèle NOE sur le terrain d'étude de l'Orb.

Le quatrième chapitre est quant à lui dédié à l'étude des liens entre les notions d'échelle spatiale (telles que définies par Blöschl and Sivapalan (1995)) et l'analyse de sensibilité basée sur la variance. On s'intéresse plus spécifiquement à l'influence du support de la sortie d'un modèle spatialisé sur la hiérarchisation des sources d'incertitude. Les développements théoriques proposés se limitent au cas des modèles ponctuels et spatialement additifs ; ils sont illustrés par une application au modèle NOE sur le terrain d'étude de l'Orb.

Enfin, nous concluons ce document par i) des remarques générales sur la portée et les perspectives d'utilisation de nos contributions méthodologiques pour l'analyse de sensibilité de modèles numériques spatialisés ; ii) des enseignements plus appliqués sur les incertitudes dans les analyses coût-bénéfice d'aménagements de protection contre les inondations ; et iii) quelques retours d'expérience sur la mise en œuvre pratique d'une analyse de sensibilité d'un modèle environnemental d'aide à la décision.

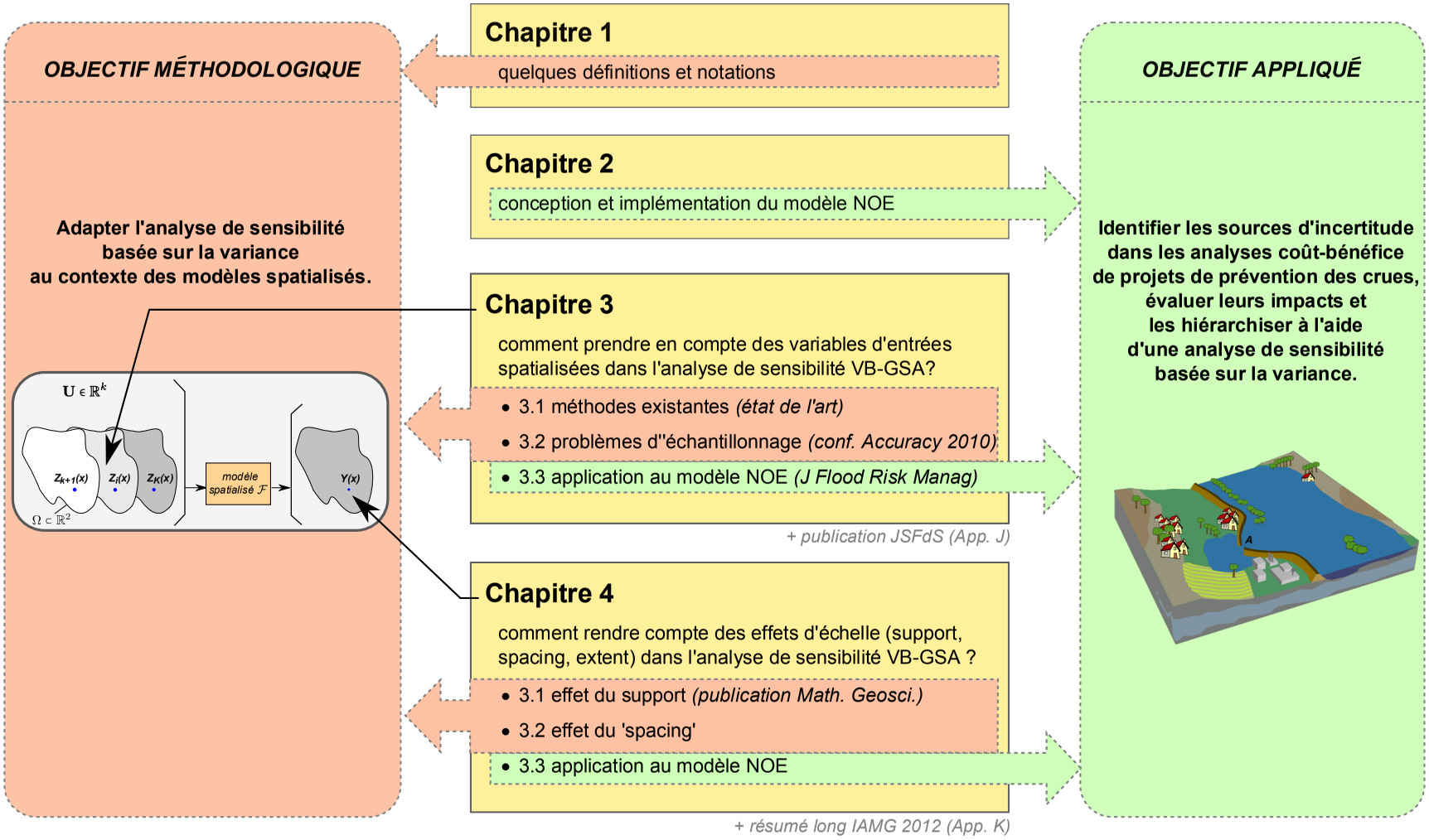

FIGURE 1 – Structure générale de la thèse : objectif méthodologique (gauche), objectif appliqué (droite), plan de la thèse (centre)

Introduction

Spatially distributed models

According to Armatte and Dalmedico (2004), the terms "*model*" and "*modelling*" can cover diverse practices that aim "*to represent a real system by a formal object allowing one to both consider and act on this system*". In this thesis, we will use the term "*model*" in a very restrictive way to designate a numerical code, which is considered as a "*black box*", that calculates output variables based on a set of input variables. More precisely, we limit ourselves to models that are based on a mechanistic description of the processes under study, as opposed to those which are empirical or data-based, and of these models, we focus solely on deterministic ones (non-stochastic). With the dramatic rise of computer performances over the last decades, numerical modelling has expanded steadily and has now established itself as a key activity in earth and environmental sciences. Numerical models are widely recognized as valuable tools to describe complex physical, biological, ecological, economic or social systems, to understand their drivers (diagnostic models), to simulate and predict their future behaviour (prognostic models), to make informed management decisions (decision-support models), and even to defend positions during international negotiations (on the climate or the future of world agriculture, for example). As Bouleau et al. (2004) clearly state it, modelling has now won recognition as the main bridge between Science and Society.

Among the numerical models used to explore environmental issues, many rely on spatially distributed data, such as Digital Terrain Models, soil maps, land use maps, etc. (Ostendorf 2011). These models, which we will refer to as "*spatial models*", have benefited from the recent development of tools and methods allowing the acquisition, structuring, exploitation and diffusion of geographic information. Scientists today, like citizen and policy makers, have at their disposal a continually expanding set of spatial environmental data, as well as an ever rising number of increasingly efficient tools to use these data: Geographic Information Systems (GIS), satellite images, webmapping, on-board geolocation technologies, etc. Modellers have quickly learnt how to master these new data and tools, and now build spatial models that allow an explicit description of the spatial structures, inter-dependencies, and dynamics involved in the physical, biological, and anthropogenic processes under study.

Uncertainty and sensitivity analysis

However, as numerical modelling assumes a leading role in numerous scientific fields, criticism highlighting the weaknesses and limits of this approach has arisen (Oreskes et al. 1994). Among the points raised is the importance of the uncertainties involved in any modelling process. These uncertainties may stem from a lack of knowledge about some of the phenomena studied, the natural variability of the quantities

of interest, measurement errors, model assumptions, or numerical approximations (Walker et al. 2003). These combine together and propagate throughout the model, leading to uncertainty regarding the outputs and indicators produced by it. When a model is used as a support tool for operational or strategic decision-making, one must then question, as emphasized by Vasseure et al. (2005), "*the value of a decision based on data, the quality of which is unknown or poorly understood by the decision-maker*". Spatial models are not exempt from these questions and "*anyone using uncertain information—meaning the overwhelming majority of mapped data users—should consider carefully the possible sources of uncertainty and how to deal with them*" (Fisher et al. 2005).

To adress this issue, the scientific community has developed qualitative and quantitative methods that allow the study of how model outputs react when input variables are uncertain; they are usually refered to as "*uncertainty and sensitivity analysis*" methods. Uncertainty analysis focuses on the propagation of uncertainties throughout the model and aims to quantify the resulting uncertainty on the output. It typically allows a confidence interval to be associated with model outputs. Sensitivity analysis goes one step further: it seeks to study how the uncertainty in a model output can be apportioned to the uncertainties in each of the model inputs. It allows input variables to be ranked according to their contribution to the variability of model outputs. Sensitivity analysis thus helps to identify the key input variables, those that determine the final decision of the model end-user, and on which further research should be carried out. Uncertainty analysis and sensitivity analysis methods have been gradually adopted by modellers in different disciplinary fields, notably in environmental research (Cariboni et al. 2007; Tarantola et al. 2002), and today are widely recognized as essential steps in model building (European Commission 2009a; CREM 2009).

However, these sensitivity analysis methods have not been applied frequently to the study of spatial models. Among the factors hindering their use in this domain are the explosion of dimensionality problems and the lack of maturity of certain models. We will examine in special detail two other limits: i) spatial model input data generally exhibit some auto-correlation, yet conventional sensitivity analysis methods only consider independent scalar variables; ii) notions of scale, support, and resolution, which play a prominent role in spatial modelling, are ignored in the formal frameworks of conventional sensitivity analysis methods. Hence, there appears to be a great need to adapt sensitivity analysis methods to the specific context of spatially distributed modelling. Some ideas have already been provided in the literature to adress this issue. First, in existing research related to sensitivity analysis, one may find some publications that deal with correlated input variables, but these studies rarely examine the particular case of spatial dependence. In addition, spatial statistics, and more specifically geostatistics, offer theoretical frameworks to describe the uncertainty weighing on spatially distributed data and to grasp the notions of spatial scale, support, or resolution. Geostatistics also provide tools to simulate these spatial uncertainties. However, this theoretical corpus has never been linked to that of sensitivity analysis of numerical models.

An example: cost-benefit analysis of flood risk management plans

In this thesis, we look into a particular situation where spatial models are used: the economic assessment of flood risk management plans.[c] More precisely, our work focuses on cost-benefit analysis (CBA) approaches that aim to characterize, using a scalar or spatially distributed indicator, the economic relevance of flood mitigation plans (dams, dikes, etc.). Such CBA studies usually require the development of a numerical model, which combines a number of modules (hydrological module, hydraulic module, land use module, damage functions) able to describe the entire chain leading to flood risk. These numerical models generally have a strong spatial component owing to the nature of their input data (topography,

map of water system, land use map, etc.), the nature of their outputs (spatial risk indicators), and the calculations that they use (notably spatial analysis operations with the help of GIS software).

Meanwhile, there is a growing consensus (Apel et al. 2004) that flood economic assessment models, and more particularly CBA studies applied to flood risk management plans, are fraught with uncertainties. There are numerous motivations for an in-depth study of these uncertainties. In the first stage of model development, such a study allows the modeller to better understand the behaviour of the tool that s/he is building, and to bring to light new questions or paths for improvement. During the model use stage, uncertainty and sensitivity analyses can lead to increase the robustness of the economic indicators produced, notably by identifying the key input variables whose values should be better specified. Lastly, during the operational stage, explicitly taking into account uncertainties allows the provision of more complete information to the model end-users (public authorities, water managers) to help them in their decision making (Ascough et al. 2008).

A number of recent studies aim to apply uncertainty and/or sensitivity analyses to all or parts of flood economic assessment models. Most of these studies are limited to the forward propagation of uncertainty (uncertainty analysis), the perimeter of which can vary from a single module of the complete model—e.g., land use (Te Linde et al. 2011), hydraulic simulation (Bales and Wagner 2009), estimation of damages (Merz et al. 2004)—up to the entire modelling chain (Apel et al. 2008). Fewer publications address the issue of ranking the various sources of uncertainty in these models by performing sensitivity analysis (Koivumäki et al. 2010; de Moel and Aerts 2011; de Moel et al. 2012; Pappenberger et al. 2008).

However, we can make three general comments regarding these studies. First, the majority of these studies are interested in the accuracy of the flood damage estimates on an area, but few examine the resulting uncertainty on the economic performance indicators of flood mitigation plans produced in a CBA. Next, these studies only rarely make use of the most recent advances in the field of sensitivity analysis, turning more frequently to naive approaches whose limits are well known—e.g., only de Moel et al. (2012) use a variance-based global sensitivity analysis in their research, although this approach is widely used in other fields. Finally, these studies generally disregard the spatial nature of the model they scrutinize, and the specific problems that this nature raises when performing sensitivity analysis. In particular, the following two points are ignored: i) the issue of spatial dependence in the input data, and ii) the issues related to the spatial scale at which these models are built. It thus appears both possible and desirable to further improve the treatment of uncertainties in numerical models dedicated to the economic assessment of flood risk, and more particularly in CBA studies applied to flood risk management plans.

Objectives of the thesis

This research work has two goals. The first objective is methodological: we want to investigate the use of sensitivity analysis methods in spatially distributed modelling. These methods should allow to grasp the specific features of spatial models, notably the presence of spatial dependence in the data and the scaling issues. The second objective is of an applied nature: after proposing a modelling framework to describe CBA studies applied to flood risk management plans, it involves studying the propagation of uncertainties in this model, and, more precisely, identifying the main sources of uncertainty with the help of suitable sensitivity analysis methods.

[c]The main reason for this choice lies in the fact that this research work revolved around three different research units: the joint research unit TETIS, that undertakes research on the use of geospatial data for the monitoring and modelling of environmental systems; the joint research unit G-EAU, which deals with the modelling of hydrosystems; and the Institute of Mathematics and Modelling in Montpellier (I3M).

Research method In an attempt to simultaneously achieve both our methodological and applied objectives, we chose an essentially inductive approach. We worked from the in-depth examination of a specific situation: the sensitivity analysis of a CBA of a flood risk management plan in the Orb Delta (Hérault, France). The first step of our work involved designing a modelling framework, named NOE, to describe CBA studies applied to flood risk management plans, followed by the implementation of a numerical code to carry out a CBA study on the Orb Delta. Based on the sensitivity analysis of this single model, we attempted to bring to light more general questions, methods, and statements regarding sensitivity analysis of spatial models. Rather than being a simple case study, the NOE model and its application to the Orb Delta thus served as a basis for our discussion: hence, they take up a large part of this thesis.

Limits We chose to focus our research on a single family of sensitivity analysis methods, namely variance-based global sensitivity analysis (VB-GSA). These methods produce sensitivity indices that measure the contribution of each model input to the variance of the model output, taking into account both the structure (or the equations) of the model under study and the uncertainty range of inputs. Three main arguments motivated our choice: i) these methods do not require any preliminary hypothesis regarding the nature of the model under study (linearity, regularity), which is considered as a simple "*black box*"; ii) they widely explore the space of input uncertainties (global methods); and iii) they describe not only the impact of uncertain input variables considered one at a time, but also the impact of interactions between these input variables. We also focused our attention on models with low CPU cost, which allow a large number of simulations to be run in Monte Carlo approaches. Therefore, we do not adress in this thesis the sensitivity analysis methods based on emulators or meta-models. Finally, the characteristics of the NOE model led us to look more specifically at "*point-based*" and "*spatially additive*" models—we will define these terms later in the document.

Research questions Within this restricted framework, more specific questions emerge: how can one define and estimate variance-based sensitivity indices for spatially distributed model inputs that exhibit spatial auto-correlation? What are the most appropriate simulation and sampling strategies for these spatially distributed inputs to estimate their sensitivity indices? Can variance-based global sensitivity analysis account for scaling issues, in particular for point-based and spatially additive models? These are the research items that we will try to adress in this thesis.

Outline of the thesis

This thesis is divided into four chapters. In these chapters, we intentionaly mixed methodological or theoretical developments with the numerical results obtained on the NOE case study. The reasons for this choice are twofolds: first, this articulation mirrors the actual unfolding of our research, in which methodological and applied work mutually enriched each other; next, this intertwined presentation allows us to discuss both sides of our work at the end of each chapter. Figure 2 on page 14 displays the outline of this document.

Chapter 1 starts with some relevant background information. We first give a number of definitions and notations on spatial models. We also display a brief introduction to the concepts of sensitivity analysis, and portray into more details the mathematical basics of variance-based global sensitivity analysis (VB-GSA). Once these foundations have been established, we articulate in greater detail our research questions in the conclusion of the chapter.

Chapter 2 presents the modelling process that led to the constitution of our case study: the cost-benefit analysis of a flood risk management plan on the Orb Delta. We first design a modelling framework named NOE to describe CBA studies of flood risk management plans based on the "*avoided damage*" approach. We then specify how the NOE modelling framework was implemented for the Orb Delta case study.

Chapter 3 concentrates on the first research question: *"how to handle spatially distributed inputs in VB-GSA?"*. We survey the existing approaches to compute variance-based sensitivity indices for spatially distributed inputs, then compare these approaches on some analytical test cases. This review is completed by some developments on the sampling of spatially distributed inputs, and by the application of VB-GSA to the NOE model on the Orb Delta case study.

Chapter 4 puts forward the second research question, by investigating the links between the spatial scale—as defined by Blöschl and Sivapalan (1995)—and variance-based global sensitivity analysis. We discuss how the ranking of the uncertain model inputs may depend on the "*support*" and the "*spacing*" of the spatially distributed model output. The theoretical developments are restricted to point-based and spatially additive models only; they are illustrated by an application to the NOE model on the Orb Delta case study.

Finally, this document is concluded by: i) general comments on the significance of our theoretical results and the possible use of our methodological contributions for sensitivity analysis of spatial models; ii) some lessons not only on the NOE modelling framework, but more generally on uncertainties in cost-benefit analyses applied to flood risk management plans; and iii) some feedback on the practice of sensitivity analysis in environmental modelling.

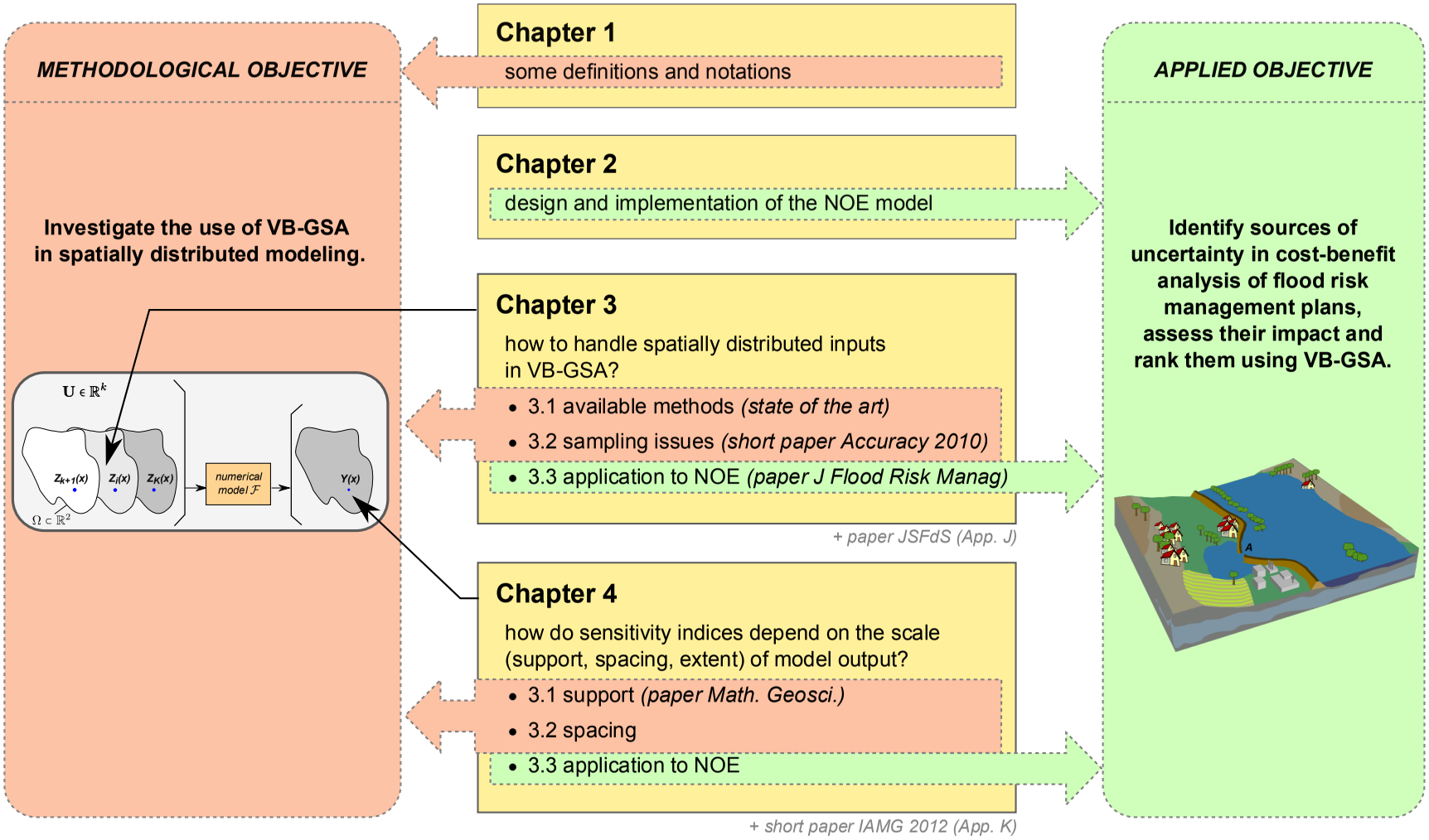

Figure 2: Outline of the thesis: methodological objective (left), applied objective (right), outline of the document (center)

Chapter 1

Theoretical background

THE goal of this first chapter is to define a number of notions that will be used throughout the thesis. In a first section §1.1, we clarify the term "*spatial models*", giving a number of notations (§1.1.1), discussing how space is represented in numerical models (§1.1.2) and presenting the various meanings of "*scale*" in this context (§1.1.3). We then focus on two specific families of spatial models that are often encountered in the literature: i) "*spatially additive models*", in which the model end user is interested in the spatial linear average or the sum of some quantity of interest over a given spatial unit (§1.1.4), and ii) "*point-based models*", in which spatial interactions in the physical processes under study can be neglected in a first approximation (§1.1.5).

In the second section §1.2, we give an overview of sensitivity analysis (SA) of numerical models. We first display the general flowchart of most SA methods (§1.2.1). We then focus on variance-based global sensitivity analysis (VB-GSA), introduce the mathematical basis of this approach and explain how to estimate variance-based sensitivity indices (§1.2.2).

1.1 Spatial models 18
1.1.1 Definitions and notations 18
1.1.2 Representing space in numerical models 19
1.1.3 Scale issues 20
1.1.4 Spatially additive models 22
1.1.5 Point-based models 23
1.2 Sensitivity analysis 24
1.2.1 An overview of sensitivity analysis methods 24
1.2.1.1 Measure of model output variability 25
1.2.1.2 Modelling uncertainty sources 26
1.2.1.3 Uncertainty propagation 26
1.2.1.4 Ranking model inputs 26
1.2.1.5 Local/global sensitivity analysis 27
1.2.1.6 Choice of an appropriate SA method 27
1.2.2 Variance-based global sensitivity analysis 29
1.2.2.1 Overview 29
1.2.2.2 A decomposition of functional variance 29
1.2.2.3 Definition and properties of variance-based sensitivity indices 30
1.2.2.4 Grouping model inputs 32
1.2.2.5 Estimation of variance-based sensitivity indices 32
1.3 Chapter conclusion 33

1.1 Spatial models

1.1.1 Definitions and notations

Our work is devoted to the study of numerical models where some of the model inputs and/or some of the model outputs are spatially distributed: we will use the term "*spatial model*" to refer to such models. They are encountered in many disciplines related to earth and environmental sciences. These models allow for a spatially explicit description of the physical or anthropogenic processes under study. Their development over the last decade is partly due to the increasing availability of spatially distributed data, to the boom of GIS tools and to the overall growth of available computing power. Here are a few examples of fields in which spatial models are often used: flood modelling (e.g., LISFLOOD, Van Der Knijff et al. (2010)), groundwater quality modelling (e.g., AquiferSim, Landcare Research (2011)), crop modelling (e.g., GLAM, Challinor et al. (2004)), air pollution modelling (e.g., AirGIS, Jensen (1998)), habitat modelling in ecology (e.g., HABITAT, Haasnoot and Van Dewolfshaar (2009)), global change modelling (Global Circulation Models), traffic modelling, etc.

In our work, we will only consider models that are: i) process-based models, i.e. models that describe the behaviour of a system in terms of lower-level mechanisms in a bottom-up approach, as opposed to empirical models, which are often based on statistical relationships extracted from observed data; ii) deterministic models, as opposed to stochastic models; and iii) models that are defined on a two-dimensional spatial domain[a]. We give here some notations that will be used throughout the document.

Numerical model For the sake of clarity, we will use in this document the term "*model*" in a very restrictive way, to denote a numerical code, considered as a black-box, which calculates a number of outputs (response variables) as a deterministic function of a given set of inputs. We will use the following general notation:

$$Y = \mathcal{F}(U_1, \ldots, U_K) \tag{1.1}$$

where K denotes the total number of model inputs, U_j denotes a given input of the model, Y denotes the model output and $\mathcal{F}$ denotes the deterministic computer code. When the numerical code $\mathcal{F}$ is based on a mathematical function, we will use the notation f to refer to this function.

Spatial domain: $\Omega \subset \mathbb{R}^2$ denotes a spatial domain, $\mathbf{x} \in \Omega$ is a point of the spatial domain (or a "*location*"), and $v \subset \Omega$ is a connected subset, or "*block*", "*support*", "*spatial unit*", "*zone*" or "*region*", of spatial domain Ω (Figure 1.1 on the facing page). $|v|$ denotes the surface area of block v.

Spatially distributed model inputs: When a model input is spatially distributed over a spatial domain Ω, we will often denote it by Z_i instead of U_i— in order to stick to classical notations used in geostatisics. The overall model input is denoted by $\{Z_i(\mathbf{x}) : \mathbf{x} \in \Omega\}$ or simply Z_i, while $Z_i(\mathbf{x})$ denotes the value of the model input at a particular point $\mathbf{x} \in \Omega$. Nevertheless, for sake of simplicity, we will sometimes imprecisely use the notation $Z_i(\mathbf{x})$ to refer to the overall model input. Non spatially distributed inputs will be refered to as "*scalar inputs*", numbered from U_1 to U_k, while spatially distributed inputs will be numbered from Z_{k+1} to Z_K. $\mathbf{U} = (U_1, \ldots, U_k)$ will denote the set of non spatially distributed inputs only.

[a] We will later discuss in §2.4 on page 72 how our case study, the NOE modelling chain, fits or not in these different categories.

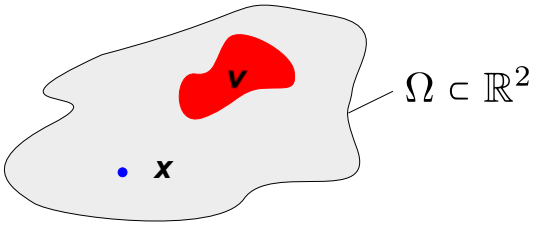

Figure 1.1: Spatial domain Ω

Spatially distributed model output: When a model output is spatially distributed over a spatial domain Ω, we will denote the overall model output by $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$ or simply Y, while $Y(\mathbf{x})$ will denote the value of the model output at a particular point $\mathbf{x} \in \Omega$. Nevertheless, for sake of simplicity, we will sometimes imprecisely use the notation $Y(\mathbf{x})$ to refer to the overall model output.

A spatial model with both spatially distributed inputs and output will thus be represented by the equation (Figure 1.2):

$$\begin{aligned} Y &= \mathcal{F}[U_1, \ldots, U_k, Z_{k+1}, \ldots, Z_K] \\ &= \mathcal{F}[\mathbf{U}, Z_{k+1}, \ldots, Z_K] \end{aligned}$$

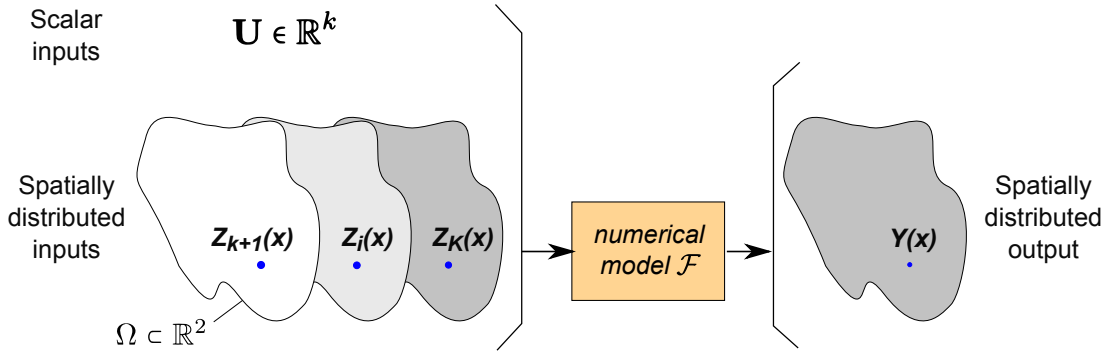

Figure 1.2: A spatial model with both spatially distributed inputs and output

1.1.2 Representing space in numerical models

In spatial models, spatially distributed data (model inputs or output) can be represented in various ways. The choice of a space representation depends on the nature of the data handled. The two most common types of representation are (Bordin 2002):

- a regular grid of cells often refered to as a "*mesh*" or "*grid*", denoted by $\mathcal{G}$. Data are given or computed at each centre point $\mathbf{x}_i$ of cell c_i of the mesh. This grid representation is mostly chosen for continuous physical quantities of interest, such as temperature, rainfall or soil properties. Data represented on a grid are usually stored as so-called "*raster data*" or simply "*raster*" in a GIS software (Figure 1.3 on the following page, left);
- a continuous and object-oriented representation, where objects under study are located in the usual 2D Euclidian space endowed with a Cartesian coordinate system. Objects can be point-shaped, linear, polygonal or have a more complex shape. This representation is usually chosen for discrete objects, such as trees, buildings, roads, plots of land. A convenient storage format for this type of representation is a vector layer in a GIS software (Figure 1.3 on the next page, right).

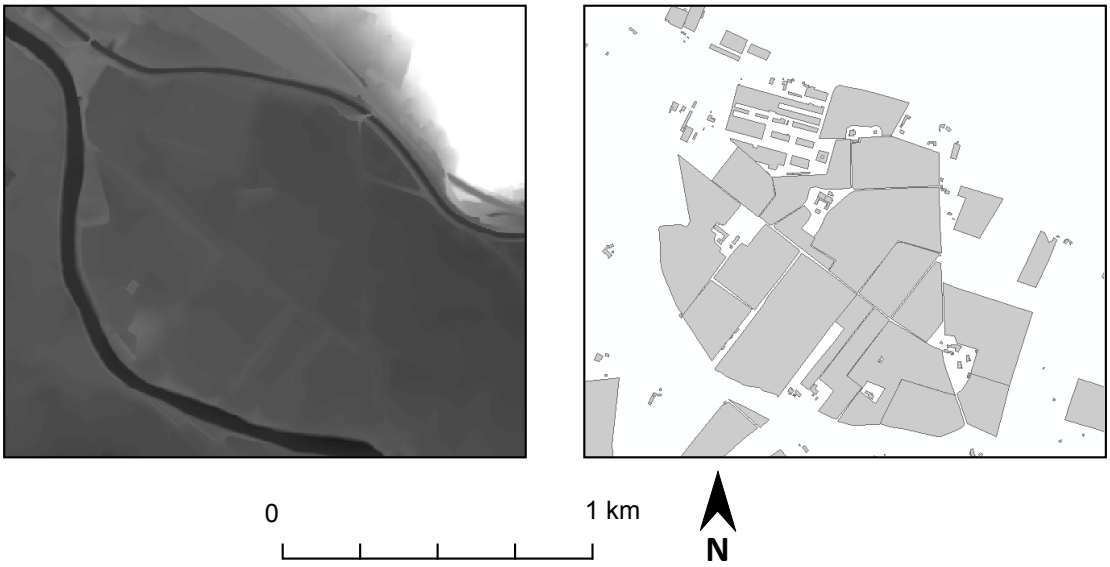

Figure 1.3: Representation of spatially distributed data: raster data (left: a zoom on the Digital Terrain Model of the Orb Delta, cell size 5 m by 5 m); vector data (right: a zoom on the map of flood-exposed buildings and plots of land on the Orb Delta)

In a single numerical model, these two ways of representing space (raster/vector) may be combined. It is not in the scope of this document to describe other types of more complex space representations, in which ad-hoc topologies are designed based on the properties of the system and processes under study[b].

1.1.3 Scale issues

A key notion in spatially distributed modelling is that of "*scale*", and more specifically "*scale in space*". This vague term can refer to a number of characteristic lengths related to a process, a set of observations, or a model. We offer to clarify this notion based on the work of Blöschl and Sivapalan (1995), who wrote an exhaustive review on scale issues in hydrology and Wu et al. (2006) who discuss scaling in ecology.

Types of scale Blöschl and Sivapalan (1995) make a first distinction between i) the "*process scale*" or "*intrinsic scale*", which is the scale at which the phenomenon of interest operates, it cannot be chosen or modified by the modeller; ii) the "*observational scale*", "*measurement scale*" or "*sampling scale*" which is the scale at which measurements are taken; and iii) the "*modelling scale*" or "*analysis scale*" which is the scale at which a model is built. These various types of scales are of course related in some ways; generally speaking, they must be commensurate with each other in order to build a relevant spatial model: processes should ideally be observed and modelled at the scale they occur.

Components of scale: the scale triplet In addition, Blöschl and Sivapalan (1995) also suggest to distinguish between three components of scale, which they refer to as the "*scale triplet*". The components of this triplet depend on the type of scale considered (process, observational or model scale). The scale triplet related to the "*process scale*" is composed of:

[b]For example, such ad-hoc topologies can be found in many 1D hydraulic models, in which the river stream is used as a spatial reference and side storage cells are only positionned by a scalar distance away from the origin of the stream.

- the "*spatial extent*" of the process, that is, the overall spatial expanse of the phenomenon under study (e.g., the total surface area of a floodplain);
- the "*space period*" of the process, if it is periodic (e.g., the distance between dunes in the desert);
- the "*correlation length*" for a stochastic process that exhibits some spatial correlation (e.g., the range parameter of a Gaussian Random Field).

The scale triplet for the "*observational scale*" and the "*modelling scale*" is slightly different; it is composed of (Figure 1.4):

- the "*spatial extent*", that is, the overall spatial expanse of a dataset or the study area covered by a model;
- the "*support*" (or "*resolution*", "*grain*"), which is the finest resolution in space of a dataset within which homogeneity is assumed. When data is represented on a regular grid (raster data), spatial support is governed by the size $|c|$ of cells c_i. When data is represented in a continuous space with an object-oriented approach (vector data), spatial support is related to the size of the smallest resolvable objects (the minimum mapping units);
- the "*spacing*", that is, the characteristic spatial gap between two data points. For raster data, the spacing coincides with the notion of support, and is governed by the size $|c|$ of cells c_i. For vector data, the notion of spacing is relevant only if the objects do not entirely cover the 2D space: spacing is then the characteristic distance between two objects.

All these notions are different but are often named with the same term "*scale*" in the literature. In this thesis, we will try to clarify the meaning of "*scale*" anytime we use it.

Scaling "*Scaling*" is a major issue in spatially distributed modelling. This term is used to refer to the problem of translating knowledge from one scale to another scale (Heuvelink 1998). For example, numerical models are sometimes developed at small spatial scale (e.g., the scale of a single plant) but are expected to produce indicators at a larger scale (e.g., a plot of cultivated land): this problem is known as "*upscaling*". The opposite issue (translating information from large scale to smaller) is known as "*downscaling*". In this thesis, we will focus on the procedure of "*upscaling*"; Blöschl and Sivapalan (1995) suggest to split this procedure into two steps (Figure 1.5 on the next page): first step consists of

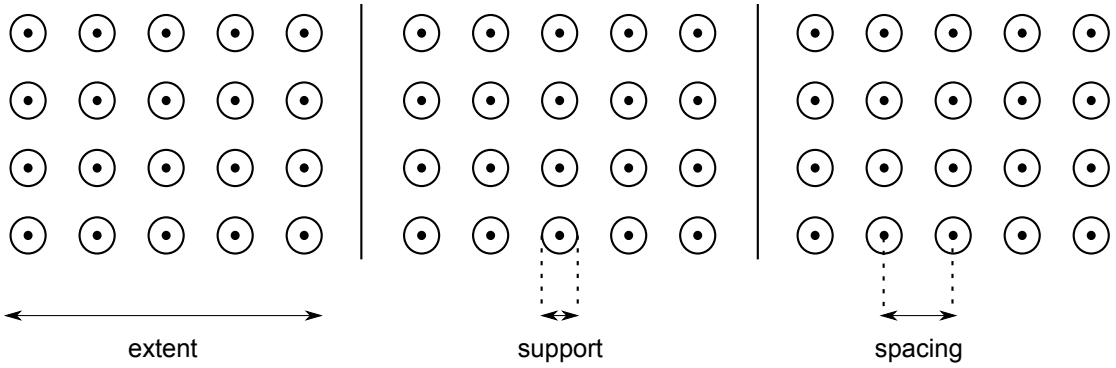

Figure 1.4: Three alternative definitions of scale in space: spatial extent (a), support (b), and spacing (c)
Source : Blöschl and Sivapalan (1995)

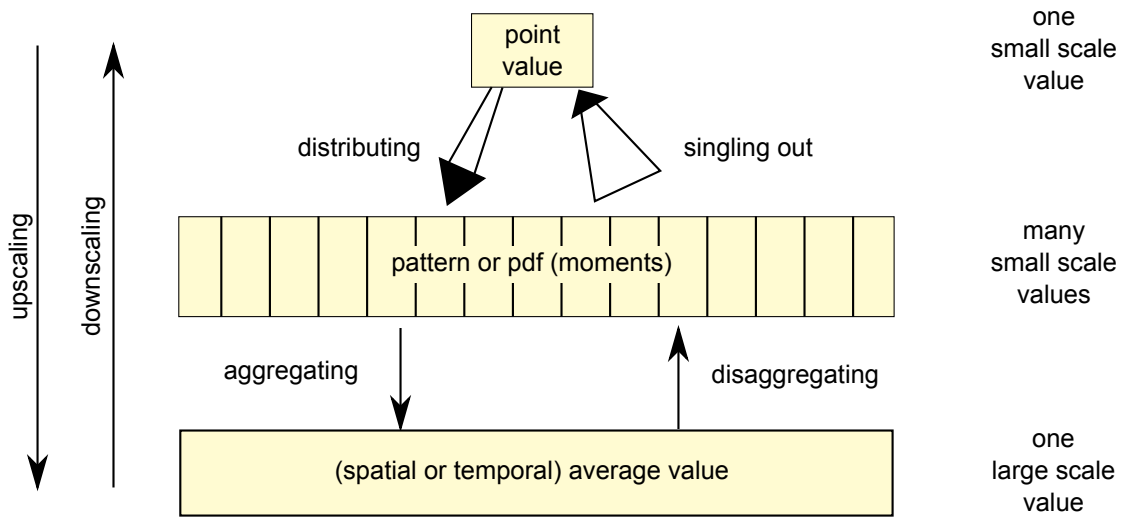

Figure 1.5: Upscaling as a two-steps procedure
Source : Blöschl and Sivapalan (1995)

"*distributing*" a number of small scale values to cover a large spatial extent with many small scale values (e.g., producing a regular grid of harvestable yield across a plot of cultivated land from a small set of point values); second step consists of "*aggregating*" these small scale values to produce a single large scale value (e.g., calculating the total yield of a plot of cultivated land). We will use these notions of "*upscaling*" and "*aggregating*" in the following chapters.

1.1.4 Spatially additive models

A family of spatial models will receive a particular attention in our research: the "*spatially additive models*".

In the case where model output $Y(\mathbf{x})$ is spatially distributed, an important issue is the choice of the spatial support over which it is observed. Model end-users may be interested in the overall spatially distributed output $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$ over spatial domain Ω (e.g., a map of annual rainfall over a catchment, a flood risk map or a pollution map). But they may also want to study a single scalar property of the output over a given spatial unit $v \subset \Omega$: for example, the sum of $Y(\mathbf{x})$ over v (e.g., total rainfall over a catchment), the average value of $Y(\mathbf{x})$ over v (e.g., the average porosity of a geological block), the maximum value of $Y(\mathbf{x})$ over v (e.g., the maximal pollutant concentration over a study area), some quantile of $Y(\mathbf{x})$ over v, or the percentage of v for which $Y(\mathbf{x})$ exceeds a certain threshold, etc. As mentioned in §1.1.3 on the preceding page, moving from spatially distributed $Y(\mathbf{x})$ to a single scalar property over a given spatial support ν is known as "*aggregating*", which is one of the two steps of the "*upscaling*" procedure.

Aggregating many small scale values into a single large scale value may be very complicated, especially when non-linearities are involved: for example, computing the hydraulic conductivity of a large support v from many values of conductivity on smaller supports is not straightforward. Nevertheless, in many environmental models, the physical quantities considered are spatially additive, that is, their large-scale properties derive from small-scale properties by simple linear averaging (Chilès and Delfiner 1999). For example, porosity, evapotranspiration or the daily amount of rainfall are spatially additive variables.

In this document, we will say that a spatial model $\mathcal{F}$ is "*spatially additive*" when two conditions are met, as shown in Figure 1.6 on the next page:

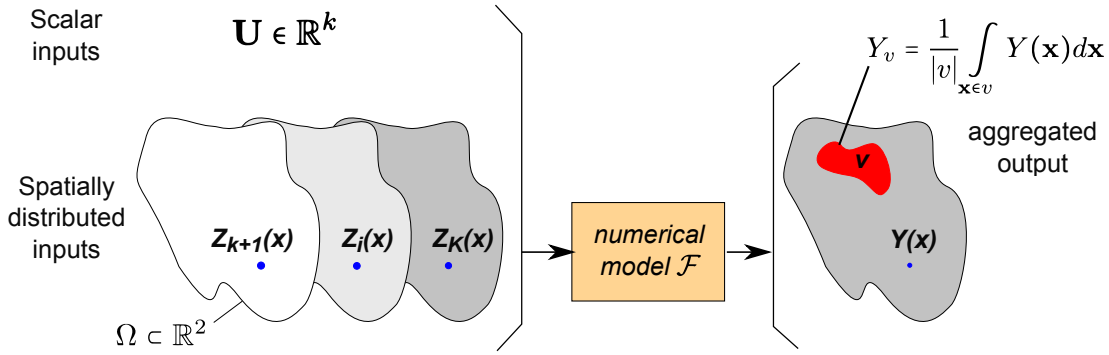

Figure 1.6: A spatially additive model

- it has a spatially distributed output $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$;
- the model end user is interested in the spatial linear average or the sum of the model output $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$ over a given spatial unit $v \subset \Omega$.

We will denote by Y_v the spatial average of the model output over a zone $v \subset \Omega$:

$$Y_v = \frac{1}{|v|}\int_{\mathbf{x}\in v} Y(\mathbf{x})d\mathbf{x} \tag{1.2}$$

When the spatially distributed output $Y(\mathbf{x})$ is only computed at a number of point $\mathbf{x}_i$ of the domain Ω, then the output of interest is a weighted sum of the form $\sum_i w_i Y(\mathbf{x}_i)$.

1.1.5 Point-based models

In our work, particular attention will also be paid to another limited class of spatial models, in which (Figure 1.7 on the following page):

- at least some of the model inputs are spatially distributed;
- model output is spatially distributed;
- the value of model output $Y(\mathbf{x})$ at a given location point $\mathbf{x} \in \Omega$ depends on the set of scalar inputs $\mathbf{U} = (U_1, \ldots, U_k)$ and on the value of spatially distributed inputs $Z_{k+1}(\mathbf{x}), \ldots, Z_K(\mathbf{x})$ **at that same location $\mathbf{x}$ only**.

Following Heuvelink et al. (2010a), we will use the term "*point-based models*" to refer to this class of models. They are encountered in various fields of environmental and earth sciences, whenever spatial interactions in the physical processes under study can be neglected in a first approximation. Some examples of point-based models are: models that predict crop growth, evapotranspiration, pesticide leaching (GeoPEARL model, Tiktak et al. (2002)) or greenhouse gas emission. On the contrary, non point-based models involve some spatial interactions in the description of the physical processes under study: for example, models that simulate river flow routing are usually not point-based, as the water flow at a location $\mathbf{x} \in \Omega$ depends on the flow at other locations $\mathbf{x}' \in \Omega$ upstream.

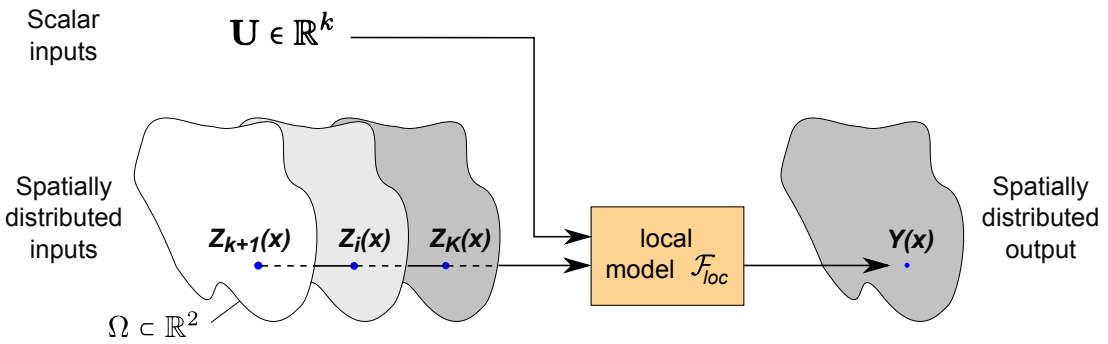

Figure 1.7: A point-based model

In a point-based model, the model output $Y(\mathbf{x})$ at a given point $\mathbf{x} \in \Omega$ is a deterministic function of scalar inputs $\mathbf{U}$ and of the values of spatially distributed inputs $Z_{k+1}(\mathbf{x}), \ldots, Z_K(\mathbf{x})$ at that same location $\mathbf{x}$; we will use the terms "*local code*" or "*local model*", denoted by $\mathcal{F}_{\text{loc}}$, to refer to this function. Hence, we will say that a numerical model $\mathcal{F}$ is "*point-based*" if there exists a local code $\mathcal{F}_{\text{loc}}$ such that:

$$\exists \mathcal{F}_{\text{loc}} : \mathbb{R}^K \to \mathbb{R}, \quad \forall \mathbf{x} \in \Omega, \quad Y(\mathbf{x}) = \mathcal{F}_{\text{loc}}\left[\mathbf{U}, Z_{k+1}(\mathbf{x}), \ldots, Z_K(\mathbf{x})\right] \tag{1.3}$$

We insist on the fact that $\mathcal{F}$ and $\mathcal{F}_{\text{loc}}$ are not the same functions: $\mathcal{F}$ is a function that takes as inputs the overall spatially distributed inputs $\{Z_i(\mathbf{x}) : \mathbf{x} \in \Omega\}$ and the scalar inputs $\mathbf{U}$ and gives as an output the overall field $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$. $\mathcal{F}_{\text{loc}}$ is a function from $\mathbb{R}^K$ to $\mathbb{R}$ that computes $Y(\mathbf{x})$ at a single point $\mathbf{x} \in \Omega$ from the scalar inputs $\mathbf{U}$ and from the values $Z_i(\mathbf{x})$ of spatially distributed inputs at that same point $\mathbf{x}$.

1.2 Sensitivity analysis

This second section aims to briefly introduce the reader to sensitivity analysis (SA) and to describe into more details one specific family of sensitivity analysis methods, on which we focused in our work: the variance-based global sensitivity analysis (VB-GSA). This section is a literature review without any innovative content. We give a broad overview rather than an exhaustive review of SA and VB-GSA methods: the interested reader will find more details in Saltelli et al. (2008) or de Rocquigny et al. (2008).

1.2.1 An overview of sensitivity analysis methods

"*Sensitivity analysis*" (SA) is better defined in relation to "*uncertainty analysis*" (UA). Both terms are closely related but refer to distinct approaches. They gather a number of methods that aim at understanding how sensitive models are to uncertain knowledge of inputs. "*Uncertainty analysis*" focuses on the propagation of uncertainty sources through the model, and tries to quantify the resulting uncertainty on model output. It allows robustness of model results to be checked. "*Sensitivity analysis*" goes one step further: it is used to study how the uncertainty of a model output can be apportioned to different sources of uncertainty in the model inputs. Sensitivity analysis aims at ranking sources of uncertainty according to their influence on the variability of the model output. This ranking helps to identify inputs that should be better scrutinized in order to reduce the variability of the model output. More generally, SA is also useful to explore the response surface of a numerical model and to prioritize the possibly numerous processes that are involved in it.

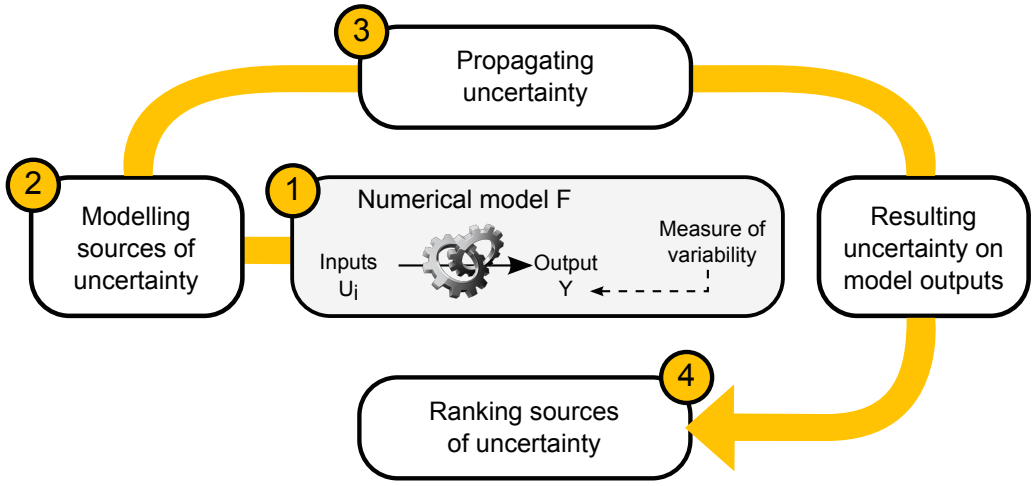

Figure 1.8: General flowchart for sensitivity analysis
Source : adapted from de Rocquigny et al. (2008)

SA is now recognized as an essential component of model building (European Commission 2009b; CREM 2009) and is widely used in different fields of environmental and earth sciences (Tarantola et al. 2002; Cariboni et al. 2007). Over the last fifteen years, a wide range of sensitivity analysis techniques have been developed, and a number of guidelines have been suggested to choose the appropriate SA method for a given problem (Saltelli et al. 2008; de Rocquigny et al. 2008; Iooss 2011; Helton and Davis 2006; Pappenberger et al. 2006; Cariboni et al. 2007). Most SA methods follow a similar flowchart composed of four steps (Figure 1.8):

1. description of the model $\mathcal{F}$ under study and choice of a measure of model output variability
2. modelling of uncertainty sources
3. uncertainty propagation
4. ranking of uncertainty sources

We give below some details on each step. For the sake of clarity, we consider a numerical model $Y = \mathcal{F}(U_1, \ldots, U_K)$ where both model inputs U_i and model output Y are all scalar variables (non spatially distributed).

1.2.1.1 Measure of model output variability

In order to study the variability of the model output Y, one must choose a measure of this variability. Various measures can be considered, such as: the variance of model output $\mathrm{var}(Y)$; the derivative of model output $\partial Y / \partial U_i$ at a given point in the input space; the probability $\mathbb{P}(Y \geq Y_{\max})$ that model output exceeds a given threshold; the overall pdf or cdf of the model output, etc. One can also consider the overall variability of the model output Y in a qualitative way, using graphical methods.

1.2.1.2 Modelling uncertainty sources

Model inputs U_i are almost always fraught with uncertainties, which arise from inaccuracy or lack of data, model assumptions, measurement errors, incomplete knowledge, etc. One main distinction can be made between aleatory uncertainty, which is related to the natural variability of the quantities of interest and which can not be reduced, and epistemic uncertainty, which stems from a lack of knowledge and can often be reduced—see Walker et al. (2003) and Refsgaard et al. (2007) for an enlightening discussion on the nature of uncertainty, which is not in the scope of this document. Numerous frameworks can be chosen to model these various types of uncertainty on model inputs U_i (de Rocquigny et al. 2008): deterministic frameworks, in which a set of alternative discrete values are associated with each input U_i; probabilistic frameworks, in which inputs U_i are considered as random variables with an identified pdf p_i (O'Hagan 2012); second-order probabilistic frameworks, in which parameters of pdf p_i are themselves uncertain; fuzzy logic frameworks, etc.

1.2.1.3 Uncertainty propagation

Once input uncertainties have been identified and modelled, they must be propagated through the numerical model $\mathcal{F}$, in order to assess their impact on the variability of the model output Y. For that purpose, it is necessary to choose a method to "*explore*" the numerical model $\mathcal{F}$. Broadly speaking, we suggest to distinguish four families of methods for model exploration:

- **intrusive methods:** analytical study of mathematical function f or modification of the numerical code $\mathcal{F}$;
- **screening methods:** exploration of numerical code $\mathcal{F}$ based on a relatively small number of simulations (typically < 1000), by varying the value of inputs U_i in a deterministic way; this family of approaches is described in the Design of Experiments (DOE) literature;
- **intensive sampling-based methods:** these methods resort to extensive exploration of the space of possible model inputs with ensuing multiple runs (typically > 1000) of numerical code $\mathcal{F}$, usually using some sort of random or quasi-random sampling in the space of input factors U_i;
- **meta-modelling:** approximation of numerical code $\mathcal{F}$ by a surrogate model $\mathcal{F}'$ with smaller CPU cost, using a small number of simulations of the original code $\mathcal{F}$ (typically < 1000). Meta-modelling (also know as "*emulation*") is not really a method for model exploration on its own: it is a way to get a simplified (less CPU intensive) numerical code that mimics the initial code $\mathcal{F}$. Surrogate model $\mathcal{F}'$ must then be explored by one of the above methods (intrusive, screening or intensive sampling). Ratto et al. (2012) present some applications of meta-modelling techniques to sensitivity analysis of environmental models.

1.2.1.4 Ranking model inputs

Once uncertainty propagation is completed (§1.2.1.3), sensitivity analysis aims at ranking uncertain model inputs U_i based on their impact on the variability of model output Y—which is measured as discussed in §1.2.1.1 on the previous page. Some SA methods produce a qualitative ranking of model inputs, usually separating them into two groups: those inputs U_i that have a large influence on the variability of Y, and those that do not. Other SA approaches are quantitative: they are based on the calculation of so-called "*importance measures*" for each model input U_i with respect to the measure of variability of

model output Y. These importance measures are often refered to as "*sensitivity indices*". They depend not only on the equation of the model $\mathcal{F}$ under study, but also on the uncertainty range of each model input U_i.[c]

1.2.1.5 Local/global sensitivity analysis

SA techniques are also often divided into "*local*" or "*global*" methods. Local methods only consider the variation of model output Y when model inputs U_i vary locally around their nominal values at a single base-point in the input space; they are related to the partial derivatives of the model $\mathcal{F}$ with respect to each input U_i. On the contrary, global methods study the variation of Y when model inputs U_i explore more widely the space of input uncertainties, with all inputs U_i varying at the same time. Global methods are generally more costly than local methods in term of computational burden, but they give a more complete information.

1.2.1.6 Choice of an appropriate SA method

We display in Table 1.1 on the following page the main methods of sensitivity analysis discussed in the literature, classified according to: i) their measure of model output variability, and ii) their approach for uncertainty propagation. In an operational context, Iooss (2011) suggests to use three criteria to choose an appropriate SA method to study a given numerical model $\mathcal{F}$:

- the type of information needed—qualitative or quantitative ranking of model inputs, identification of the most/least influent inputs, etc. Saltelli et al. (2008) use the term "*SA setting*" to refer to the various objectives of a sensitivity analysis (factor fixing FF, factor prioritizing FP);
- possible hypotheses on model complexity—linearity, monotonicity, regularity, etc.;
- the CPU cost of the method—it usually depends on the cost of a single simulation of model $\mathcal{F}$, on the number of simulations of $\mathcal{F}$ needed, and on the number K of model inputs.

[c] In that sense, they differ from other "*sensitivity*" meaures which are only related to the partial derivatives of the equations of a model and which do not depend on the uncertainty range or pdf of the inputs.

Table 1.1: Main methods of sensitivity analysis.

Measure of variability	**Method for uncertainty propagation**		
	Intrusive	**Screening**	**Intensive sampling**
Qualitative (graphical approaches)			cobweb plots scatterplots contribution to the sample mean plot (Bolado-Lavin et al. 2009)
Derivatives or Regression	Automatic differentiation, adjoint methods (Vidard 2011) Equation-based local sensitivity analysis (Delenne et al. 2012) Global adjoint SA	One-at-a-time DoE Factorial DoE Morris DoE or Elementary Effects (Morris 1991; Campolongo et al. 2007)	Linear regression (SRC, PCC, Pearson's CC) Rank regression (SRCC, PRCC) (Carlos García-Díaz and Gozalvez-Zafrilla 2012) Derivative-based global sensitivity measures (Kucherenko et al. 2009; Lamboni et al. 2012)
Variance	Analytical computation of variance-based sensitivity indices		Sobol' (Sobol' 1967) Winding stairs (Jansen et al. 1994; Chan et al. 2000) FAST and extended FAST (Cukier et al. 1978; Saltelli et al. 1999; Xu and Gertner 2011) Random Balance Design (Tarantola et al. 2006)
Extreme probabilities	Analytical computation of probabilities		Approximate Reliability Algoritms (FORM, SORM)
Other			Moment independent GSA (Borgonovo et al. 2011, 2012) Importance measures based on CDF (Liu and Homma 2010) Classification and Regression Trees (Breiman et al. 1984; Mishra et al. 2003) Entropy tests Statistical tests (non grid based: Regional Sensitivity Analysis, Cramer Von Mises test, Man-Whitney test; grid based: common means, common distribution or location, common medians, statistical independence)

1.2.2 Variance-based global sensitivity analysis

We focus in this thesis on a specific family of sensitivity analysis methods: the variance-based global sensitivity analysis (VB-GSA). This choice was made at the very beginning of our research in order to narrow down the scope of the work. VB-GSA was chosen because it is one of the most common and popular SA methods. It explores widely the space of uncertain input variables and is suitable for complex models with non-linear effects and interactions among inputs. Other methods of sensitivity analysis (one-at-a-time, adjoint SA ...) are not considered nor discussed in this thesis. In this section, we give a broad presentation of the mathematical basis of VB-GSA. Please refer to Saltelli et al. (2008); Lamboni (2009); Da-Veiga (2007); Lavergne (2006) for more details.

1.2.2.1 Overview

In VB-GSA, the measure of variability of model output Y is its variance $\mathrm{var}(Y)$. Uncertainty on model inputs U_i is described in a probabilistic framework, by identifying a pdf p_i for each model input U_i. Besides, inputs U_i are supposed to be independent—Kucherenko et al. (2012), Li et al. (2010) or Mara and Tarantola (2012) suggest ways to overcome this restrictive hypothesis. Uncertainty is propagated through the model $\mathcal{F}$ either with an intensive sampling-based approach, or using a meta-model. Finally, VB-GSA is quantitative, and leads to the computation of importance measures named "*variance-based sensitivity indices*" of various orders—in most cases, only "*first*" and "*total*" order sensitivity indices are considered.

1.2.2.2 A decomposition of functional variance

In order to introduce the VB-GSA approach, let us first describe the space of model inputs U_i as a K-dimensional unit cube. We consider a square-integrable function $f : [0,1]^K \to \mathbb{R}$. VB-GSA is based on the expansion of f into a sum of elementary functions f_α of increasing dimensionality:

$$\forall (U_1, \cdots, U_K) \in [0;1]^K, \quad f(U_1, \cdots, U_K) = \sum_{\alpha \subseteq \{1,\ldots,K\}} f_\alpha(\mathbf{U}_\alpha) \tag{1.4}$$

where $\mathbf{U}_\alpha = (U_i)_{i \in \alpha}$ denotes a subset of the set of model inputs $(U_1, \cdots, U_K)$, the empty set is denoted by $\mathbf{U}_\varnothing$ and $f_\varnothing$ is a constant. The expansion given in Eqn. (1.4) always exists but is not necessarily unique. It is sometimes refered to as a High-Dimensional Model Representation (HDMR) (Saltelli et al. 2008), and was initially proposed by Hoeffding (1948).

A particular HDMR expansion is obtained when the condition Eqn. (1.5) is met:

$$\forall \alpha \subseteq \{1, \ldots, K\}, \quad \forall i \in \alpha, \quad \int_0^1 f_\alpha \cdot dU_i = 0 \tag{1.5}$$

Under condition Eqn. (1.5), the expansion given in Eqn. (1.4) exists and is unique. Elementary functions f_α are then orthogonal:

$$\forall \alpha, \beta \subseteq \{1, \ldots, K\}^2 \quad \text{with} \quad \alpha \neq \beta, \quad \int_{\mathbf{U} \in [0;1]^K} f_\alpha \cdot f_\beta \cdot d\mathbf{U} = 0 \tag{1.6}$$

To understand this particular and unique HDMR decomposition in terms of sensitivity analysis, let us now switch to a different framework. We now study a model $Y = f(U_1, \ldots, U_K)$, in which U_i are i.i.d.

scalar random variables with uniform pdf in $[0;1]$, such that $\mathbb{E}Y^2 < +\infty$. From Eqn. (1.4) on the previous page and from the property of orthogonality [Eqn. (1.6)], one can derive an expansion of the variance of f as a sum of the variances of elementary functions f_α:

$$\mathrm{var}(f) = \sum_{\alpha \subseteq \{1,\cdots,K\}} \mathrm{var}(f_\alpha) \tag{1.7}$$

Besides, by integrating Eqn. (1.4) with respect to each model input U_i, under condition Eqn. (1.5), we can get an expression of elementary functions f_α:

$$\begin{aligned} f_\varnothing &= \int_{[0;1]^K} f(U_1,\ldots,U_K)\, dU_1 \ldots dU_K \quad \text{[Eqn. (1.4) \& Eqn. (1.5)]} \\ &= \mathbb{E}(Y) \quad \text{(by definition of } \mathbb{E}(Y)\text{)} \end{aligned}$$

$$\begin{aligned} f_1(U_1) &= \int_{[0;1]^{K-1}} f(U_1,\ldots,U_K)\, dU_2 \ldots dU_K - f_\varnothing \quad \text{[Eqn. (1.4) \& Eqn. (1.5)]} \\ &= \mathbb{E}(Y \mid U_1) - \mathbb{E}(Y) \quad \text{(by definition of } \mathbb{E}(Y \mid U_1)\text{)} \end{aligned}$$

$$\begin{aligned} f_{1,2}(U_1,U_2) &= \int_{[0;1]^{K-2}} f(U_1,\ldots,U_K)\, dU_3 \ldots dU_K - f_1(U_1) - f_2(U_2) - f_\varnothing \\ &= \mathbb{E}(Y \mid U_1,U_2) - \mathbb{E}(Y \mid U_1) - \mathbb{E}(Y \mid U_2) + \mathbb{E}(Y) \end{aligned}$$

Elementary functions f_α can thus be written as a linear combination of conditional expectations of model output Y given model inputs U_i. A general expression of elementary function f_α is (see Appendix §A on page 200 for a proof):

$$\forall \alpha \subseteq \{1,\cdots,K\}, \quad f_\alpha = \sum_{\beta \subseteq \alpha} (-1)^{|\alpha|-|\beta|} \cdot \mathbb{E}(Y \mid \mathbf{U}_\beta) \tag{1.8}$$

in which $|\alpha|$ and $|\beta|$ denote the cardinal of subsets α and β, respectively.

It must be noted that the unique HDMR decomposition we obtain is very similar to the ANOVA decomposition scheme (Archer et al. 1997).

1.2.2.3 Definition and properties of variance-based sensitivity indices

Sobol' (1993) used the functional variance decomposition of f given in Eqn. (1.7) to define importance measures named "*variance-based sensitivity indices*"—sometimes now found in the literature as "*Sobol' sensitivity indices*" or simply "*sensitivity indices*".

Definition (First-order sensitivity indices)**.** Sobol' (1993) defines first-order sensitivity index of model input U_i with respect to model output Y, denoted by S_i or S_{U_i}, as the following ratio:

$$\forall i \in \{1,\cdots,K\}, \quad S_i = \frac{\mathrm{var}(f_i)}{\mathrm{var}(Y)} = \frac{\mathrm{var}\left[\mathbb{E}(Y \mid U_i)\right]}{\mathrm{var}(Y)} \tag{1.9}$$

First-order sensitivity index $S_i \in [0;1]$ measures the main effect contribution of the uncertain model input U_i to the variance of model output Y. It is the expected part of output variance that could be reduced by fixing the value of the uncertain input U_i.

Definition (Higher order sensitivity indices)**.** For any subset $\mathbf{U}_\alpha = (U_i)_{i\in\alpha}$ of model inputs, with $\alpha \subseteq \{1,\dots,K\}$, high order sensitivity index S_α is defined as:

$$\forall \alpha \subseteq \{1,\dots,K\}, \quad S_\alpha = \frac{\text{var}(f_\alpha)}{\text{var}(f)} \tag{1.10}$$

The "*order*" of sensitivity index S_α is equal to the cardinal $|\alpha|$ of subset α, that is, the number of inputs U_i included in subset $\mathbf{U}_\alpha$.

It follows from variance decomposition [Eqn. (1.7) on the facing page] that sensitivity indices S_α of all orders sum up to 1:

$$\sum_{\alpha\subseteq\{1,\dots,K\}} S_\alpha = 1 \tag{1.11}$$

As a consequence, sum of first-order sensitivity indices S_i is always equal or smaller than 1:

$$\sum_{i\in\{1,\dots,K\}} S_i \leq 1 \tag{1.12}$$

The difference $1 - \sum_i S_i$ accounts for the contribution of all interactions between model inputs U_i to the model output variance $\text{var}(Y)$.

Definition (Total-order sensitivity indices)**.** Finally, Homma and Saltelli (1996) suggest to define another importance measure named "*total-order sensitivity indices*", denoted by ST_i or ST_{U_i}, as the sum of all sensitivity indices in which model input U_i is involved:

$$\forall i \in \{1,\dots,K\}, \quad ST_i = \sum_{\alpha\subseteq\{1,\dots,K\},\, i\in\alpha} S_\alpha \tag{1.13}$$

$ST_i \in [0;1]$ represents the part of output variance that is explained by model input U_i and all its interactions with other inputs U_j. It is the expected residual part of output variance if all model inputs but U_i were fixed. It was shown (Saltelli et al. 2008) that total-order sensitivity indices ST_i can also be written as:

$$ST_i = \frac{\mathbb{E}\left[\text{var}(Y \mid \mathbf{U}_{\sim i})\right]}{\text{var}(Y)} \tag{1.14}$$

in which $\mathbf{U}_{\sim i} = (U_j)_{j\neq i}$ denotes all model inputs but U_i. It follows from Eqn. (1.11) and Eqn. (1.13) that the sum of total-order sensitivity indices always sum up to more than 1:

$$\sum_{i\in\{1,\dots,K\}} ST_i \geq 1 \tag{1.15}$$

Sensitivity indices can be used to identify the model inputs that account for most of the model output variability (U_i with high first-order indices S_i); they may lead to model simplification by identifying model inputs that have little influence on the model output variance (U_i with low total-order sensitivity indices ST_i); they also allow discussing the contribution of interactions between model inputs to the model output variance (comparison between first and total-order sensitivity indices).

1.2.2.4 Grouping model inputs

VB-GSA offers the possibility to define first and total-order variance-based sensitivity indices associated with a "*group of model inputs*". Let consider a group $\mathbf{U}_\alpha = (U_i)_{i\in\alpha}$ of model inputs with α a subset of $\{1,\dots,K\}$. First and total-order indices of the group, which we denote by $S_\alpha^{\text{gr.}}$ and $ST_\alpha^{\text{gr.}}$, are defined by:

$$S_\alpha^{\text{gr.}} = \sum_{\beta\subseteq\alpha} S_\beta \quad \text{and} \quad ST_\alpha^{\text{gr.}} = \sum_{\substack{\beta\subseteq\{1,\dots,K\} \\ \alpha\cap\beta\neq\emptyset}} S_\beta \tag{1.16}$$

Grouping model inputs is often used to lower the CPU cost of the estimation of sensitivity indices when the number K of inputs is too large (typically $K \geq 20$). Besides, grouping is useful to cope with correlated inputs (such inputs can be gathered into a single group). Finally, we suggest that grouping can also be helpful to perform VB-GSA in a sequential way: first, a small number of input groups is identifed—usually, the composition of the groups has some physical meaning for the modeller, e.g., in a crop model a parameter describing clay content of a soil horizon will be grouped with other model inputs describing soil properties, while an input related to daily rainfall will be grouped with other climate variables. Then, a first VB-GSA is performed to estimate sensitivity indices associated with each group of inputs. If a group appears to have a large influence on the variance of model output (high group indices $S_\alpha^{\text{gr.}}$ and $ST_\alpha^{\text{gr.}}$), then it will be split into a number of smaller groups. A second VB-GSA will be performed to identify which of these smaller groups are the most influential, and so on, until the most influential individual inputs U_i are identified.

1.2.2.5 Estimation of variance-based sensitivity indices

Reminder: we only briefly present here the estimation of variance-based sensitivity indices for a numerical model $Y = \mathcal{F}(U_1,\dots,U_K)$ where all model inputs U_i are scalar random variables (non spatially distributed). The case of spatially distributed inputs is one of our research items and will be discussed in Chapter 3.

The calculation of variance-based sensitivity indices is closely related to the choice of a method for the exploration of numerical model $\mathcal{F}$ (§1.2.1.3 on page 26): intrusive methods, screening methods, intensive sampling-based methods or meta-modelling. First, for some models $\mathcal{F}$, it may possible to derive the exact analytical expression of sensitivity indices from the equations of the model (intrusive approach). Regarding non-intensive screening methods, there is, to our knowledge, no specific design of experiments nor any ad-hoc estimators to calculate variance-based sensitivity indices[d]. On the contrary, a large body of scientific literature is available on various intensive sampling strategies and associated estimators to calculate variance-based first-order and total-order sensitivity indices from a large number of simulations of a numerical code $\mathcal{F}$. These techniques follow a similar three steps procedure: i) generate an input matrix through an appropriate random sampling method, usually some Quasi Monte Carlo sampling scheme or other space-filling design (Pronzato and Mueller 2012); ii) calculate an output vector by evaluating the numerical code $\mathcal{F}$ at each line of the input matrix ; iii) estimate variance-based sensitivity indices from the output vector. Here are some of the most popular methods encountered in the literature: Sobol' estimators using LP-τ samples (Sobol' 1993), Winding Stairs approach (Jansen et al. 1994; Chan et al. 2000), FAST sampling (Cukier et al. 1978; Saltelli et al. 1999), Random Balance Design (Tarantola et al. 2006), use of replicated Latin Hypercube Sampling (Tong 2010), etc. It is not in the scope of this document to make a

[d] Campolongo et al. (2011) shows the similarity between the Elementary Effects approach and VB-GSA and suggest to use a "*radial one-at-a-time*" design of experiment to estimate total-order sensitivity indices.

complete review of these sampling strategies and estimators: the interested reader will find more details in Saltelli et al. (2008); de Rocquigny et al. (2008); Lilburne and Tarantola (2009); Gatelli et al. (2009).

In our research, we have always used the same algorithm to estimate first and total-order variance-based sensitivity indices, using the estimation procedure suggested by Saltelli et al. (2010):

- generate two input samples M_1 and M_2 of size $K \times N$ in the space of model inputs, where K is the number of model inputs (or groups of model inputs) and N will be refered to as the "*base sample size*". The two samples are $LP-\tau$ sequences where each input U_i is sampled from its pdf p_i. The choice of a base sample size N depends on the accuracy needed for sensitivity indices estimates; it has to be of the form 2^m to ensure that the desired properties of the $LP-\tau$ sequences hold;
- combine these two samples M_1 and M_2 to generate a new, longer sample M_{tot} of size $K \times N_{\text{tot}}$ where N_{tot} will be refered to as the "*total sample size*". N_{tot} depends on the base sample size and on the number of model inputs (or groups of model inputs): $N_{\text{tot}} = (K+2) \cdot N$. Details on the combination procedure are given in Appendix §A on page 200;
- calculate an output vector by evaluating the numerical code $\mathcal{F}$ at each line of the total sample M_{tot};
- estimate sensitivity indices for each model input U_i (or each group of model inputs) from the output vector. The estimators are those suggested by Saltelli et al. (2010); they are given in Appendix §A on page 200.

1.3 Chapter conclusion

In this chapter, we have given some elements of theoretical background, definitions and notations to better specify the subject of our research.

First we have clarified the notion of "*spatial model*": for us, this term will refer to a numerical model in which some of the inputs and/or some of the outputs are spatially dsitributed over a 2D spatial domain. In such numerical models, spatially distributed inputs/output can be stored as raster or vector data. Besides, the notion of "*spatial scale*" can be better grasped through the scale triplet of "*extent*", "*support*" and "*spacing*". We have also given the definitions of two subclasses of spatial models that we will focus on: i) the spatially additive models, in which the model end user is interested in the spatial linear average or the sum of some quantity of interest over a given spatial unit, and ii) the point-based models, in which spatial interactions in the physical processes under study can be neglected in a first approximation.

Then, we have displayed a brief overview of sensitivity analysis techniques and presented the four steps they are composed of: i) definition of the model under study and choice of measure of model output variability; ii) modelling of uncertainty sources; iii) uncertainty propagation; and iv) ranking of model inputs. We finally portrayed into more details the theoretical basis of variance-based global sensitivity analysis (VB-GSA). We defined the first and total-order variance-based sensivity indices, and explained which estimation procedure will be used throughout our work.

From this material, we are now able to give a more detailed description of the two methodological questions that we will try to answer in the following chapters (Figure 2 on page 14):

(1) handling spatially distributed inputs in VB-GSA: the estimation procedure presented in §1.2.2.5 on the preceding page to compute variance-based sensitivity indices is only appropriate for scalar

and independent model inputs U_i. A first question is: *how to compute variance-based sensitivity indices S_Z and ST_Z for a spatially distributed model input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, that may exhibit spatial auto-correlation?* We will try to adress this issue in Chapter 3 with a very pragmatic perspective, by exploring the various numerical tricks that can be used to compute S_Z and ST_Z. Besides, we will focus our research on intensive sampling-based methods to estimate S_Z and ST_Z, thus ignoring a number of other possibilities discussed in §1.2.1.3 on page 26 (intrusive methods, screening or meta-modelling);

(2) scale issues in VB-GSA: the theoretical background of VB-GSA presented in this chapter does not account for the notion of "*scale*", which is of utmost importance in spatial modelling (§1.1.3 on page 20). We will try to adress this issue in Chapter 4 by using the concepts, notations and results from the geostatistics theory. More specifically, we will focus our research on the specific case of spatially additive and point-based models. We will try to answer the following question: *in a spatially additive and point-based model, how do variance-based sensitivity indices depend on the scale of the model output Y, and more precisely on the three components of the scale triplet: support, spacing and extent?*

Chapter 2

Building a modelling framework for cost-benefit analysis of flood management plans

FLOOD risk research makes intensive use of numerical modelling. Over the last decades, many numerical models have been developed in order to i) forecast flooding events, ii) assess the impacts of potential flooding events in terms of human well-being, economic and social development, and iii) design efficient flood management policies. These models simulate hydrological, hydraulic and economic processes over a given study area. They are usually spatially distributed and often make use of GIS tools.

In this thesis, we look into one specific family of flood-related models, which are based on a common approach for the economic appraisal of flood risk management plans: the "*cost-benefit analysis based on avoided damages*" approach (CBA-AD). As mentioned in the general introduction, the applied objective of the thesis is to investigate the propagation of uncertainty through this CBA-AD approach. For that purpose, we first had to design a general modelling framework to describe the CBA-AD approach, in order to perform its uncertainty and sensitivity analysis. This modelling framework, named NOE, was then implemented into a computer code and applied on a number of case studies, including the Orb Delta study site, which will be used as a real-world test case for VB-GSA of spatial models throughout this document.

The goal of this chapter is to present both the NOE modelling framework and its application to the Orb Delta study site. It is divided into three sections. The first section §2.1 starts with some elements of context on flood risk management and the economic appraisal of flood mitigation projects. It introduces the reader to the CBA-AD approach and surveys the literature on the subject. Next, the second section §2.2 explains why and how we built a modelling framework named NOE to describe the CBA-AD approach. As we will detail it, the NOE modelling framework does not pretend to cover the entire variety of CBA-AD studies, but rather to clarify the structure, inputs and outputs of the CBA-AD flowchart, in the view toward performing its uncertainty and sensitivity analysis. Then, section §2.3 briefly mentions how we implemented the NOE modelling chain into an efficient computer code, and presents the various case studies on which the NOE code has been applied. In particular, it portrays the Orb Delta study site, which will be used as the main case study in the thesis. Finally, in the chapter conclusion (§2.4), we make some general comments on the scope and limitations of the NOE modelling framework, and stress a few key points that will prove important to carry out its sensitivity analysis.

2.1 Cost-benefit analysis of flood risk management plans 38
2.1.1 Flood risk and flood management policies 38
2.1.2 Economic appraisal of flood management policies 38
2.1.3 Cost-benefit analysis based on avoided damages (CBA-AD) 38
2.2 NOE modelling framework for cost-benefit analysis based on avoided damages 40
2.2.1 Motivations & methods 40
2.2.2 Some definitions first 41
2.2.3 Overview of the NOE modelling framework 41
2.2.4 Comparison of costs and benefits 42
2.2.5 Average Annual Avoided Damages 44
2.2.5.1 Modelling flooding events as random variables 45
2.2.5.2 Definition of average annual damages: multidimensional case 46
2.2.5.3 Unidimensional case 47
2.2.5.4 Approximation 48
2.2.6 Choice of flood scenarios 48
2.2.7 Weights of flood scenarios 49
2.2.7.1 Unidimensional case: flood frequency analysis 49
2.2.7.2 Multi-dimensional case 50
2.2.7.3 Two comments 51
2.2.8 Flood damage estimation 51
2.2.8.1 Flood hazard modelling 53
2.2.8.2 Flood exposure modelling 54
2.2.8.3 Damage costs estimation 56
2.3 Case studies 58
2.3.1 Development of the NOE code 58
2.3.2 The Orb Delta case study 59
2.3.2.1 Study site 59
2.3.2.2 Model flowchart 60
2.3.2.3 Model inputs 64
2.3.2.4 Model outputs 65
2.3.2.5 Model scales 67
2.3.3 Other case studies 70
2.4 Chapter conclusion 72
2.4.1 A brief summary 72
2.4.2 Main features of the NOE modelling framework 72
2.4.3 Key limitations of the NOE modelling framework 73

2.1 Cost-benefit analysis of flood risk management plans

2.1.1 Flood risk and flood management policies

Flooding is one of the most damaging natural hazards in Europe. In France, floods threaten around two millions people, one third of municipalities and cost over 250M€ per year—approximately 80% of economic losses due to natural hazards (MEDD 2004). These figures are worryingly likely to rise in the next decades, as flood-prone areas continue to be developed under the pressure of demographic and economic growth. In addition, climate change may result in more extreme meteorological events and increase flood frequencies and magnitudes (Parry et al. 2007).

In order to prevent floods, protect people and assets from flood damage and prepare our societies to face such extreme events, public authorities implement various flood risk management policies. In the past, these policies were mainly concerned with structural measures designed to prevent flood hazard (e.g., levees, dams, channel improvement, etc.). Nowadays, flood risk management policies also include non-structural measures that aim at i) reducing vulnerability of flood-exposed assets; ii) controlling land planning in flood-prone areas; and iii) improving crisis management (Ledoux 2006). In Europe, since the approval of the EU Flood Directive (2007/60/EC) in 2007, member states have to establish flood risk management plans combining prevention, protection and prepardness, in all flood-prone river basins and coastal areas. In France, a similar framework was adopted in 2003, when integrated flood management plans named PAPI ("*Programmes d'Actions de Prévention des Inondations*") were introduced (MEDDTL 2011).

2.1.2 Economic appraisal of flood management policies

For project prioritisation, planning and monitoring, the assessment of flood risk management plans is needed. These appraisals must consider the various aspects of a flood management plan, including its economic efficiency, its technical feasibility, its environmental and social impacts, etc. They require a good understanding of the components of flood risk (flood hazard, vulnerability, resilience, etc.), and use knowledge from numerous fields of Science such as hydrology, hydraulic, economy (Messner et al. 2007; Hubert and Ledoux 1999; Ledoux et al. 2003).

One of the most common methods for the appraisal of flood risk management plans is the "*cost-benefit analysis*" approach (CBA). CBA assesses the economic efficiency of a policy by comparing its costs and benefits, both converted to a monetary unit, over a conventional length of time (European Commission 2008). This approach is recommended for the economic appraisal of flood risk management plans in the EU Flood Directive. In France, it is compulsory for local authorities and water managers to produce a CBA of their PAPI flood management plan when they claim national subsidies (MEDDTL 2011).

2.1.3 Cost-benefit analysis based on avoided damages (CBA-AD)

Generally speaking, all costs related to a flood management policy should be included in a CBA: direct investment and maintenance costs, but also monetized negative social impacts or environmental impacts of the policy, such as landscape deterioration. In the same way, all benefits of the policy should be considered, including possible social benefits such as the recreational use of a lake resulting from a dam construction. However, for practical reasons, most CBA studies only account for the "*main*"—in terms on monetary amount—costs and benefits of the policy. Hence, assessing policy costs often comes down

to the estimation of the direct investment and maintenance costs. In addition, it is most often assumed that the largest share of the benefits can be measured by the monetized amount of flood damages that will be spared thanks to the flood risk management policy. We will use the term "*Cost-Benefit Analysis based on Avoided Damages*" (CBA-AD) to refer to this somehow restrictive framework for CBA of flood risk management policies (Erdlenbruch et al. 2008). CBA-AD is almost the only approach encountered in French case studies for the appraisal of structural flood management plans, and is also widely used in other European countries.

In the literature, CBA-AD studies vary in their aim, their scope, their scale and the data used. In order to propose a general modelling framework for CBA-AD studies—as we will explain it later in §2.2 on the following page—we surveyed a number of academic papers and reports that discuss the use of this approach for the economic assessment of flood risk management policies. Messner et al. (2007) display guidance and recommendations for the economic appraisal of flood risk, based on the outcomes of the European project FLOODsite[a]; they propose a flowchart for cost-benefit analysis based on avoided damages (Messner et al. 2007 p.21), give a clear overview of the best practices in Europe and survey a number of CBA-AD studies and other guidelines across Europe (Messner et al. 2007 p. 63). Merz et al. (2010) extensively discuss the issue of flood damage assessment. The Flood Hazard Reseach Center[b] published a handbook on assessment techniques to evaluate the benefits of a flood risk management policy, known as the Multi-Coloured Manual (Penning-Rowsell et al. 2005). The Queensland Government also issued guidance for flood damage assessment (DNRM 2002), as well as the US Army Corps of Engineers (Baecher et al. 2000). Bournot (2008) reviewed thirty French projects of flood risk management and presented how their economic relevance was assessed. Jonkman et al. (2004) discussed the use of cost-benefit analysis for flood damage mitigation policies in the Netherlands. Erdlenbruch et al. (2008) presented a CBA-AD study on the Orb river (Hérault, France). Finally, Achleitner et al. (2010) studied the flood protection structures in the Ötztal valley (Tyrol, Austria).

We also looked into technical reports that describe into details how cost-benefit analyses were carried out for four different flood risk management plans in France. A first group of three case studies are located along the Rhône River.[c] These studies include: a project of dike strengthening and heightening on the river reach between Fourques and Beaucaire; a project to improve the water storage capacity of the two small islands of La Motte and L'Oiselet, upstream of Avignon; a larger project that aims at renovating old floodplains along the lower reaches of the Rhône river. A last case study was considered in Bretagne in the Vilaine floodplain close to the city of Redon.[d]

Finally, we also read the existing guidelines that French national or local authorities have published on the CBA-AD approach for the economic appraisal of flood risk management plans. These guidelines aim to help experts in environmental consultancy firms produce relevant and rigorous CBA-AD assessments, and assist the local contracting authorities in their claim for public subsidies. We surveyed four guidelines, published respectively by: the Plan Rhône, a public water management body along the Rhône river (Ledoux Consultants 2010); the European Center for Flood Risk Prevention[e] (CEPRI 2011); the Gard Water Committee (GERI 2012); and the Ministry of Ecology (MEDDTL 2011 Appendix 4).

This literature review does not pretend to be exhaustive. Its main objective was to give us the foundations to build a general modelling framework that could depict the CBA-AD studies. This modelling framework named NOE is presented in next subsection §2.2 on the following page.

[a] http://www.floodsite.net

[b] An interdisciplinary centre based at Middlesex University, http://www.mdx.ac.uk/research/areas/geography/flood-hazard

[c] We worked on these case studies as part of a larger project on uncertainties in cost-benefit analysis of flood risk management plans, funded by the Plan Rhône (http://www.planrhone.fr).

[d] These four case studies will be further described in §2.3.3 on page 70.

[e] CEPRI, http://www.cepri.net

2.2 NOE modelling framework for cost-benefit analysis based on avoided damages

In this section, we give a detailed description of the NOE modelling framework, whose purpose is to depict cost-benefit analyses of flood risk management plans based on the "*avoided damages*" approach (CBA-AD studies, §2.1.3 on page 38). The motivations and methods of this modelling work are given in §2.2.1. They are followed by some important definitions in §2.2.2, and by a detailed description of the NOE modelling chain from §2.2.3 to §2.2.8. The aspects related to the numerical implementation of the NOE model and its application to different case studies will be discussed in the next section §2.3 on page 58.

2.2.1 Motivations & methods

As mentioned in the chapter introduction, the applied goal of this thesis is to carry out an uncertainty and sensitivity analysis of the CBA-AD studies. As a preliminary step, we had to design and implement a modelling framework to describe these CBA-AD studies. We briefly expound here our motivations and objectives for this modelling work.

Motivations Two reasons motivated our need to design a modelling framework for CBA-AD studies. First, as explained by Saltelli et al. (2008) or de Rocquigny et al. (2008), the very first step of an uncertainty and sensitivity analysis is to properly define the system under study: one needs to specify the boundary of the model, its inputs, its outputs. Second, as mentioned in §1.2.2.5, most sensitivity analysis techniques require to run many simulations of the model under study (e.g., more than thousands simulations are needed in the sensitivity analysis methods based on intensive sampling). Hence, another preliminary step of uncertainty/sensitivity analysis is to implement the model into a convenient computer code, that can be run a thousand times or more with reasonable effort from the modeller. To meet these two requirements for sensitivity analysis, it thus appeared necessary to design a modelling framework for CBA-AD studies and to develop an efficient computer code for it.

Methods CBA-AD studies broadly consist of a comparison of the costs of a flood management plan with its benefits, which are measured by the monetized amount of flood damages that will be avoided, each year on average, thanks to the plan. In the academic papers, technical reports and guidelines we surveyed, the CBA-AD studies vary in their aim, their scope, their scale and the data used (§2.1.3 on page 38). Nevertheless, most of them follow a similar flowchart that combines hydrological, hydraulic and economic modelling as well as GIS-based spatial analysis. Our method here is to propose a general modelling framework, named NOE, that could depict this common flowchart. We insist on the fact that our work is by no means an attempt to make an exhaustive review of cost-benefit analyses applied to flood management policies. Instead, the NOE modelling framework aims to give a clear description of the CBA-AD studies in order to:

- define a number of terms that will be used throughout this thesis;
- cleary define the boundaries, the inputs and the outputs of the system under study, which is a requirement to carry out sensitivity analysis.

We built the NOE model so that it could account for the various practices that we encountered in our literature review. We will describe how well these various practices fit into the modelling framework we suggest to adopt. However, we do not pretend to cover entirely the variety of CBA-AD studies applied to flood risk assessments: hence, we will also discuss the limitations of the NOE framework when needed.

2.2.2 Some definitions first

Before describing the flowchart of the NOE model, we give definitions of some important terms. It is not the scope of this manuscript to define general and widely used notions such as "*flood risk*", "*flood hazard*", "*exposure*" or "*vulnerability*". The interested reader will find such definitions in European Commission (2010) or Ledoux (2006).

Study area We will use the term "*study area*" to refer to a well-identified and limited floodplain, characterised by its geophysical attributes, but also its landuse and all man-made infrastructures, buildings and activities. Typical size of floodplains considered in CBA-AD studies ranges from 10 to 1 000 sq. km.

Flood management policy We will equally use the terms "*flood management plan*" or "*flood management policy*" to refer to a set of structural and non-structural measures designed to reduce flood risk on a study area, by preventing flood hazard and/or reducing vulnerability of assets. We will more precisely use the term "*flood-control measures*" to refer to structural measures only, such as levee and dam construction or channel improvement.

Present/future situation We will use the terms "*present situation*" and "*future situation*" to describe the state of a study area in relation to a given flood management plan: "*present situation*" (resp. "*future situation*") will refer to the state of the area before (resp. after) the implementation of the flood risk management plan (Figure 2.1).

2.2.3 Overview of the NOE modelling framework

We choose to describe the NOE model as a combination of "*steps*" or "*modules*" (Figure 2.2 on page 43). These steps may be not be clearly separated in all the CBA-AD studies we reviewed, but they will prove useful to clarify the description of input data, intermediate outputs and final outputs. The very last step of

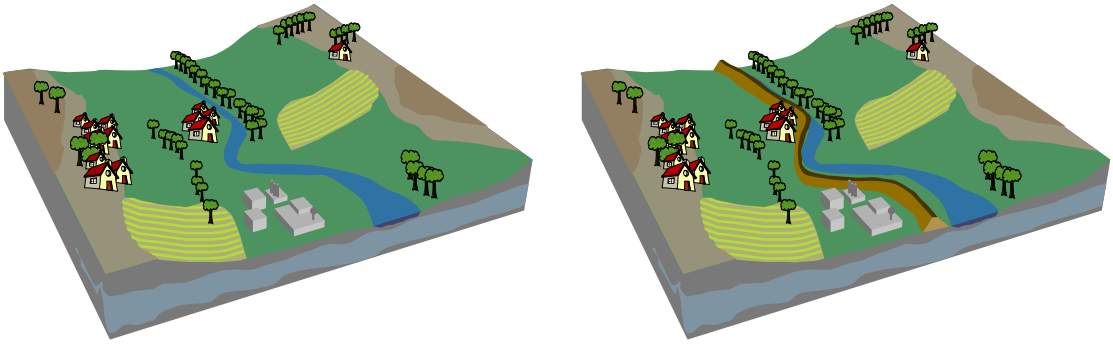

Figure 2.1: Flood-prone study area: present situation (left); future situation after levee construction (right)

the NOE flowchart is the comparison of the costs and benefits of a flood management policy, that leads to the computation of performance indicators such as the Net Present Value (NPV [€]). The expected benefits of a flood management policy are partly based on the estimation of the so-called "*Average Annual Avoided Damages*" (ΔAAD) [€/year]), which is the amount of annual expected flood losses that is reduced thanks to the policy. The computation of the ΔAAD indicator requires to estimate flood damages and probabilities of occurence for a range a flood scenarios of various magnitudes. All these steps are described into more details in §2.2.4 to §2.2.8 on pages 42–51. We made the deliberate choice to start the description from the end (bottom) of the NOE modelling flowchart.

2.2.4 Comparison of costs and benefits

The ultimate goal of CBA-AD studies is to produce an indicator measuring the economic efficiency of a flood management plan, by comparing its benefits with its costs. Both flow of benefits and flow of costs are observed over a given length of time R (usually R ranges from 30 to 50 years). B_i (resp. C_i) denotes the expected benefits (resp. costs) of the plan at year i. Benefits and costs are adjusted for the time value of money and converted into present value amounts using a discount rate, denoted by τ_i at year i. Both R and discount rates τ_i have conventional values that are usually fixed by national or European authorities (European Commission 2008). Two main efficiency indicators are met in the literature to compare discounted costs and benefits:[f]

- the Net Present Value of the policy (NPV [€]), defined as:

$$NPV = \sum_{i=0}^{R} \tau_i \cdot (B_i - C_i) \tag{2.1}$$

- the Hicks ratio (H [dimensionless]), defined as:

$$H = \frac{\sum_{i=0}^{R} \tau_i \cdot B_i}{\sum_{i=0}^{R} \tau_i \cdot C_i} \tag{2.2}$$

A positive NPV—or a Hicks ratio greater than 1—indicates that the benefits generated by the flood risk management plan outweigh its costs. The larger the NPV value is, the more efficient the policy is.

Costs of the policy The costs of a flood management plan usually include at least the initial investment costs CI and the maintenance costs CM—CEPRI (2011) details the various components of these two costs. A common assumption found in CBA-AD studies is to attribute investment costs CI to time step $i = 0$, and to consider that maintenance costs CM are constant over the length of time R. This is just an average view of reality; if detailed data are available on these costs—e.g., if successive phases of the flood management plan are scheduled— investment and maintenance costs can be assumed to take different values CI_i and CM_i at each time step i.

Other costs related to the flood management plan, such as environmental impacts or landscape degradation, are most often neglected. One notable exception is the CBA-AD study carried out on the ZEC

[f] Other indicators that will not be discussed in this thesis include the return on investment and the internal rate of return (Bournot 2008).

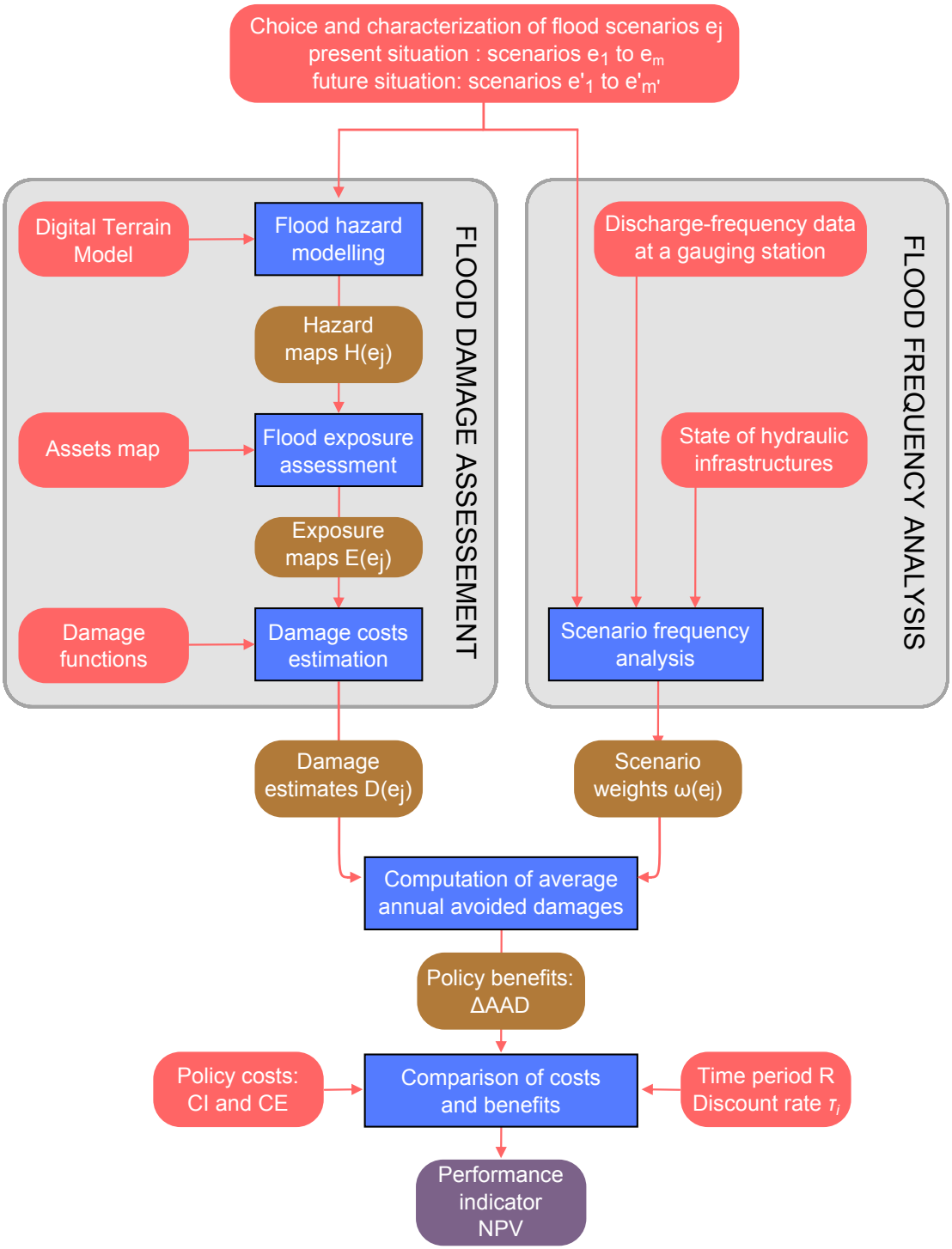

Figure 2.2: General flowchart of the NOE modelling framework

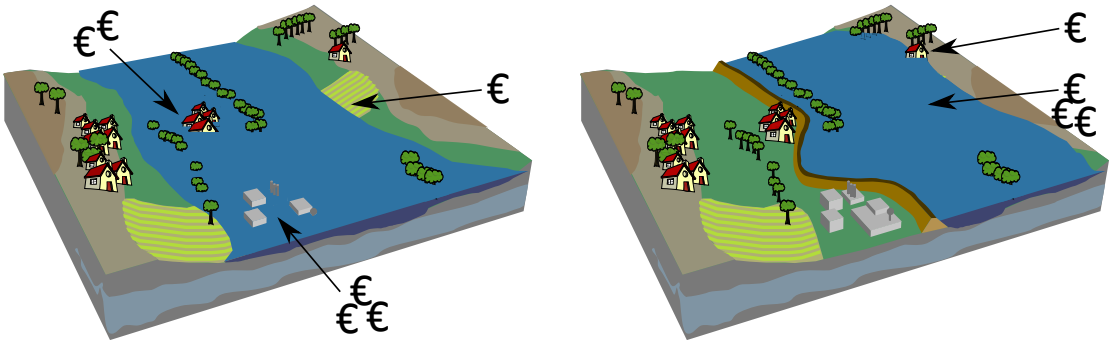

Figure 2.3: Flood-prone study area: reduction of total flood damages thanks to a flood risk management plan. Present situation (left) and future situation (right)

project (§2.3.3 on page 70): in this study, the subsidies received by some farmers in compensation for the overflooding of their agricultural parcels was included as an additional cost in the analysis.

However, in the NOE modelling framework, we suggest to stick to the most common case, in which only investment costs CI and maintenance costs CM are considered, with the following assumptions:

$$\forall i \in [\![0; R]\!], \quad C_i = \begin{cases} CI & \text{if } i = 0 \\ CM & \text{if } i > 0 \end{cases} \tag{2.3}$$

Benefits of the policy In the CBA-AD approach, the benefits of a flood management plan are measured by the amount of flood damages that it can avoid yearly on average. We will use the term "*Average Annual Avoided Damages*" and the notation ΔAAD [€/year] to refer to this amount. Benefits are supposed to be equal to the ΔAAD indicator at each time step i:

$$\forall i \in [\![0; R]\!], \quad B_i = \Delta AAD \tag{2.4}$$

The ΔAAD indicator is an algebraic value: it may be positive when flood damages are spared thanks to the flood management plan, or negative if extra damages are caused by the policy. We explain in the next subsection how to rigorously define and compute it.

To sum up, in the NOE modelling framework, under the assumptions we made on the costs and the benefits of the flood risk management plan, the NPV indicator is equal to:

$$NPV = -CI + \sum_{i=1}^{R} \tau_i \cdot (\Delta AAD - CM) \tag{2.5}$$

2.2.5 Average Annual Avoided Damages

The ΔAAD indicator is obtained by comparing the amount of expected annual flood damages between the present and the future situation (Figure 2.3). We will denote by AAD [€/year] (resp. AAD') the average annual damages in present (resp. future) situation. The ΔAAD indicator is defined as:

$$\Delta AAD = AAD - AAD' \tag{2.6}$$

The notion of "*average annual damages*" has been used for a long time as a quantitative indicator for the assessment of flood risk. Early works on this subject were notably initiated in the late 1950ies in the

Netherlands by Van Dantzig (1956) who carried out a probabilistic analysis of Dutch flood defences and tried to estimate the expectation of the flood damage per year. In the literature, the same notion is also refered to as "*expected annual damages*" (NRC 2000), the "*annual average flood losses*" (Messner et al. 2007) or the "*Weighted Annual Average Damages*" (WAAD) (Messner et al. 2007).[g]

The AAD indicator is used in almost all the CBA-AD studies we reviewed. Some of these studies cast doubt on its relevance as a synthetic risk indicator, or question its definition. Indeed, defining rigorously the "*average annual flood damages*" proves to be a challenging issue. One contribution of the NOE modelling framework is to propose a clear frame to define the AAD indicator. This definition proves useful to better identify the underlying assumptions that are hidden in the notion of average annual damages, and to highlight its limitations. This definition is summarized in the following subsections §2.2.5.2 to §2.2.5.4; the interested reader will find an extended description of our contribution on this subject in Appendix §C on page 206.

2.2.5.1 Modelling flooding events as random variables

A flood can be defined as an overflow or inundation that comes from a river or other body of water onto normally dry land and causes or threatens damage (Figure 2.4 on the following page). In our research work, we only considered fluvial floods, as opposed to coastal floods, groundwater floods or surface water floods[h]. In the NOE modelling framework, we assume that a "*flooding event*" can be entirely described by a finite number $\kappa \in \mathbb{N}$ of scalar parameters. These parameters usually include:

- a set of parameters describing the hydrological load associated with the flooding event, such as the discretised hydrogram at a reference gauging station, the peak discharge or the inflow volume;
- a set of parameters describing the state and behaviour of hydraulic infrastructures along the river stream during the flooding event: e.g., water level beyond dams, failure or not failure of levees;
- the season of occurence of the event (summer, autumn, etc.).

It is of course impossible to predict if and which flooding events will occur during one given year on a study area. This uncertainty related to flooding events can be classified as "*aleatory uncertainty*", that is, natural variability which is associated with the phenomenon under study and which cannot be reduced (Refsgaard et al. 2007). To represent this aleatory uncertainty, we suggest to identify each flooding event with a single realisation e of a random vector $\mathbf{E} = (E_1, \ldots, E_\kappa)$ with values in $\mathbb{R}^\kappa$. We assume that random vector $\mathbf{E}$ has a probability density function p_E. For a given year, $p_E(e)$ represents the probability that flooding event $e \in \mathbb{R}^\kappa$ occurs this year.

All CBA-AD studies and guidelines use a similar probabilistic framework. One specific feature of the NOE modelling framework is that flooding events are not reduced to a real-valued random variable describing their peak discharge q, but are modelled as a random vector $\mathbf{E}$ of dimension $\kappa \geq 1$. This enlarged framework makes it possible to account for complex situations in a unified manner. For example, in the Fourques-Beaucaire case study (§2.3.3 on page 70), flooding events can be modelled by a 4-dimensional random vector, which describes: the peak discharge of the scenario, the season of flood occurence, the state of levee (failure/no failure), and the location of the possible levee failure.

[g] Jonkman et al. (2004) do not use the average annual damages as a risk indicator, but rather compute the (discounted) present value of flood damages for some flooding events.

[h] However, most of the material displayed in the following sections could be adapted to other types of floods with small efforts.

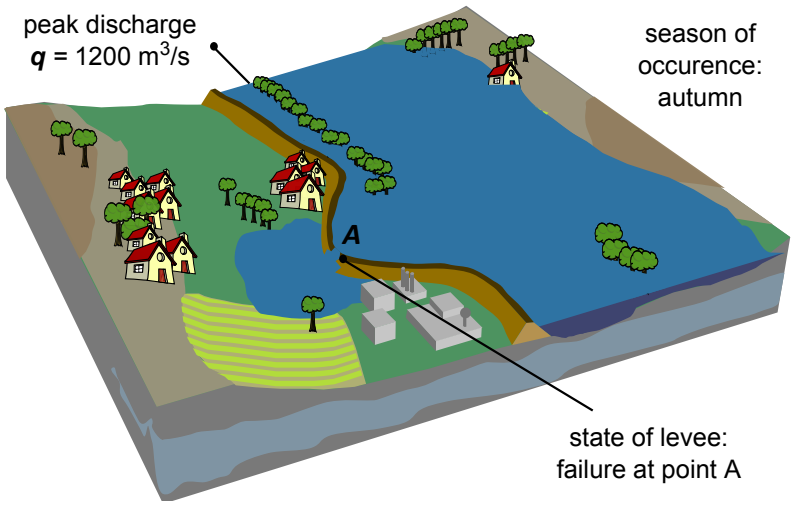

Figure 2.4: Flood-prone study area: a flooding event

2.2.5.2 Definition of average annual damages: multidimensional case

We assume that, on a given study area, the monetized damages associated with flooding events can be represented by a continuous function $D : \mathbb{R}^\kappa \to \mathbb{R}$. For a given flooding event $e \in \mathbb{R}^\kappa$, $D(e)$ is the amount of damage costs [€] induced by this event:

$$\begin{array}{rccc} D: & \mathbb{R}^\kappa & \to & \mathbb{R} \\ & e & \mapsto & D(e) \end{array} \tag{2.7}$$

The Average Annual Damages (AAD) are defined as the expectation of damage D over the space of random flooding events e with pdf $p_E(\cdot)$:

$$AAD = \mathbb{E}(\mathbf{D}) = \int_{\mathbb{R}^\kappa} D(e)\, p_E(e)\, de \tag{2.8}$$

This definition of the AAD indicator is based on at least two restrictive assumptions.

First, we assume that flood damages can be represented as a deterministic function $D(e)$ of a flooding event e characterised by a finite number of scalar descriptors: its hydrological load, the state of hydraulic infrastructures along the stream (dam, levees, etc.), its season of occurence (§2.2.5.1 on the preceding page). Function D must include knowledge about the assets that are exposed to floods in the study area and their vulnerability. This function is supposed to be "*fixed*" over time. Nevertheless, in the real world, total flood damages related to a single flood event $e \in \mathbb{R}^\kappa$ also depend on a number of other random characteristics that are not taken into account in our definition, such as the quality of crisis management or the proximity (in space or time) of other flooding events that may result in a greater vulnerability of the exposed assets, etc. A possible extension would be to include these characteristics as extra scalar descriptors of a flooding event.

Next, we also assume that random variable $D(e)$ has a finite expectation: this assumption requires flood damages not to grow "*too quickly*" for extreme flooding events e with very small probabilities $p_E(e)$, so that the integral Eqn. (2.8) is well defined. In most of the CBA-AD studies we reviewed, the function $D(\cdot)$ is simply assumed to be bounded: this assumption is fully justified when flood damages are measured with repair costs only.

2.2.5.3 Unidimensional case

In most of the CBA-AD studies we surveyed, the probability of occurence of a flooding event e is defined with respect to a single scalar descriptor: its peak discharge q at a reference gauging station. In this case, the dimension κ of the set of flood event descriptors is reduced to 1. Somewhat imprecisely, we will then equally use the notations $\mathbf{E}$ or $\mathbf{Q}$ to denote the scalar random variable representing flooding events, the notations e or q to denote a single flooding event and its associated peak discharge, and the notations $p_E(\cdot)$ or $p_Q(\cdot)$ to denote its pdf. The AAD indicator is then equal to:

$$AAD = \mathbb{E}(\mathbf{D}) = \int_{\mathbb{R}} D(q)\, p_Q(q)\, dq \tag{2.9}$$

Hydrologists usually prefer to consider the cumulative distribution function of random variable $\mathbf{Q}$, or more precisely the function $F_Q : q \mapsto \int_q^{\infty} p_Q(x)dx$. $F_Q(q)$ is refered to as the "*annual exceedance probability*" of peak discharge q. Using this function, Eqn. (2.9) can be written as:

$$AAD = \int_0^1 D \circ F_Q^{-1}(u)\, du \tag{2.10}$$

Eqn. (2.10) was notably used by Arnell (1989) to define the Average Annual Damages. It can be understood as the area under the damage-frequency curve, which is the graph of flood damage $D(q)$ against discharge exceedance probability $F_Q(q)$ (Figure 2.5).

We will use this presentation in some parts of this thesis, when flooding events e will simply be described by their peak discharge q, notably on the Orb Delta study site. Nevertheless, we insist on the fact that in our broader definition of the AAD indicator [Eqn. (2.8) on the preceding page], a flooding event is defined both by a set of hydrological descriptors and a set of parameters describing the state of hydraulic infrastructures along the river, and is represented by a vector $e \in \mathbb{R}^{\kappa}$ with $\kappa \geq 1$.

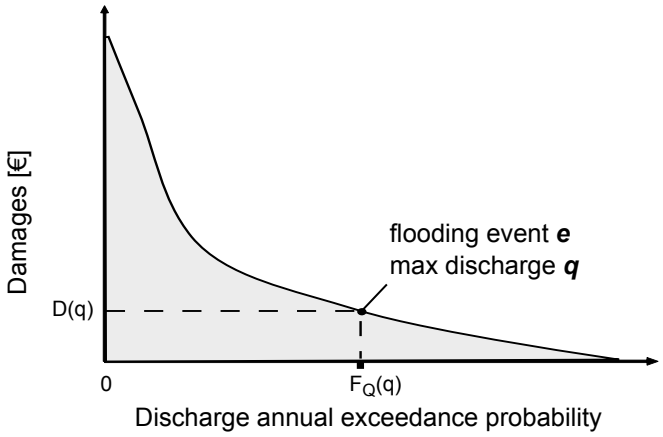

Figure 2.5: Average Annual Damages (AAD [€/year]) in unidimensional case: the AAD indicator is equal to the (shaded) area under the damage-frequency curve

2.2.5.4 Approximation

The various CBA-AD studies and guidelines we surveyed do not exactly use the same calculation method to approximate the AAD indicator. However, these various methods all fit in a common frame: for each situation (present and future), the computation of the AAD indicator is based on the approximation of a unidimensional or multi-dimensional integral [Eqn. (2.8) on page 46 and Eqn. (2.10) on the previous page]. Flood damage estimates are evaluated for a finite set of points e_j that we suggest to call "*flooding scenarios*". A weighted sum of these values is then used to approximate the integral:

$$AAD \approx \sum_j \omega(e_j) \cdot D(e_j) \tag{2.11}$$

where $D(e_j)$ denotes the estimated damage for flood scenario e_j and $\omega(e_j)$ denotes the weight of this scenario. For each situation (present and future), the computation of the AAD indicator thus requires:

- a set a flood scenarios e_j (that may be different from present to future situation);
- the estimation of scenario weights $\omega(e_j)$;
- the estimation of flood damages $D(e_j)$ for each scenario.

The first element may be considered as part of the model itself: it will be discussed in §2.2.6. The computation of weights $\omega(e_j)$ will be discussed in §2.2.7 on the next page. The computation of flood damage estimates $D(e_j)$ will be detailed in §2.2.8 on page 51.

2.2.6 Choice of flood scenarios

As mentioned in §2.2.5.4, the calculation of the AAD indicator requires damage estimation for a number of relevant flood scenarios e_j. The first step of the NOE modelling framework is thus to choose two sets of potential flooding events of various magnitudes: one set of m scenarios, denoted by e_1 to e_m, for the present situation, and one set of m' scenarios, denoted by e'_1 to $e'_{m'}$, for the future situation (Figure 2.6 on the facing page).

In the NOE modelling framework, we suggest to characterise each flood scenario e_j (or e'_j for future situation) by i) a complete description of the hydrological load (hydrogram, peak discharge q_j, inflow volume); ii) a description of the state and behaviour of hydraulic infrastructures along the river reach (failure or not failure of levees, water level in dam reservoir); and iii) its season of occurence. For each situation (present and future), at least two flood scenarios must be considered: e_1 (resp. e'_1) is supposed to be the "*smallest*" flooding event that induces damage in present (resp. future) situation—"*smallest*" is taken here in the sense "*with the smallest peak discharge*"; e_m (resp. $e'_{m'}$) is supposed to be an extreme flood, which would result in an over-topping of all flood-control infrastructures.

This framework fits well to all the CBA-AD studies and guidelines we reviewed, except for the work of Achleitner et al. (2010) (see below). In all studies, the flood scenarios usually include i) some "*historical floods*" that are modelled from ex-post data collected after past flooding events, and ii) some "*synthetic floods*" that are usually related to construction and safety standards of flood-control infrastructures (for example, a synthetic 100-year or 1 000-year flood). The number of scenarios for present and future situation usually falls between 3 and 20[i]. This initial choice (number and characteristics of scenarios) can be considered as part of the model structure; results of a CBA-AD study heavily depend on it.

[i] The guidelines published by the French Ministry of Ecology for the economic appraisal of PAPI management plans requires at least three flood scenarios to be considered (MEDDTL 2011 Appendix 4).

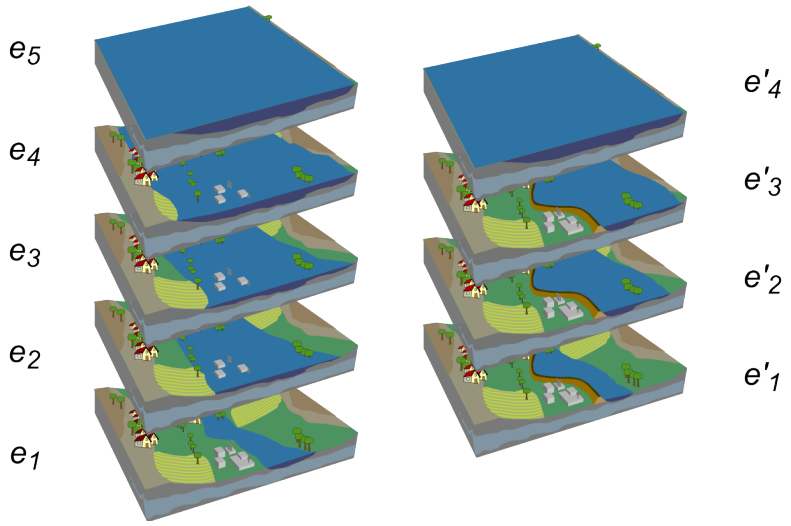

Figure 2.6: Flood-prone study area: set of flood scenarios for the present (left) and future (right) situations

A notable exception: continuous scenarios. It can be noted that a very different framework could be chosen to estimate the AAD indicator and represent the aleatory uncertainty associated with flood hazard. Flood scenarios e_j could be randomly generated over a very large length of time ($\sim 1\,000$ years), in order to build a plausible chronicle of flooding events over time. This kind of continuous simulation of flooding events over time may for example be obtained from rainfall simulations. The number of flood scenarios considered would then be very large ($m, m' \sim 1\,000$). We will use the term "*continuous scenarios*" to denote this approach, which is seldom encountered in the literature—Achleitner et al. (2010) use a similar approach over a short period of 100 years and call it "*stochastic flood and loss modelling framework*".

2.2.7 Weights of flood scenarios

As mentioned in §2.2.5.4 on the preceding page, the calculation of the AAD indicator is based, for each situation (present and future), on a weighted sum of damage estimates $D(e_1), \ldots, D(e_m)$ for a set of flood scenarios $e_1, \ldots, e_m$ (resp. $D(e'_1), \ldots, D(e'_{m'})$ and $e'_1, \ldots, e'_{m'}$ for future situation). The weights associated with each flood scenario e_j, denoted by $\omega(e_j)$, have to be determined according to the approximation technique chosen to calculate the unidimensional or multi-dimensional integral defining the AAD indicator [Eqn. (2.8) on page 46 and Eqn. (2.10) on page 47].

2.2.7.1 Unidimensional case: flood frequency analysis

In the unidimensional case, a flood scenario e_j is only described by its peak discharge q_j. In this case, weights $\omega(e_j)$ are obtained from a hydrological frequency analysis. A discharge-frequency curve (**Q-f**) is first fitted to observed **Q-f** data (annual maximum flow serie on a given gauging station) with a chosen extreme value distribution (Gumbel, gev, etc.) (Figure 2.17 on page 62). Then, each peak discharge q_j is associated with an estimated annual exceedance probability $F_Q(q_j)$ and corresponding return interval $T_j = 1/F_Q(q_j)$. Weights $\omega(e_j)$ are then computed from non-exceedance probabilities $F_Q(q_j)$ in order to approximate the unidimensional integral [Eqn. (2.10) on page 47]. In the NOE modelling framework, we

suggest to use a classical trapezoïdal rule (Figure 2.7):[j]

$$ADD \approx \frac{1}{2} \cdot [D(e_1) + D(e_2)] \cdot [F_Q(e_1) - F_Q(e_2)] + \ldots$$
$$+ \frac{1}{2} \cdot [D(e_{m-1}) + D(e_m)] \cdot [F_Q(e_{m-1}) - F_Q(e_m)] + D(e_m) \cdot F_Q(e_m)$$

which can be written as:

$$ADD \approx \sum_j \omega(e_j) \cdot D(e_j) \tag{2.12}$$

in which weights $\omega(e_j)$ are equal to:

$$\begin{cases} \omega(e_j) = \dfrac{F_Q(e_{j-1}) - F_Q(e_{j+1})}{2}, & \forall i \in \{2, \ldots, m-1\} \\ \omega(e_1) = \dfrac{1}{2} \cdot [F_Q(e_1) - F_Q(e_2)] \\ \omega(e_m) = \dfrac{1}{2} \cdot [F_Q(e_{m-1}) + F_Q(e_m)] \end{cases} \tag{2.13}$$

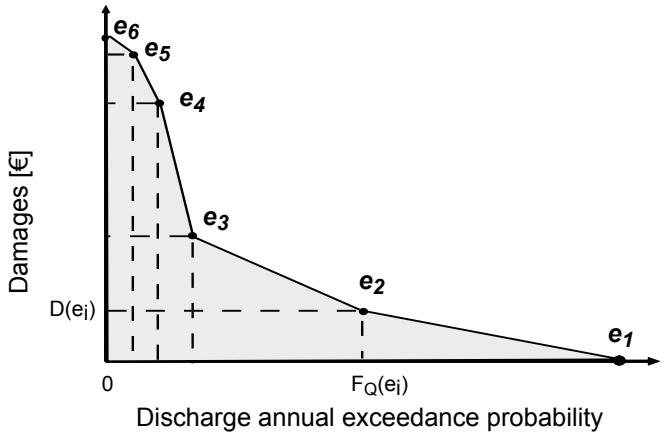

Figure 2.7: Approximation of the AAD indicator [€/year] with trapezoidal rule (unidimensional case)

2.2.7.2 Multi-dimensional case

In the broader definition we suggest to use in the NOE modelling framework, the AAD indicator is equal to a κ-dimensional integral [Eqn. (2.8) on page 46]. We give in Appendix §C on page 206 a rigorous presentation of the way this κ-dimensional integral can be approximated. In short, it can be turned into a unidimensional integral based only on flood exceedance probabilities $F_Q(q)$, but then it requires to

[j] This unidimensional case is by far the most often encountered in the CBA-AD studies we reviewed. However, we observed slight variations in the formula chosen to approximate the unidimensional integral Eqn. (2.10) on page 47. For example, the Weighted Annual Average Damages defined in Messner et al. (2007) approximate the integral with a Riemann sum (rectangle method). This is also the method chosen in the Fourques-Beaucaire case study (ISL 2011). Erdlenbruch et al. (2008) use a different weight $\omega(e_m)$ for the largest flooding event: $\omega(e_m) = 1/2 \cdot \left[F_Q(e_{m-1}) + d_\infty \cdot F_Q(e_m)\right]$ in which d_∞ is an ad-hoc coefficient which is assumed to represent the ratio between the flood damages $D(e_m)$ associated with flood scenario e_m and the flood damages D_∞ associated with a virtual flood event with an exceedance probability equal to 0; they suggest to use a value of $d_\infty = 2$. MEDDTL (2011 Appendix 4) recommends to use the same computation scheme with $d_\infty = 1.5$.

estimate the conditional expectation of flood damages D given a fixed peak discharge q, while other $\kappa - 1$ flood descriptors vary. From a practical perspective, in this case, hydrological flood frequency analysis based on **Q-f** data analysis has to be completed to calculate the overall weight $\omega(e_j)$ of each flood scenario e_j. This overall weight $\omega(e_j)$ must account not only for the exceedance probability of the flood peak discharge $q(e_j)$, but also for the probabilities associated with the other scalar descriptors of the flooding scenario e_j, such as the state of hydraulic infrastructures along the river, or the season.

In the various CBA-AD studies we reviewed, we found only one example in which such multi-dimensional weights $\omega(e_j)$ were computed: the Fourques-Beaucaire study (§2.3.3 on page 70). In this study, the simple weights $\omega(e_j)$ based on flood frequency analysis were completed by: i) a probabilistic levee reliability assessment to estimate probabilities of levee failure for each flood scenario e_j; ii) a seasonal hydrological analysis to estimate the probability that a flood scenario occurs at a given season (Allamano et al. 2011). In an extended framework, it would also be possible to further describe and weigh flooding events e_j with other aleatory characteristics that are not strictly related to flood hazard but that have an impact on the monetary amount of flood damages, such as the quality of crisis management or the state of flood-exposed assets.

2.2.7.3 Two comments

Continuous scenarios If "*continuous scenarios*" framework is chosen to design flood scenarios e_j in a CBA-AD study, then in the unidimensional case the Average Annual Damages is simply the non-weighted average of flood damage estimates for all randomly generated flood scenarios—scenario weights $\omega(e_j)$ are all equal to $1/m$ where m is the number of scenarios.

Spatially heterogeneous scenario weights In some studies, it is not possible to rigorously define a spatially homogeneous weight $\omega(e_j)$ for each flood scenario e_j over a large study area, because hydrological return intervals T_j may spatially depend on the contribution of various river tributaries and lateral inflows. It is then necessary to divide the study area into a number of homogeneous zones where scenario weights will be separately computed. This is the case in the CBA-AD study that was carried out on the ZEC project (§2.3.3 on page 70).

2.2.8 Flood damage estimation

We briefly describe in this subsection the data and flowchart used to calculate flood damage estimates $D(e_j)$ (resp. $D(e'_j)$) for the set of flood scenarios $e_1, \ldots, e_m$ (resp. $e'_1, \ldots, e'_{m'}$) in the present (resp. future) situation. Once again, our purpose is not to give an exhaustive view of the approaches encountered in the literature for flood damage estimation, but rather to clarify a number of terms and notions in order to perform sensitivity analysis of the NOE modelling chain. Generally speaking, the flood damage estimation process for a given flood scenario e_j can be divided into three steps shown in Figure 2.8 on the following page: i) flood hazard modelling; ii) flood exposure modelling; and iii) damage costs estimation.

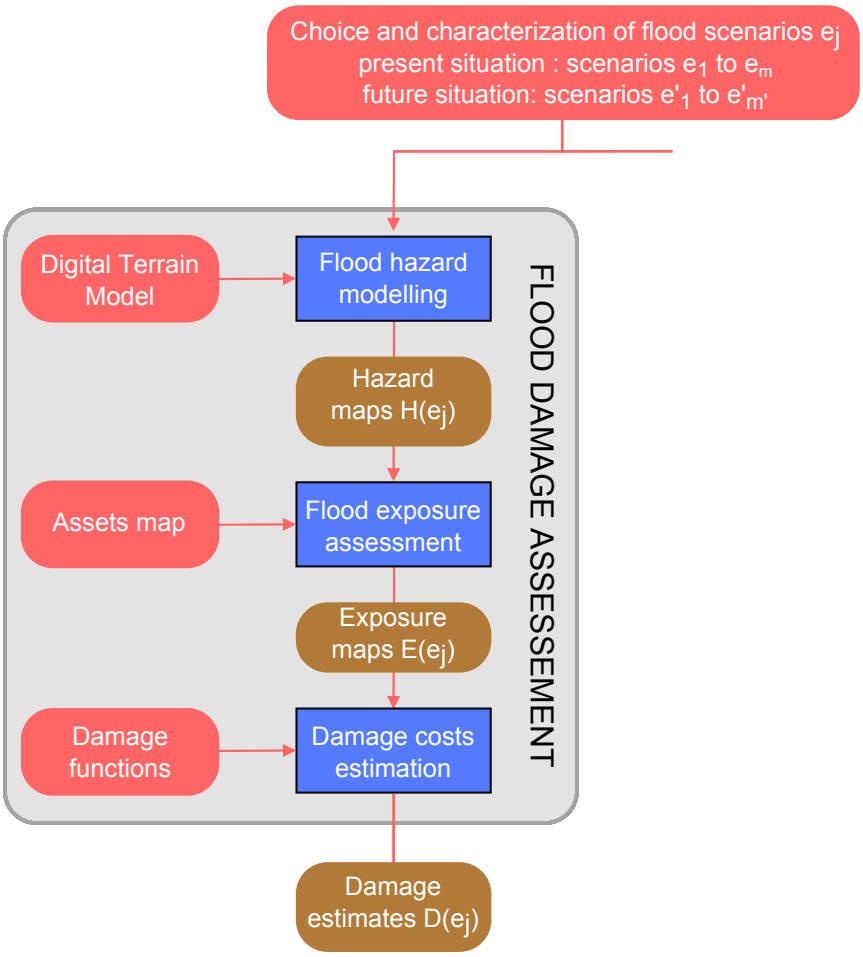

Figure 2.8: Flowchart of the NOE modelling framework: flood damage estimation

2.2.8.1 Flood hazard modelling

For a given flood scenario e_j (present or future situation), flood hazard modelling consists of using an hydraulic model to simulate the flood inundation process (Figure 2.9) based on the scenario characteristics (hydrological load and state of hydraulic infrastructures). A wide range of hydraulic models can be used for that purpose: 1D models, 2D models, storage cells models, with steady or unsteady flow conditions. These models solve various versions of free-surface flow equations. Input data used by hydraulic models fall into the following categories: i) a set of flow boundary conditions, including an hydrogram $Q(t)$ at a reference gauge, ii) a Digital Terrain Model with a precise description of natural and man-made structural elements that control water flow, and iii) a set of spatially distributed friction coefficients used for model calibration (Strickler coefficients). It is not in the scope of this thesis to describe these hydraulic models into more details. The interested reader will find further explanations in Novak et al. (2010).

Flow simulations are then usually combined with a high resolution Digital Terrain Model (DTM), to produce maps giving spatially explicit values of the main flood intensity parameters: water depth [m], water velocity [m/s] and flood duration [h] over the study area.[k] We will use the general term "*hazard maps*" to refer to these maps of flood intensity parameters. The notations $\mathcal{H}(e_j)$ (resp. $\mathcal{H}(e'_j)$) will denote the hazard maps associated with flood scenario e_j (resp. e'_j for future situation). These maps can be produced as GIS vector layers with polygonal features representing small storage cells or finite elements, or can be transformed into raster data (Figure 2.18 on page 63).

Note: in the NOE modelling framework, we will consider the hazard maps $\mathcal{H}(e_j)$ as inputs. The hydraulic model that produces them will not be considered as a part of the NOE model. This restrictive choice was made to reduce the CPU cost of sensitivity analysis and to narrow down the scope of our research.

[k] In some case studies, maps od flood intensity parameters are real-valued, but in other cases they just give intervals—e.g., the maps of water depth used by Erdlenbruch et al. (2008) show classes of water depths in centimeters: $[0; 50]$, $]50; 100]$, etc.

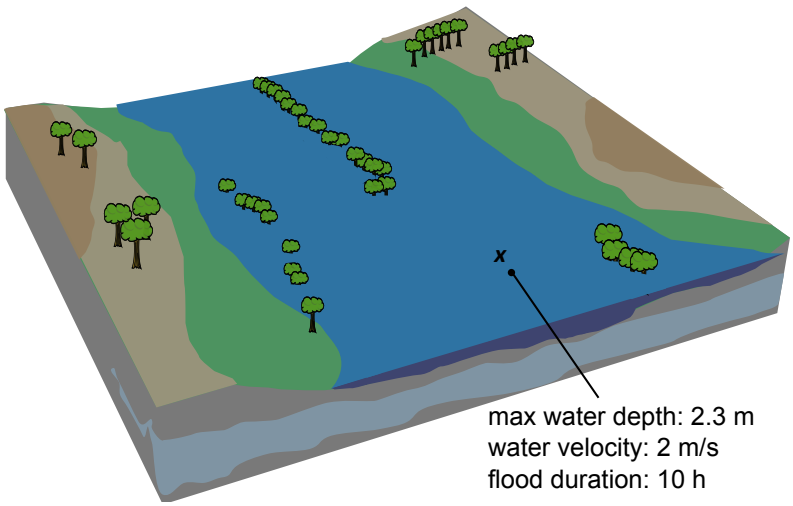

Figure 2.9: Flood-prone study area: inundation process for a given flood scenario

2.2.8.2 Flood exposure modelling

For a given flood scenario e_j (present or future situation), flood exposure modelling consists of: i) describing the assets which are exposed to the flood on the study area; ii) determining for each identified asset the intensity of flood hazard, in terms of water depth, velocity and flood duration.

Description of assets Many types of assets are exposed to floods and may be considered in a CBA-AD study: private housing, agricultural land, economic activities, road network, power supply networks, etc. (Figure 2.10)[1]. In order to derive flood exposure from hazard maps, the description of assets has to be spatially explicit. In this thesis, we will use the term "*assets map*" to refer to this description. In most cases, the assets map is a GIS vector layer, each asset being represented by a point, a line or a polygonal feature. These assets maps are usually created from a combination of data sources, including land use maps, cadastral maps, field surveys, national statistics, etc. From one CBA-AD study to another, assets maps may differ with respect to the following criteria:

- **scale (support):** flood exposure may be assessed at the scale (here understood as the spatial support, see §1.1.3 on page 20) of individual assets (buildings, plots of cultivated land), or at a coarser level (e.g., district level or regional level). The choice of a spatial support for the description of assets usually depends on the size of the area under investigation and on the available data. As developed into details by Messner et al. (2007 Figure 3.5), the finest spatial supports for the description of assets are often restricted to study areas of local size (e.g., municipalities or single floodplains). On the contrary, studies for areas of regional size (e.g., a part of a big river or the catchment of a smaller river) or even national size (e.g., a national coastline or a river basin of a transboundary river) have to rely on approaches which require less effort per unit of area and, consequently, consider a larger spatial support to describe the elements at risk;

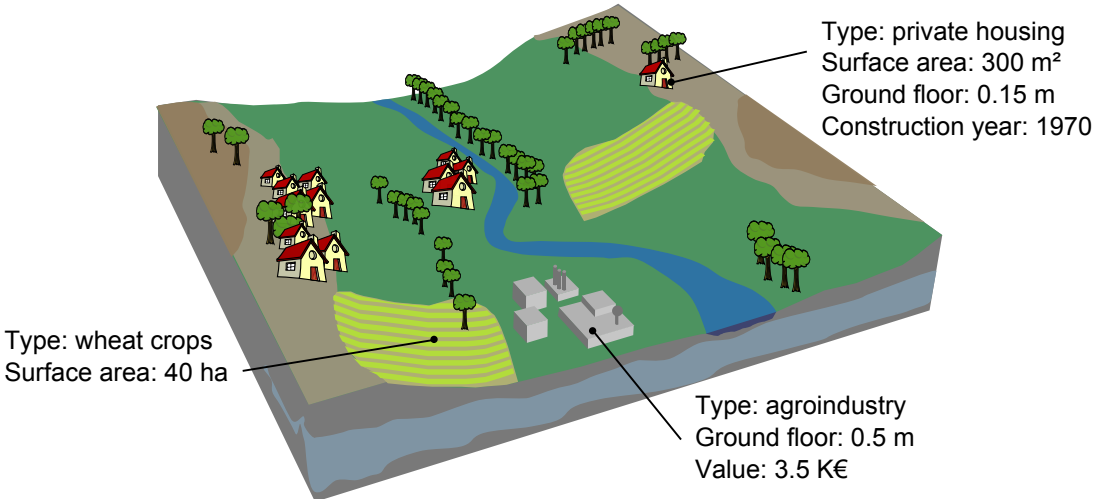

Figure 2.10: Flood-prone study area: spatially explicit description of assets

[1] the guidelines published by the French Ministry of Ecology for the economic appraisal of PAPI management plans requires to consider at least the following types of assets: private housing, economic activities, agricultural land and public buildings (MEDDTL 2011 Appendix 4)

- **nomenclature of assets:** a more or less detailed nomenclature may be used to classify assets according to their type. Nomenclature can range from very coarse (e.g., distinguishing only from housing, agricultural land and economic activities) to very detailed (using various subtypes to further describe crops, types of buildings, types of industry, shops, etc.). The nomenclature has to be chosen in accordance with damage functions (§2.2.8.3 on the next page);
- **attribute data:** each asset is usually characterised by a number of attributes, generally including its ground floor elevation, its monetized value, its number of levels (for a building), its surface area, etc. These attribute data have to be chosen in accordance with their subsequent use in damage cost estimation (§2.2.8.3 on the following page). From one CBA-AD study to one another, assets maps may vary in the presence, completeness and quality of attribute data.

A more complete discussion on the characterisation of flood-exposed assets is given by Merz et al. (2010). Figure 2.11 shows two different examples of assets maps. On the left is an extract of the assets map used for a CBA-AD on the ZEC case study (Gilbert and Ledoux 2012): assets are identified by points (economic activities) or polygonal features (private housing units, agricultural land) on a GIS vector layer built from Corine Land Cover land use map completed with field data; typical surface area of polygonal features is approximately 30 ha. On the right is an extract of the assets map used for a CBA-AD on the Orb Delta case study (SMVOL 2011): assets are all identified by polygonal features, of smaller spatial support—typical surface area of assets is approximately 100 sq. m. for private housing, 3 ha. for agricultural land and 0.1 ha. for other economic activities. The choice of a description of assets will of course influence the subsequent computation of flood damages: Eleuterio et al. (2008) discuss this issue by investigating a case study in Bas-Rhin, France.

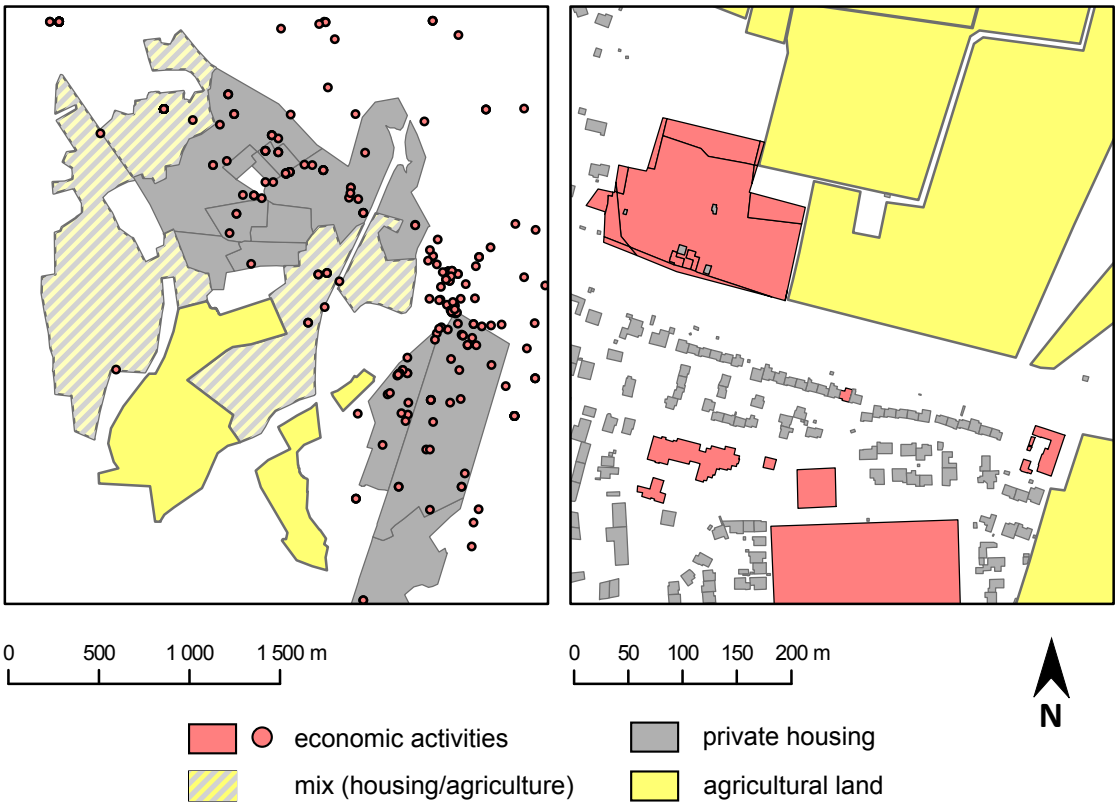

Figure 2.11: Two assets maps with different spatial support on two study sites: ZEC (left), Orb Delta (right)

Exposure assessment For a given flood scenario e_j (present or future situation), exposure is measured by confronting the assets map with flood hazard maps $\mathcal{H}(e_j)$. For each exposed asset (usually represented by a point or a polygonal feature in a GIS layer), the flood caracteristics (water depth, water velocity, flood duration) are extracted by an overlay analysis using GIS tools (Figure 2.12). Depending on the nature of the data used, various GIS techniques can be considered to perform this overlay analysis between the assets map and the hazard maps: extraction of mean or maximum values of flood characteristics over the asset, preliminary clipping of polygonal features to the flood extent, etc. All these techniques will result in the computation of flood intensity characteristics for each flood-exposed asset, but important variations may be observed from one technique to one another. However, in spite of this variety of techniques, none of the CBA-AD studies or guidelines we surveyed provide the reader with a detailed description of the overlay procedure that was used, or should be used, to assess flood exposure.

One contribution of this thesis is to investigate this weak point of CBA-AD studies. We supervised the master's thesis of Thibaud Langer (Langer 2011), who listed ten techniques that can be considered to perform the overlay analysis for flood exposure assessment, depending on: i) the nature of the assets (point, line or polygonal features in a GIS vector layer); and ii) the nature of the hazard maps (point vector data, polygonal vector data, raster data). These different overlay procedures were then applied to the same case study (Fourques Beaucaire case study, §2.3.3 on page 70): we found relative differences in exposure data that could rise up to 10%. The interested reader will find a brief discussion on this issue in appendix §D on page 212.

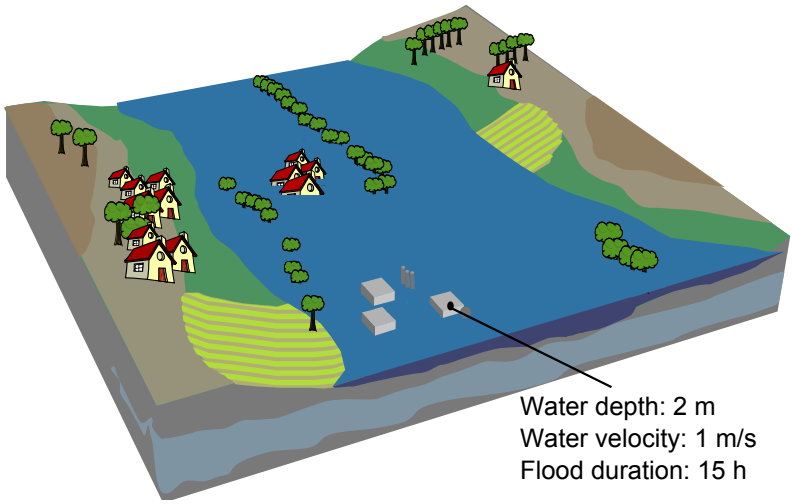

Figure 2.12: Flood-prone study area: flood exposure

2.2.8.3 Damage costs estimation

Damage costs estimation is the last step of flood damage assessment (Figure 2.8 on page 52): it consists of estimating damage costs for each flood scenario e_j and for each asset within the study area, from flood exposure assessment (Figure 2.14 on the facing page).[m] Generally speaking, damage estimates should include both direct and indirect, tangible and intangible damages—Merz et al. (2010) or DNRM (2002) explain these notions into details. Nevertheless, in most of the CBA-AD studies we surveyed, only direct and tangible monetary losses are taken into account because of lack of data and/or lack of suitable

[m] In the NOE modelling framework, flood damages are always assumed to be equal to zero for flood scenario e_1 (resp. e'_1) that is defined as the "*smallest*" flood event that induces damage in present (resp. future) situation (§2.2.6 on page 48).

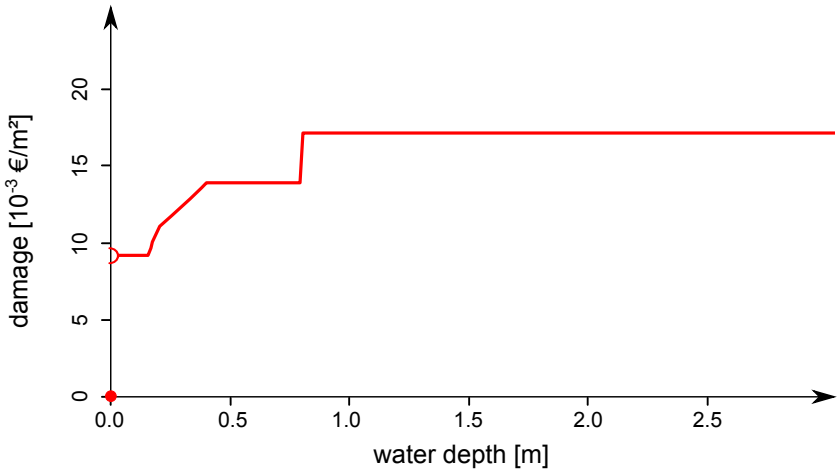

Figure 2.13: Absolute and surface-dependent depth-damage curve used for wheat crop on the Orb Delta case study
Source : Grelot et al. (2012)

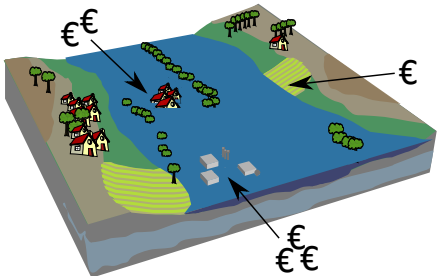

Figure 2.14: Flood-prone study area: damage costs estimation for a given flood scenario

methodology. So-called "*damage functions*" are then used to model the relation between flood intensity characteristics (water depth, velocity, flood duration), asset characteristics (type, surface area, ground floor elevation, value, etc.) and direct monetary losses (Figure 2.13). These damage functions are either built from i) statistical analysis of ex-post data collected after flood events (e.g., from insurance companies), or ii) by a synthetic approach where the elementary effects of flood on the various components of the asset are modelled (e.g., damage on furniture, walls and doors of a house). Damage functions can be:

- **absolute/relative:** absolute damage functions give absolute values of monetary losses in euros; relative damage functions give monetary losses as a percentage of the total value of the asset;
- **surface-dependent/surface-independent:** surface-dependent damage functions give monetary losses per surface area unit [€/m^2], while surface-independent damage functions give monetary losses for the whole asset regardless of its surface area.

The interested reader can refer to Merz et al. (2010) for more information on damage functions. Bournot (2008) also gives a list of references related to the main damage functions used in French CBA-AD studies. For a given flood scenario e_j, damages are estimated individually for each asset identified on the assets map. Hence, flood damages can be mapped to give a spatially explicit portray of flood risk on the study area. We will use the term "*damage map*" and the notation $\mathcal{D}(e_j)$ to refer to it. Damages can also be summed up over the study area to give total damage estimate $D(e_j)$. This total amount is then used in the computation of the ΔAAD indicator (§2.2.5 on page 44).

2.3 Case studies

The NOE modelling framework that we described in §2.2 has been applied to a number of different case studies. The purpose of this section is to: i) briefly explain how we implemented the NOE modelling framework into a convenient computer code (§2.3.1); and ii) present the study sites on which the NOE code has been applied (§2.3.2 and §2.3.3). Among these study sites, we will focus on the application of the NOE code to the Orb Delta (§2.3.2). This case study will be used as a real-world test case for variance-based global sensitivity analysis of spatial models throughout this document. The other case studies are shortly portrayed in §2.3.3.

2.3.1 Development of the NOE code

Most sensitivity analysis techniques require to run many simulations (hundreds or thousands) of the model under study. To carry out a sensitivity analysis of the NOE modelling framework, it was thus necessary to implement it into a convenient and efficient computer code. We will refer to this code as the "*NOE code*". This code development is not at the core of our research, hence we only present it very briefly in the following paragraphs.

Structure of the NOE code A first part of the NOE code consists of a set of tools that perform spatial analysis operations to produce flood exposure maps from the assets map and the hazard maps, as described in §2.2.8.2 on page 56. These tools were coded in `Python`™, an interpreted and object-oriented programming language[n]. They make use of the `ArcPy` library, which provides access to the geoprocessing functions available in `ArcGis`®, a commercial GIS software[o]. These tools were packaged into a single toolbox named `geonoe`, which can easily be loaded and used from a standard `ArcGis`® user session.

A second part of the NOE code consists of a set of scripts coded in the R open-source programming language[p]. These scripts code for: the computation of flood damages based on exposure data and damage functions, as described in §2.2.8.3 on page 56; the calculation of the average annual avoided damages (ΔAAD indicator) from a set of flood scenarios e_j and associated weights ω_i, as explained in §2.2.5 to §2.2.7 on pages 44–49; the computation of the Net Present Value of a flood risk management plan (§2.2.4 on page 42). These scripts were packaged into a single library named `noe`.

Note: it must be noted that no hydraulic model is included in the NOE code. The hazard maps $\mathcal{H}(e_j)$ associated to each flood scenario e_j (§2.2.8.1 on page 53) are considered as initial inputs of the NOE code.

Computing time The computing time associated with one run of the NOE code depends on the characteristics of the input data (spatial extent, resolution, etc.). On the Orb Delta case study, which will be used as the main test-case in our thesis, one single run costs around 30 seconds on a computer with average performances. This duration may be considered as "*short*" with respect to the requirements of most sensitivity analysis techniques. It allows lauching many simulations of the NOE code in any pseudo Monte Carlo approach: for example, running the NOE code a thousand times would last slightly more

[n] http://www.python.org
[o] http://www.esri.com/software/arcgis
[p] http://cran.r-project.org/

than height hours. However, in order to lower the CPU cost even more, we used the NOE code on a grid computing cluster: calculation times were divided by six.

2.3.2 The Orb Delta case study

In this section, we describe the application of the NOE code on the Orb Delta (Hérault, France). This case study will be used as a real-world test case for VB-GSA of spatial models throughout this document.

2.3.2.1 Study site

As a study area, we selected the lower Orb river fluvial plain, known as the Orb Delta, located in the south of France. We focused on a 15 km reach from Béziers to the Mediterranean sea that is bounded by an area of 63 sq. km and includes the cities of Béziers, Portiragnes, Sauvian, Sérignan, Valras-Plage and Villeneuve-lès-Béziers (Figure 2.15). The Orb catchment has a typical Mediterranean subhumid regime. The annual maximum discharge in Béziers (Tabarka gauge) varies from year to year between 100 and 1 500 m^3/s (BCEOM 2000). The flood prone area in the Orb Delta is home to approximately 16 290 permanent people (total population of the six localities: 90 000 people), 774 companies and 30 seaside campgrounds (which attract up to 100 000 tourists in summertime). Approximately one-third of the area is devoted to agriculture. The flood of December 1995 - January 1996, with a peak discharge of 1 700 m^3/s at the Tabarka gauge, caused a total amount of damage of 53 M€ (SMVOL 2011).

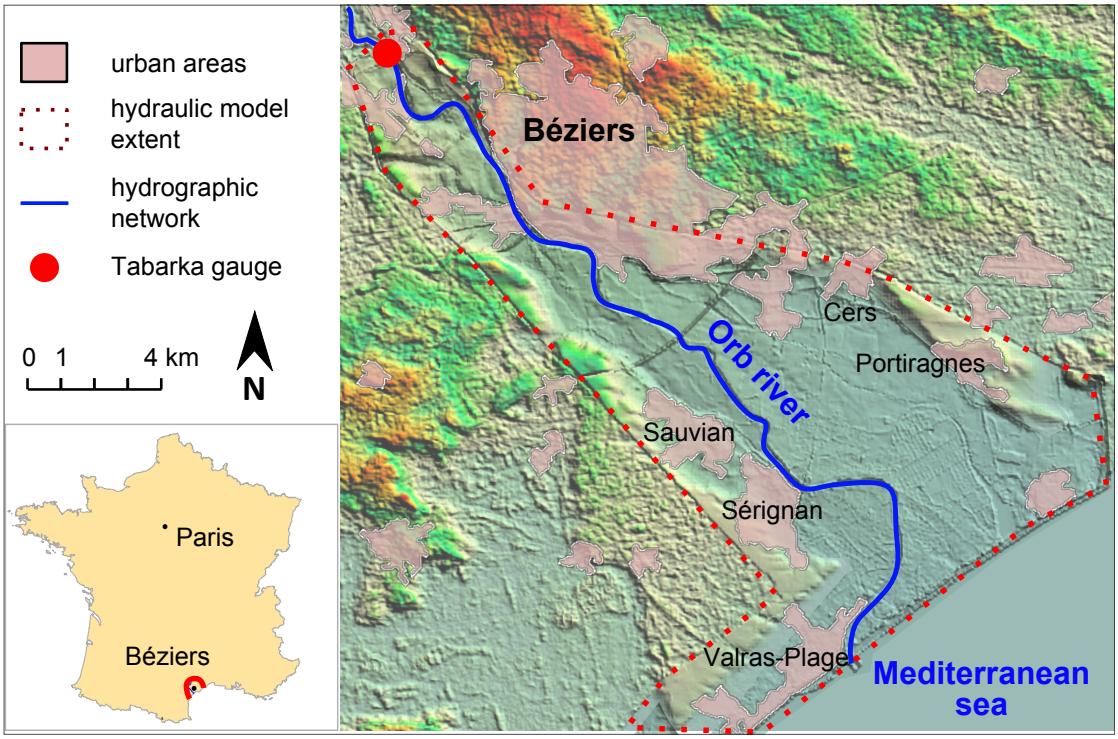

Figure 2.15: The Orb Delta study site is located in Hérault département, south of France. The Orb River flows southward.
Source : www.geoportail.fr

Flood risk management plan In 2001, local authorities launched a flood risk management plan on the Orb Delta, mainly based on various structural mitigation measures, including levee strengthening around urban areas, restoration of sea outfalls and channel improvement (Figure 2.16). In 2011, to claim national subsidies, they completed a cost-benefit analysis of their project (Grelot et al. 2012).

This study site was mainly chosen because it is a "*real*" case study, with a flood risk management plan under construction and a cost-benefit analysis produced by the local authorities. Moreover, the area was already documented with numerous available data. These data included aerial photographs, a high-resolution Digital Terrain Model (DTM) built from photogrammetry, the annual maximum flow series from 1967 to 2009 at the Tabarka gauge, and various spatial datasets on buildings, agricultural land and economic activities in the area (Erdlenbruch et al. 2008).

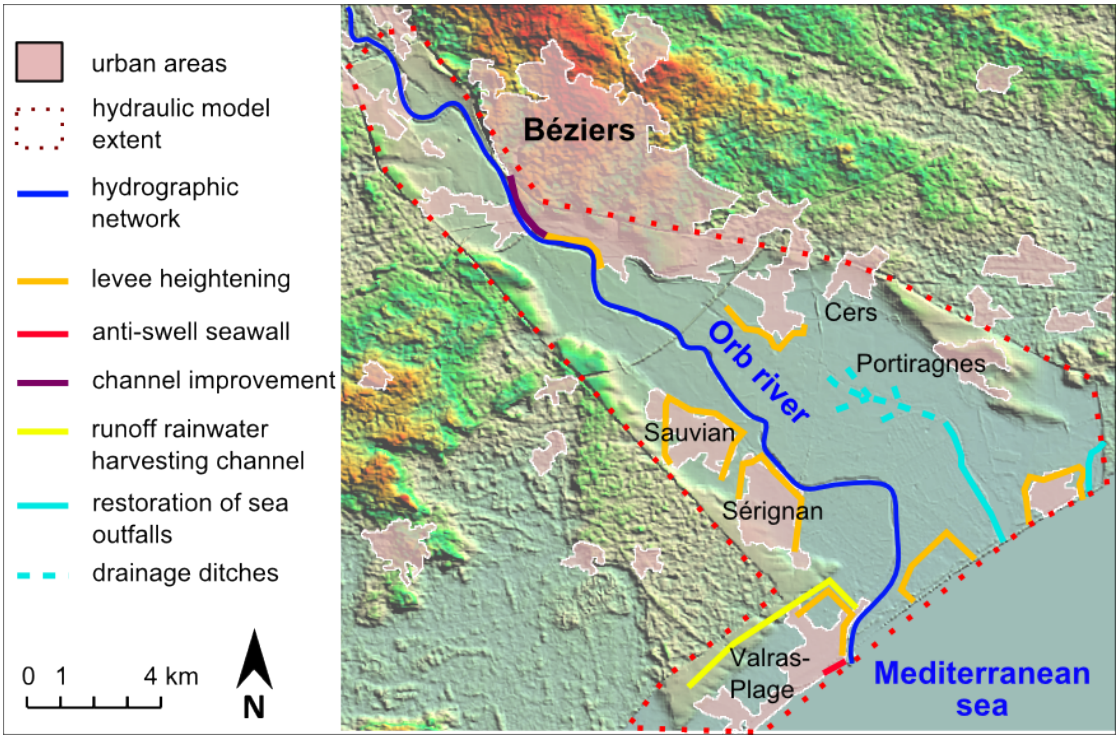

Figure 2.16: Orb Delta case study: structural flood-control measures
Source : Erdlenbruch et al. (2007)

2.3.2.2 Model flowchart

The application of the NOE code to the Orb Delta case study follows the general flowchart that we described in Figure 2.2 on page 43. We give here, when needed, more details on the data used and the realisation of each step of the flowchart.

Comparison of costs and benefits In the NOE modelling framework, Net Present Value (NPV) is chosen to compare costs and benefits of the flood risk management plan. On the Orb Delta case study, it is computed over a period of R = 30 years. Only investment and maintenance costs are considered (CI = 35.2M€, CM = 1.6M€/year), and project benefits only include avoided damages.

Choice of flood scenarios Six flood scenarios e_1 to e_6 were selected to describe flood hazard in the present situation, and other six similar scenarios e'_1 to e'_6 for the future situation. They are only characterised by their peak discharge $q(e_j)$ at the Tabarka gauging station (Table 2.1) and by an hydrogram. For each $j = 1, \ldots, 6$, present and future scenarios e_j and e'_j have the same values of peak discharge and same hydrograms. Besides, the state of hydraulic infrastructures is assumed to be the same for all flood scenarios: according to a previous hydraulic and civil engineering survey, flood-control levees on the Orb river can be overtopped in case of an extreme flooding event, but are never supposed to break (BCEOM 2000).

Weights of flood scenarios — flood frequency analysis In this case study, flood scenarios are only characterised by their peak discharge $q(e_j)$. Hence, their weigths $\omega(e_j)$ are completely characterised by a classical flood frequency analysis, based on the annual maximum flow series at Tabarka gauging station from 1967 to 2009 (Figure 2.17 on the following page). Estimated exceedance probabilities $F_Q(q_j)$, return intervals $T(q_j)$ and associated scenario weights $\omega(e_j)$ are given in Table 2.1[q].

Table 2.1: Orb Delta case study: flood scenarios

	Scenario description	Peak discharge q [m^3/s]	Exceedance frequency f	Return interval $T = 1/f$ [years]	Weight ω
e_1 or e'_1	Smallest flooding event that induces damage	1 018	0.2	5	0.05
e_2 or e'_2	10-year synthetic flood	1 287	0.1	10	0.08333
e_3 or e'_3	Historical flood (December, 1987)	1 696	0.0333	30	0.04
e_4 or e'_4	Historical flood (January, 1996)	1 882	0.02	50	0.01165
e_5 or e'_5	Large synthetic flood	2 133	0.01	100	0.0095
e_6 or e'_6	Probable maximum flood (over-topping dykes)	3 000	0.001	1 000	0.006

Flood hazard modelling The hydraulic model used for the Orb Delta case study is the 1D step-backwater model ISIS Flow. ISIS Flow computes flow levels and discharges using a method based on the Saint-Venant equations (ISIS 2012). The rough outputs produced by the hydraulic model are GIS vector layers that give water levels, water velocity and flood duration along the river stream and on a number of storage cells represented by polygonal features, with a typical surface area of 1 sq. km. Water depth maps were obtained from these rough outputs through a simple substraction of water levels with a high-resolution DTM of 5 m cell size. This DTM (raster data) was initially built from stereophotogrammetry. In the end, for each flood scenario e_j, the hazard maps $\mathcal{H}(e_j)$ consist of a set a three rasters of 5 m cell size giving respectively water depths, water velocity and flood duration over the study area (Figure 2.18 on page 63). It should be noted that we do not consider the ISIS Flow computer code as part of the NOE code; water depth maps are considered as inputs.

[q]The figures given in Table 2.1 are slightly different from the figures used in previous studies on the Orb Delta (Erdlenbruch et al. 2007; Grelot et al. 2012). The difference lies in the estimation of flood return intervals from the annual maximum flow series at Tabarka gauging station: the serie used in this document is longer than the one used in previous studies, and the fitted Q-f curve is different.

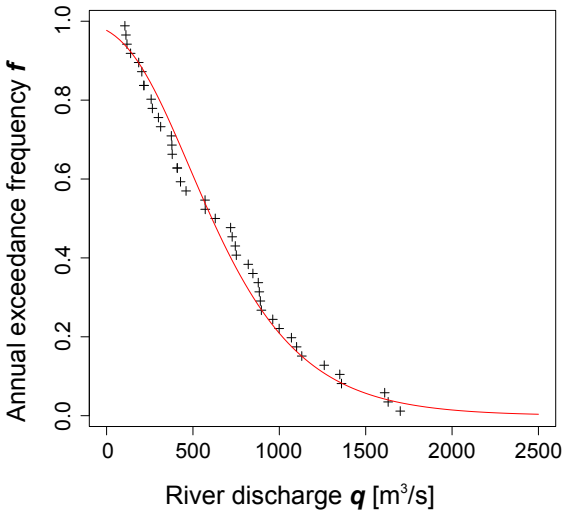

Figure 2.17: Orb Delta case study: observed **Q-f** data and fitted discharge-frequency curve

Flood exposure analysis In the application of the NOE code to Orb Delta case study, four economic sectors were considered for the exposure analysis: private housing, agricultural land, campgrounds and other economic activities (industry, shops, etc.). Flood exposure was assessed at the scale of small individual assets (buildings, plots of cultivated land, etc.). Data from various sources was collected to build the assets map: digital cadastral maps, a dataset of the regional Chamber of Commerce and Industry (2009), and the national agricultural land use statistics (RPG dataset, 2009). An extensive field survey was also conducted to collect additional data on assets, such as ground floor elevation of buildings. In the end, the assets map describes private housing units (individual buildings), plots of cultivated land, campgrounds and other economic activities by individual polygonal features in a single GIS vector layer (Table 2.2, Figure 2.19 on page 64). Plots of cultivated land were further characterised by a subtype (wheat, vineyard, etc.), while economic activities were classified into sixty categories following the French classification of economic activities NAF2008 (INSEE 2008).

Table 2.2: Orb Delta case study: content of the assets map

Type of assets	Data source	Number of objects	Total surface [sq. km]	Average surface [sq. m]
Private housing	Cadastral map + field survey	16 436	1.37	83
Agricultural land	National agricultural land use statistics (2009)	707	23.36	33 044
Campgrounds	Cadastral map + field survey	111	1.02	9 203
Other economic activities	Cadastral map + CCI dataset (2009)	691	0.62	904

Flood exposure of assets was then assessed by confronting the assets map with water depth maps $\mathcal{H}(e_1)$ to $\mathcal{H}(e_6)$ and $\mathcal{H}(e'_1)$ to $\mathcal{H}(e'_6)$. The water depth associated with each flood-exposed asset for each flood scenario e_j was calculated with the following overlay procedure:

1. in order to handle very large assets (e.g., large plots of cultivated land), we first divided all objects of the assets map into small pieces, by intersecting the assets map with a regular square grid of 200 m cell size. After this operation, all the polygonal objects of the assets map have a maximum

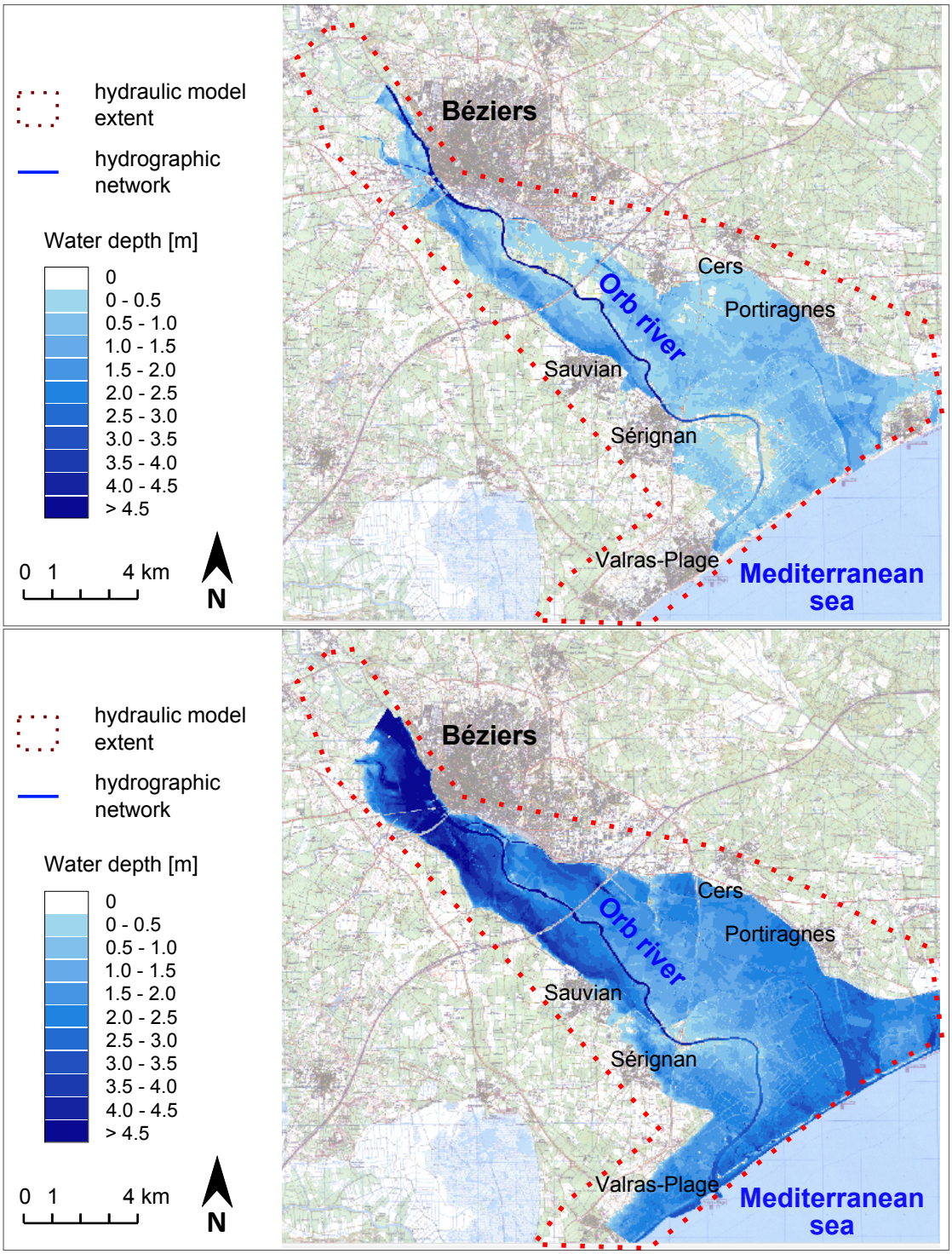

Figure 2.18: Orb Delta case study: hazard maps for scenario e_2' (10-years flood in future situation, top) and e_6 (very large flood in present situation, bottom)

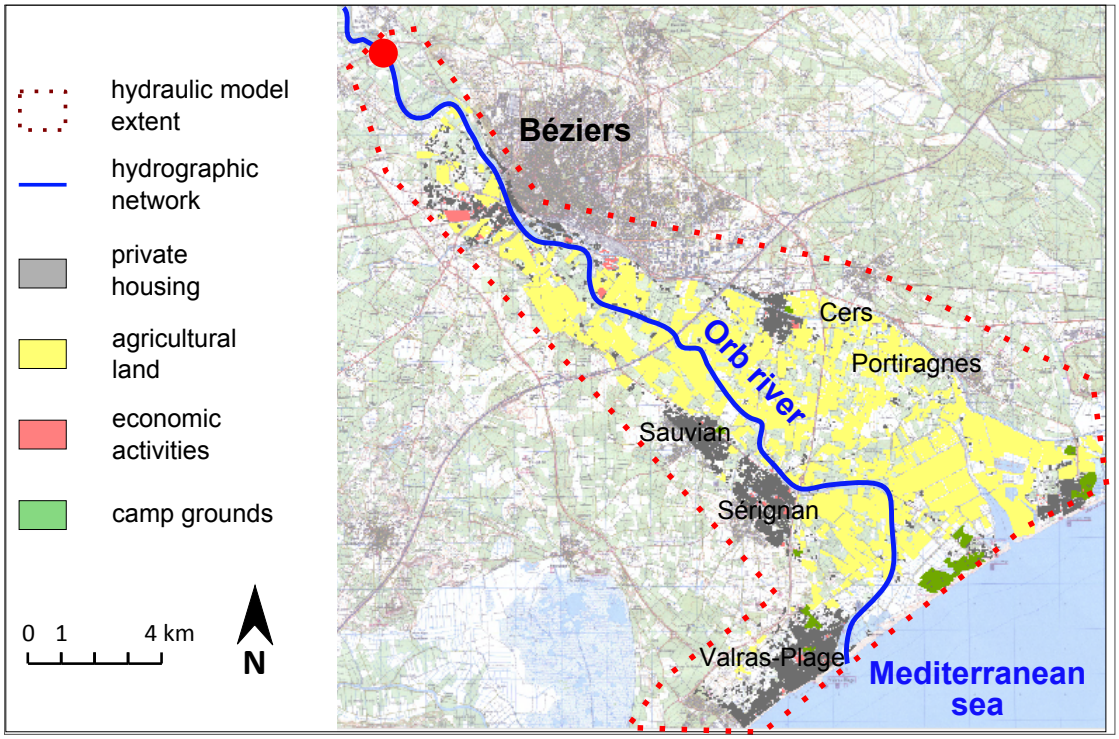

Figure 2.19: Orb Delta case study: assets map

surface area of 200 m by 200 m;

2. for each polygonal object of the assets map, the average water depth over the object was computed from the raster hazard map $\mathcal{H}(e_j)$ using the `Zonal Statistics` tool provided in `ArcGis`®;
3. the average water depth associated with each asset was then extracted as a new attribute data in the original assets map (vector layer).

Flood damage estimation For damage costs estimation, we used absolute damage functions (§2.2.8.3 on page 56) which depend on the following parameters: type and floor surface area of the exposed asset, average water depth. Damage functions for private housing, campground and cultivated land are surface-dependent, while the damage functions used for other economic activities are surface-independent. Flood velocity and flood duration were considered as homogeneous. These damage functions were taken from the recommendations of French Ministry of Ecology, Sustainable Development and Energy (MEDDTL 2011), for a complete description see the original study (Grelot et al. 2012). In the end, a total of 94 depth-damage relations were used, one for each land use type and subtype (Table 2.3 on the next page).

2.3.2.3 Model inputs

As mentioned in §2.2.1 on page 40, a preliminary step to carry out sensitivity analysis of the NOE code on the Orb Delta case study is to describe properly its inputs and its outputs. In order to cope with the large number of input data involved in the NOE code, we decided to group them into five groups that can be considered as independent (Table 2.4 on the facing page). We will somehow imprecisely use the term

Table 2.3: Orb Delta case study: damage functions

Type of assets	Type of approach	Sub-types number	Parameters
Private housing	Empirical (data collected after flood events)	1	water depth; floor surface area
Agricultural land	Synthetic approach (based on questionnaires)	15	water depth; surface area
Campgrounds	Synthetic approach (based on questionnaires)	18	water depth; surface area
Other economic activities	Synthetic approach	60	water depth

Table 2.4: Orb Delta case study: model inputs of the NOE code

Notation	Name	Nature	Details
U_1	Scenario weights	Group of scalar inputs	Set of scenario weights $\omega(e_1)$ to $\omega(e_6)$ and $\omega(e'_1)$ to $\omega(e'_6)$ for all flooding scenarios in present and future situations (Table 2.1 on page 61).
U_2	Damage functions	Group of scalar inputs	Group of parameters describing depth-damage functions for all types of assets.
U_3	Project costs	Group of scalar inputs	Investment costs CI and maintenance costs CE related to the flood risk management plan
Z_4	Assets map	GIS vector layer	Spatially explicit description of assets exposed to floods on the study area. Each asset is represented by a polygonal feature with the following attributes: type (qualitative); surface area [m^2]; groundfloor elevation [m].
Z_5	Hazard maps	Group of GIS raster data	Set of water depth maps $\mathcal{H}(e_1)$ to $\mathcal{H}(e_6)$ and $\mathcal{H}(e'_1)$ to $\mathcal{H}(e'_6)$ (5 m cell size) for all flooding scenarios in present and future situations.

"*model inputs*" to refer to these five groups. Their composition was chosen in accordance with the general flowchart of the NOE modelling framework[r]. Three groups (damage functions, project costs and scenario weights) are composed of scalar inputs. The two other groups include spatially distributed inputs (hazard maps, assets map).

2.3.2.4 Model outputs

As developed in §2.2, the main intermediate and final outputs of the NOE modelling framework are: i) the damage estimates $D(e_j)$ for each flood scenario e_j; ii) the Average Annual Damages for the present and future situation: AAD and AAD' indicators; iii) the ΔAAD indicator; and iv) the Net Present Value of the flood risk management plan under study. A key point is that, except for the NPV indicator, all these outputs can be aggregated on different spatial supports: individual assets (vector map), regular grid (raster map), or the entire floodplain (scalar).

Vector map of flood damages First, flood damage estimates are initially computed for each flood-exposed asset over the study area. Hence, they can be mapped as a GIS vector layer (Figure 2.20 on the next page).

[r] We will later discuss in the conclusion of Chapter 3 how the composition of these groups may influence the results of sensitivity analysis of the NOE code.

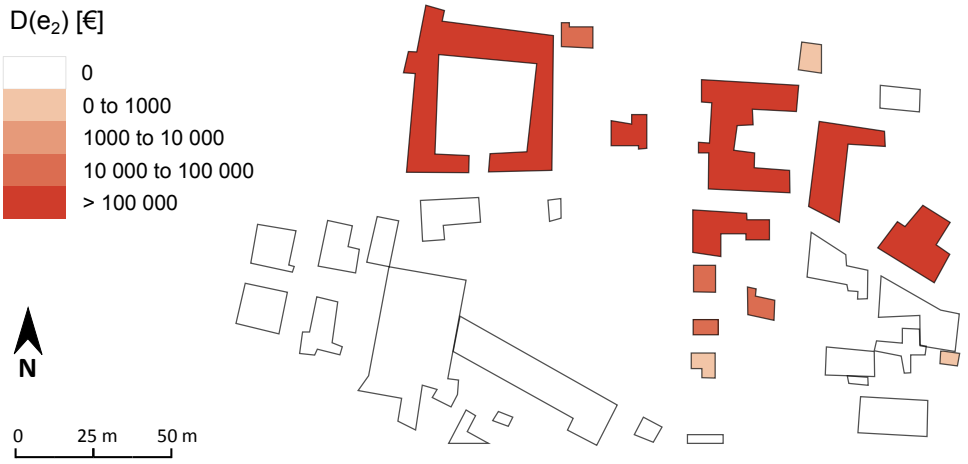

Figure 2.20: Zoom on the vector map of damages $\mathcal{D}(e_2)$ (10-year flood in present situation), GIS vector layer

Raster map of flood damages Next, flood damage estimates (and AAD, AAD' or ΔAAD indicators) can also be represented on a raster map of any cell size: in this case, the value given on a cell c_i of the grid is equal to the sum of the flood damage over all assets (or parts of assets) contained in the cell. If an asset has a large surface area and overlaps many cells of the grid, then the value of the output of interest over this asset is shared out among the cells in proportion to the overlaped areas (Figure 2.21). Here the "*spatial support*" of the model output is simply related to the cell size $|c|$ of the raster. We will in our research consider the following cell sizes: 200 m × 200 m, 400 m × 400 m, 800 m × 800 m and 1 600 m × 1 600 m, with corresponding surface areas $|c|$ = 0.04, 0.16, 0.64 and 2.56 sq. km, respectively. Figure 2.22 on the next page displays such raster maps for the AAD and AAD' indicators (fixed cell size $|c|$ = 0.04 sq. km). It shows the strong spatial variability of flood damages; in particular, urban areas and campgrounds display much larger amounts of damages than cells only covered with cultivated land. Similar raster maps for the ΔAAD indicator are given in Figure 2.23 on page 68 (cell sizes $|c|$ = 0.04 sq. km and $|c|$ = 2.56 sq. km). They clearly suggest that the flood risk management plan will not result in a spatially homogeneous reduction of average annual flood damages. Urban areas display positive values of the ΔAAD indicator, proving that they will benefit from the flood risk management plan. On the contrary, cells covered with cultivated land or campgrounds mostly display negative values of the ΔAAD indicator: in these areas, the flood risk management plan will unfortunately result in an increase of the average annual damages.

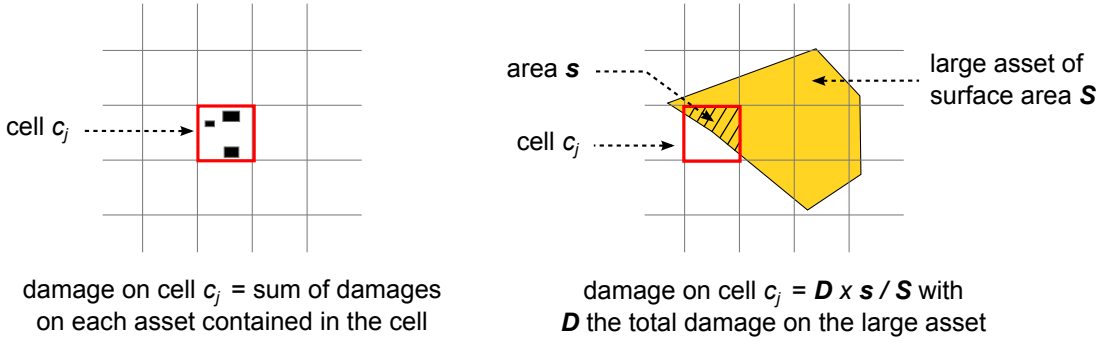

Figure 2.21: From vector map of damages to raster map of damages: computation of flood damages on a cell. Case of small assets (left) and case of a large asset that overlaps many cells of the grid (right)

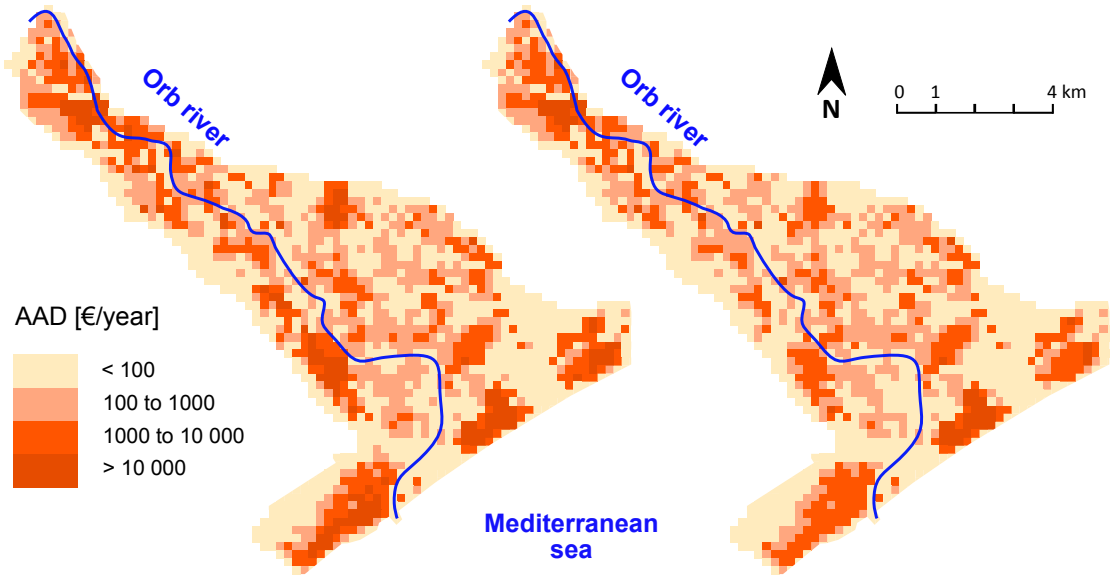

Figure 2.22: Raster maps of the AAD indicator in present (left) and future (right) situations, cell size $|c|$ = 0.04 sq. km

Table 2.5: Orb Delta case study: nominal output values of the AAD, AAD', ΔAAD and NPV indicators over the entire floodplain. Subtotals are also given for each economic sector considered.

	AAD [M€/year]	AAD' [M€/year]	ΔAAD [M€/year]	NPV [M€]
Private housing	2.678930	1.12995	1.54899000	-
Agricultural land	0.177004	0.18182	-0.00481636	-
Campgrounds	1.739160	1.66792	0.07124420	-
Other economic activities	7.634780	2.72724	4.90754000	-
Total	12.229874	5.70693	6.52295784	49.92795

Total damages over the entire floodplain Finally, the various outputs produced by the NOE code can also be aggregated over the entire floodplain to produce scalar outputs. Figure 2.24 on page 69 shows the nominal estimates of total flood damages $D(e_j)$ over the Orb Delta for each flood scenario e_j. Nominal output values of the AAD, AAD', ΔAAD and NPV indicators over the entire floodplain are given in Table 2.5.

To perform the sensitivity analysis of the NOE code on the Orb Delta case study, we will alternatively consider the model outputs over different spatial supports. In Chapter 3 we will pay attention to the model outputs aggregated over the entire floodplain. Next in Chapter 4 we will study raster maps of the ΔAAD indicator for increasing cell sizes $|c|$.

2.3.2.5 Model scales

Table 2.6 on page 69 displays the spatial scales at which the NOE code operate, for the Orb Delta case study. This description is based on the "*scale triplet*" discussed in §1.1.3 on page 20.

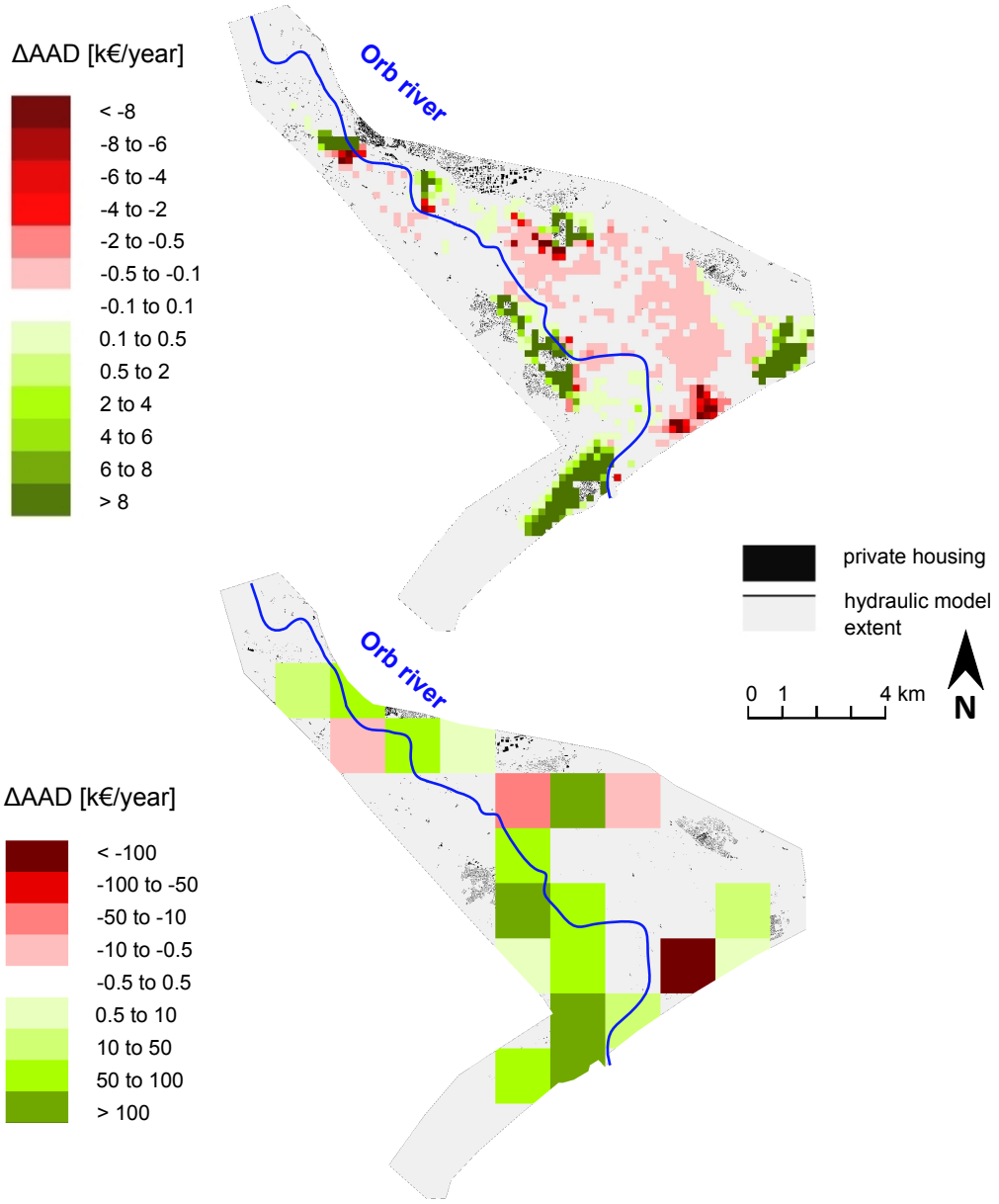

Figure 2.23: Raster maps of the ΔAAD indicator for nominal values of the model inputs, cell size $|c|$ = 0.04 sq. km (top) and $|c|$ = 2.56 sq. km (bottom)

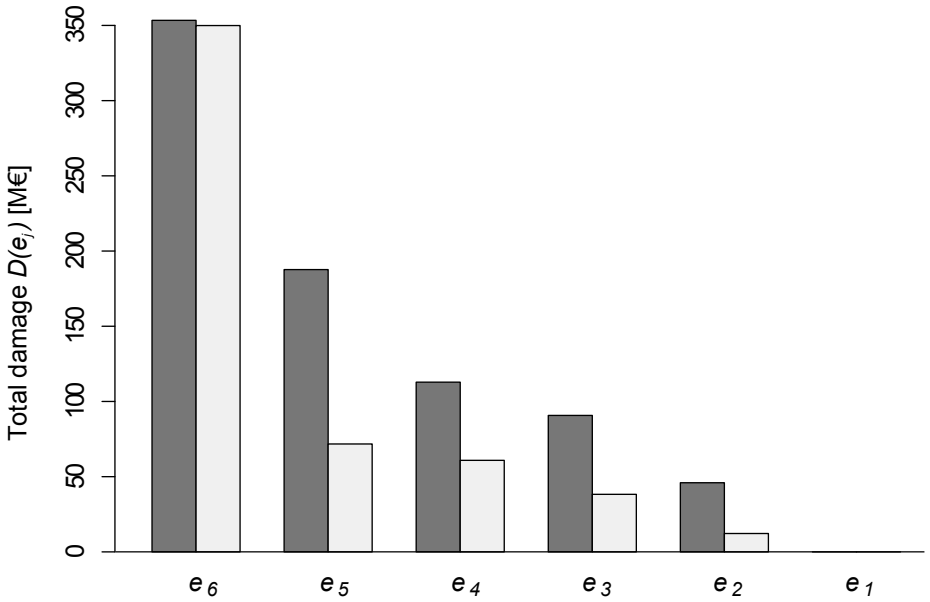

Figure 2.24: Orb Delta case study: nominal results for total flood damage estimates $D(e_j)$. Present situation (dark grey bars) ad future situation (light grey bars)

Table 2.6: Orb Delta case study: spatial scales in the NOE modelling chain

Data	Role	Format	Spatial extent	Support	Spacing
Project costs	input	scalar		non spatially distributed	
Scenario weights	input	scalar		non spatially distributed	
Flood hazard modelling					
hydraulic model outputs	input	vector	floodplain	storage cells ~ 1 sq. km.	–
DTM	input	raster	floodplain	point	5 m
water depth maps	input	raster	floodplain	cell (5 m)	–
Flood exposure modelling					
Assets map	input	vector	floodplain	smallest resolvable units ~ 100 sq. m.	–
Flood damage estimation: output indicators $D(e_j)$, AAD, AAD' and ΔAAD on different spatial supports					
Vector map of damages	output	vector	floodplain	smallest resolvable units ~ 100 sq. m.	–
Raster maps of damages	output	raster	floodplain	cells of increasing sizes: $\|c\| = 0.04, 0.16, 0.64$ and 2.56 sq. km	–
Aggregated damages over the floodplain	output	scalar		non spatially distributed	
Net Present Value	output	scalar		non spatially distributed	

2.3.3 Other case studies

Apart from the Orb Delta, the NOE modelling framework was also applied to another three case studies. We only display here the main characteristics of these extra studies, which are summarized in Table 2.7 on the next page. A more complete description can be found in Appendix §E on page 215. We will use these extra case studies in various parts of this thesis to make comparative analyses with the results obtained on the Orb Delta.

The Vilaine floodplain The first extra case study is the Vilaine floodplain (Île-et-Vilaine, France). This study area lies around the city of Redon, and has a small extent of only 17 sq. km. This case study was investigated as part of the RDT research program (Grelot 2009).

The Rhône river from Fourques to Beaucaire The second extra case study investigates the economic relevance of a project of dike strengthening and heightening on the Rhône river reach between the cities of Fourques and Beaucaire (ISL 2011). The total extent of the study area is 125 sq. km. We worked on this case study as part of a larger expert mission on uncertainties in cost-benefit analysis of flood risk management plans, funded by the Plan Rhône[s].

The ZEC case study The third extra case study looks into a larger flood risk management plan, which we will refer to as the "*ZEC project*", that aims at renovating old floodplains along the lower reaches of the Rhône river (Gilbert and Ledoux 2012). The total extent of the study area is 650 sq. km. We also worked on this case study as part of a larger expert mission for the Plan Rhône.

The main characteristics of each case study are summarized in Table 2.7 on the next page.

[s] http://www.planrhone.fr

Table 2.7: Main characteristics of the extra case studies

	Vilaine floodplain	**Fourques - Beaucaire**	**ZEC**
Extent	17 sq. km	125 sq. km	650 sq. km
Flood scenarios			
Number of scenarios	6	14	12
Descriptors of scenarios	peak discharge q (κ = 1)	peak discharge q, levee failure/not failure, position of levee failure, season (κ = 4)	peak discharge, season (κ = 2)
Computation of scenario weights	hydrological frequency analysis	hydrological frequency analysis + seasonal analysis + probabilistic levee reliability assessment	hydrological frequency analysis + seasonal analysis
Approximation of AAD indicator	trapezoidal rule	Riemann sum (rectangle method)	trapezoidal rule
Flood hazard modelling			
Type of hydraulic model	1D model (InfoWorks)	2D model (RUBAR 20)	hybrid 1D model with storage cells
Flood exposure modelling			
Types of assets	private housing (points), economic activities (points)	private housing (points), economic activities (points), agricultural land (polygons), farm buidlings (points)	private housing (polygons), agricultural land (polygons), economic activities (points)
Number of assets	~ 500	~ 5 000	~ 10 000
Flood damage functions			
Intensity parameters	water depth	water depth, velocity (for agricultural assets only) and flood duration (idem)	water depth, velocity (for agricultural assets only) and flood duration (idem)

2.4 Chapter conclusion

The goal of this chapter was to present the NOE modelling framework for cost-benefit analysis of flood risk management plans based on the "*avoided damages*" approach (CBA-AD approach), and to describe its application to the Orb Delta case study. To conclude this chapter, we will first display a brief summary of our contributions, then enumerate the main characteristics of the NOE modelling framework, and finally stress its key limits.

2.4.1 A brief summary

We started in §2.1 with some elements of context on flood risk and economic appraisal of flood management policies, and surveyed the literature on CBA-AD studies, mostly in Europe and more particularly in France. Next, we explained our need for a general modelling framework that could describe the CBA-AD studies: such a modelling framework is necessary to clarify the boundaries, the inputs and the outputs of these studies; it is an essential preliminary step in order to carry out a proper sensitivity analysis of the CBA-AD approach, which is the applied objective of this thesis.

We gave in §2.2 a detailed description of the steps composing the NOE modelling framework, which involves hydraulic, hydrological and economic modelling as well as spatial analysis using GIS tools. These steps are: i) choice of flood scenarios; ii) estimation of scenario weights; iii) flood hazard modelling; iv) flood exposure modelling; v) damage costs estimation; vi) computation of the Average Annual Avoided Damages; and vii) comparison of the costs and benefits of the plan.

In the NOE modelling framework, we brought two original contributions compared to the existing literature on cost-benefit analysis applied to flood risk management plans. The first contribution is an attempt to extend the definition of the average annual damages, by modelling flooding events as random vectors rather than just real-valued random variables: this attempt proved useful to better identify and discuss the assumptions and limitations that are hidden when average annual damages are used as a risk indicator. Our second contribution is the investigation of the various spatial overlay techniques that can be used to compute the flood exposure map from the assets map and the hazard maps.

Finally, we briefly explained in §2.3 how we developed a computer code to implement the NOE modelling framework, and how we applied this NOE code to different case studies. The NOE code may be released for public use in the near future. We focused on the Orb Delta case study, which we will use as a real-world test case for VB-GSA of spatial models throughout this thesis.

2.4.2 Main features of the NOE modelling framework

We give here a list of the main features of the NOE modelling framework, especially of those characteristics that matter most to carry out an uncertainty and sensitivity analysis. We use the terms and notions that were defined in Chapter 1.

A deterministic model The NOE modelling framework is deterministic. Even if some of the quantities of interest are modelled as random variables (e.g., the flooding events e and their associated peak discharge q), the way this randomness is treated is totally deterministic. In particular, the definition of the average annual damages (ΔAAD indicator, §2.2.5 on page 44) and its approximation from a set of fixed flood scenarios e_j is a way to reduce the aleatory uncertainty related to flooding events and to summarize

it into a single deterministic indicator. We have briefly mentioned in §2.2.6 on page 49 the possibility to estimate the AAD indicator from a plausible random chronicle of flooding events over a very large length of time. In that case, the computer code that calculates the AAD indicator would be stochastic. This particular situation will not be considered in our research.

A spatial model The NOE modelling framework has both spatially distributed inputs and outputs. Spatially distributed inputs include the assets map and the hazard maps, which are usually vector and raster GIS data, respectively. These input data are expected to show some sort of spatial auto-correlation—the error on water depth at site $\mathbf{x}$ is necessarily correlated to the error on water depth at a neigbouring site $\mathbf{x}'$. Besides, most output quantities of interest (damage per flood scenario $D(e_j)$, AAD, AAD', and ΔAAD indicators) can also be spatially distributed (vector or raster data), as it was developed in §2.3.2.4 on page 65.

A spatially additive model with different supports for the model output The NOE modelling chain is a "*spatially additive model*" as defined in §1.1.4 on page 22: the aggregation of an output damage indicator (flood damage estimates $D(e_j)$, AAD, AAD', or ΔAAD indicators) over a given spatial support v is simply the sum of this damage indicator over all flood-exposed assets included in the support. As mentioned in §2.3.2.4 on page 65, we will consider in our analysis various spatial supports for the NOE model outputs: first the floodplain Ω as a whole, but also a number of raster maps with different cell sizes.

A point-based model The NOE modelling chain is also a "*point-based*" model, according to the definition given in §1.1.5 on page 23. Indeed, flood damage estimates at a given point $\mathbf{x} \in \Omega$ only depend on: i) flood intensity parameters at the same point $\mathbf{x} \in \Omega$; ii) the asset description at the same point $\mathbf{x} \in \Omega$; and iii) damage functions. It can be noted that some authors developed more complex frameworks, in which flood damage assessments are not point-based because they consider flood-exposed assets as spatially connected systems and take into account a number of induced damages; in such frameworks, flood damage on an asset located at point $\mathbf{x} \in \Omega$ (e.g., a factory) may depend on flood intensity parameters at other locations $\mathbf{x}'$ (e.g., water depth on a distant warehouse). See Brémond (2011) for an example on flood damages to farms.

A code with low CPU cost As detailed in §2.3.1 on page 58, one simulation of the NOE code on the Orb Delta case study costs around 4 seconds on a grid computing cluster. Hence, it is possible to run the code more than a thousand times, and to use intensive sampling techniques for sensitivity analysis.

2.4.3 Key limitations of the NOE modelling framework

As mentioned in §2.2.1 on page 40, the NOE modelling framework does not pretend to cover the entire variety and complexity of cost-benefit analyses applied to flood risk management plans. In particular, we must highlight the following limitations.

Hydraulic model The hydraulic model, which is used to produce hazard maps associated with flood scenarios, is not considered as a part of the NOE modelling framework. It is an external module, whose outputs (the hazard maps) are the inputs of the NOE code. Hence, the uncertainty and sensitivity analysis we will perform on the NOE code in the following chapters will not look into the hydraulic model uncertainties, but will only try to describe the resulting uncertainty on the hazard maps.

Evolution of landuse over time An important hypothesis in our framework is that all the input data are assumed to keep a fixed value over the time interval R over which the flood risk management plan is evaluated. In particular, the assets map is not supposed to change over the next 30 to 50 years. This strong assumption may seem indefensible, yet it is a common hypothesis in all cost-benefit analyses applied to flod risk management plans. We will discuss later how this hypothesis might be relaxed in further research.

Fluvial floods As mentioned in §2.2.5.1 on page 45, we designed the NOE modelling framework to deal with fluvial floods. However, this framework could be easily adapted to other types of inundation, such as groundwater floods, coastal floods, surface water floods, etc. It would just be necessary to add new descriptors for flooding events (e.g., the sea level for a coastal flood), and to compute scenario weights that would take these new descriptors into account.

Chapter 3

Spatially distributed inputs in variance-based global sensitivity analysis

As mentioned in Chapter 1, VB-GSA is initially designed for models with independent and scalar random inputs only: this is a first obstacle that limits its extension to spatial models. Indeed, in a spatial model, some model inputs are not scalar values but 2D raster or vector data, and may exhibit some sort of spatial auto-correlation; the original framework of VB-GSA does not fit any longer in this case. In this chapter, we explore this issue and investigate how variance-based sensitivity indices can be calculated in numerical models with spatially distributed inputs. Only models with scalar outputs are considered here (spatially distributed outputs will be studied in Chapter 4).

This chapter is composed of three sections. The first section §3.1 is a review of the methods encountered in the literature for VB-GSA of models with spatially distributed inputs. It is completed by a numerical study of some of these methods and by a discussion on their pros and cons. In the second section §3.2, we offer to discuss the issue of sampling of spatially distributed inputs in VB-GSA. We start with a brief overview of the main strategies to model uncertainty in a spatially distributed input and sample random realisations of it. We then explore the impact of sampling size and technique on the estimation of variance-based sensivity indices, with two numerical studies of analytical test cases, one of which was published in the *Proceedings of the ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (Saint-Geours et al. 2010). Next, the third section §3.3 is dedicated to an application of VB-GSA with spatial inputs to the NOE code on the Orb Delta case study. It was submitted in July 2012 to the *Journal of Flood Risk Management*. It starts with a review on uncertainty treatment in flood damage assessment studies. Finally, in the chapter conclusion (§2.4), we give some insights on i) how to extend the methods discussed in this chapter to other types of model inputs; and ii) how to extend the results observed on the NOE code for the Orb Delta case study to other CBA-AD studies.

3.1 State of the art and comparison of the available methods 79
3.1.1 Introduction 79
3.1.2 A review 79
3.1.2.1 Macro-parameter 79
3.1.2.2 Dimension reduction 81
3.1.2.3 Switch or trigger input 83
3.1.2.4 Map labelling 84
3.1.2.5 Joint meta-models 86
3.1.2.6 Second level uncertainty modelling 86
3.1.3 Numerical study 88
3.1.3.1 Analytical test cases 88
3.1.3.2 Methods under study 90
3.1.3.3 Results 90
3.1.4 Discussion 92
3.1.4.1 Methods do not produce the same information 92
3.1.4.2 Computational cost 94
3.1.4.3 Spatial structure of uncertainty 94
3.1.4.4 Coupling with a meta-model 95
3.1.4.5 Decision tree to choose the appropriate method 95
3.2 Sampling issues 98
3.2.1 Introduction 98
3.2.2 Modelling uncertainty on spatially distributed inputs: a review 98
3.2.2.1 Uncertainty models for raster data 99
3.2.2.2 Uncertainty models for vector data 100
3.2.3 Convergence of sensitivity indices estimates with map labelling method 100
3.2.3.1 Introduction 100
3.2.3.2 Method 100
3.2.3.3 Results 101
3.2.4 Comparing SRS and LHS to generate random realisations of $Z(\mathbf{x})$ 101
3.2.4.1 Introduction 103
3.2.4.2 A simple spatial model 103
3.2.4.3 Spatial global sensitivity analysis: method 104
3.2.4.4 Results and discussion 107
3.2.4.5 Conclusion 109
3.2.5 Discussion 110
3.3 Application to the NOE code on the Orb Delta case study 111
3.3.1 Introduction 111
3.3.1.1 Study site 113
3.3.2 Description of the NOE modelling framework 114
3.3.2.1 Flood scenarios 114
3.3.2.2 Flood frequency analysis 115
3.3.2.3 Flood hazard modelling 115
3.3.2.4 Flood exposure modelling 115
3.3.2.5 Damage estimation 115

3.3.2.6 Average Annual Avoided Damages 116
3.3.2.7 Net Present Value of the mitigation measures 116
3.3.3 Uncertainty and sensitivity analysis 117
3.3.3.1 Modelling sources of uncertainty 117
3.3.3.2 Propagating uncertainty 124
3.3.3.3 Variance-based sensitivity indices 124
3.3.4 Results 125
3.3.4.1 Uncertainty analysis 125
3.3.4.2 Sensitivity analysis 125
3.3.5 Discussion 128
3.3.5.1 Improving the NOE modelling framework 128
3.3.5.2 Averaging-out effects 129
3.3.5.3 Limits 130
3.3.6 Conclusion 130
3.3.7 Extra comments: results on other case studies 131
3.4 Chapter conclusion 133
3.4.1 On VB-GSA with spatially distributed inputs 133
3.4.2 On the NOE modelling framework and CBA-AD studies 134

3.1 State of the art and comparison of the available methods

3.1.1 Introduction

In its initial form summarized in §1.2 on page 24, variance-based global sensitivity analysis (VB-GSA) is designed to deal with scalar inputs U_i only, which are modelled as independent random variables. However, some recent research aim to extend the definition of variance-based sensitivity indices to correlated inputs (Li et al. 2010; Mara and Tarantola 2012; Kucherenko et al. 2012). Besides, a number of authors have recently introduced methods to compute variance-based sensitivity indices associated to one or several spatially distributed inputs $Z_i(\mathbf{x})$, that possibly exhibit spatial auto-correlation. Lilburne and Tarantola (2009), Iooss and Ribatet (2009), or Iooss (2008) make a partial state of the art of these approaches. In this section, we offer to:

- make an updated state of the art of these methods (§3.1.2);
- compare these methods on some analytical test cases (§3.1.3 on page 88) ;
- discuss their pros and cons and give practice-oriented criteria to choose the appropriate method for a given problem (§3.1.4 on page 92).

We consider in this whole section a numerical model $\mathcal{F}$ with a single spatially distributed input $Z(\mathbf{x})$ and a scalar output Y, as it was described in §1.1.1 on page 18:

$$Y = \mathcal{F}(U_1, \ldots, U_k, Z) \tag{3.1}$$

Besides, in some parts of this section, we will assume that the uncertainty on the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ can be simulated by a stochastic process, which will be denoted by $\mathcal{P}$ (e.g., a geostatistical algorithm). This stochastic process $\mathcal{P}$ may account for spatial auto-correlation or any complex structure of variability in $Z(\mathbf{x})$. This point will be further discussed in §3.2.2 on page 98.

3.1.2 A review

We found in the SA literature a number of papers dealing with VB-GSA applied to models with spatially distributed inputs. They display various methods that intend to define variance-based sensitivity indices that could measure the influence of an uncertain spatially distributed input $Z(\mathbf{x})$ on the variance of model output Y. These methods can be classified into six categories, which we describe into details in the following paragraphs §3.1.2.1 to §3.1.2.6 on pages 79–86.

3.1.2.1 Macro-parameter

A first approach (Figure 3.1 on the next page) is to consider a spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, or more precisely its numerical representation, as a finite set of d scalar parameters $\lambda_1 \ldots \lambda_d$. When $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is stored as a GIS raster, scalar parameters $(\lambda_j)_{j \in [\![1;d]\!]}$ are simply the values $Z(c_1), \ldots, Z(c_d)$ at the centers of cells c_i and d is the total number of cells. When $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is represented by a GIS vector layer, scalar parameters λ_j are the values of the attributes of each spatial object in the layer, and d is a multiple of the number of objects. Following Iooss and Ribatet (2009), we will use the term "*macro-parameter*" to refer to this method.

In some applications, it may possible to consider that the λ_i are independent random variables. Spatial correlation of uncertainty in model input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is then neglected. In other applications, random variables $(\lambda_j)_{j\in[\![1;d]\!]}$ are not statistically independent, because there is spatial auto-correlation in $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. In the first case (parameters λ_i independent), variance-based sensitivity indices can be computed either for each scalar parameter λ_j, or for the whole input group $(\lambda_j)_{j\in[\![1;d]\!]}$ (see §1.2.2.4 on page 32 for a discussion on grouping in VB-GSA). We will refer to these two options as "*macro-parameter without grouping*" and "*macro-parameter with grouping*". When scalar parameters $(\lambda_j)_{j\in[\![1;d]\!]}$ are correlated, only the "*with grouping*" option is appropriate. The steps of these two options are as follows:

Macro-parameter without grouping

Step 1 represent spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ as a finite set of d scalar parameters $\lambda_1 \dots \lambda_d$;

Step 2 identify a pdf p_{λ_i} for each scalar input λ_i (assumed to be independent);

Step 3 consider the new code $\mathcal{F}^\star$ (modified version of $\mathcal{F}$):

$$Y = \mathcal{F}^\star(U_1, \dots, U_k, \lambda_1, \dots, \lambda_d) \tag{3.2}$$

Step 4 create an input sample M_{tot} with inputs $(\lambda_j)_{j\in[\![1;d]\!]}$ along with other scalar model inputs U_i (following the pseudo-Monte Carlo procedure described in §1.2.2.5 on page 32) and evaluate code $\mathcal{F}^\star$ for each line of the sample;

Step 4 estimate first and total-order sensitivity indices for each scalar parameter λ_j (§1.2.2.5 on page 32).

Macro-parameter with grouping

Step 1 represent spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ as a finite set of d scalar parameters $\lambda_1 \dots \lambda_d$;

Step 2 identify a joint pdf p_λ for the set of parameters $(\lambda_j)_{j\in[\![1;d]\!]}$; if the parameters $(\lambda_j)_{j\in[\![1;d]\!]}$ are assumed to be independent, joint pdf p_λ is simply the product of the marginal pdfs p_{λ_i}; if not, the joint pdf p_λ will describe the correlated structure of these parameters;

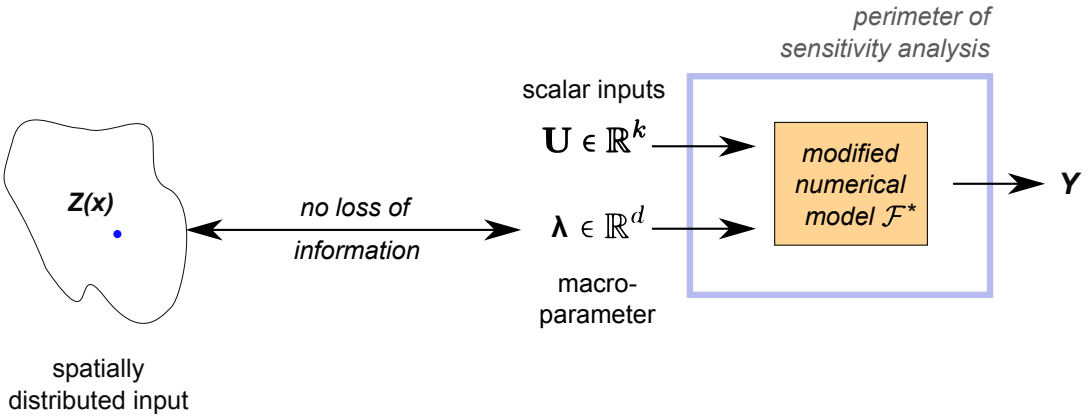

Figure 3.1: Macro-parameter method

Step 3 consider the new code $\mathcal{F}^\star$ (modified version of $\mathcal{F}$):

$$Y = \mathcal{F}^\star(U_1, \ldots, U_k, \lambda_1, \ldots, \lambda_d) \tag{3.3}$$

Step 4 create an input sample M_{tot} with the group of inputs $(\lambda_j)_{j\in[\![1;d]\!]}$ along with other scalar model inputs U_i (following the pseudo-Monte Carlo procedure described in §1.2.2.5 on page 32) and evaluate code $\mathcal{F}^\star$ for each line of the sample;

Step 5 compute the first and total-order sensitivity indices of the whole group of inputs $(\lambda_j)_{j\in[\![1;d]\!]}$ (§1.2.2.5 on page 32).

The "*macro-parameter with grouping*" approach was recently used by Heuvelink et al. (2010b) to carry out a sensitivity analysis of the GeoPEARL pesticide leaching model. In this work, each spatially distributed soil property (soil horizon thickness, clay or organic matter content, etc.) was represented on a spatial grid by a set of 258 scalar parameters λ_i that were assumed to be statistically independent. Hence, spatial auto-correlation of uncertainty was neglected. For each soil property, sensitivity indices were estimated for the whole group $(\lambda_j)_{j\in[\![1;d]\!]}$ by a pseudo Monte Carlo procedure.

The main drawback of the macro-parameter approach is its computational burden. The number of scalar parameters that must be sampled to estimate sensitivity indices is equal to $k + d$ where k is the number of non-spatially distributed model inputs $U_1, \ldots, U_k$ and d is the "*dimension*" of spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$—number of cells in a grid or number of spatial objects in a GIS vector layer. As soon as the dimension d gets too large ($d \gg 100$), the large samples required to estimate sensitivity indices may be difficult to handle.

3.1.2.2 Dimension reduction

The "*dimension reduction*" approach is similar to the "*macro-parameter*" method, except that it results in a reduction of the total information contained in the initial spatially distributed data $Z(\mathbf{x})$ (Figure 3.2 on the next page). Its idea is to find some way to approximate a spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ by a deterministic function ϕ of a small number of scalar inputs $\lambda_1, \ldots, \lambda_d$, with d small (typically $d \leq 100$, much less than the dimension of the initial spatially distributed data). As in the "*macro-parameter*" method, scalar parameters $\lambda_1, \ldots, \lambda_d$ may be assumed statistically independent or not, and may be grouped for sensitivity analysis or not (see §1.2.2.4 on page 32 for a discussion on grouping in VB-GSA). The steps of these two options are as follows:

Dimension reduction without grouping

Step 1 approximate spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ by a deterministic function ϕ of a small number of scalar inputs $\lambda_1, \ldots, \lambda_d$:

$$Z(\mathbf{x}) \approx \phi(\lambda_1, \ldots, \lambda_d, \mathbf{x}) \tag{3.4}$$

Step 2 identify a pdf p_{λ_i} for each scalar input λ_i (assumed to be independent);

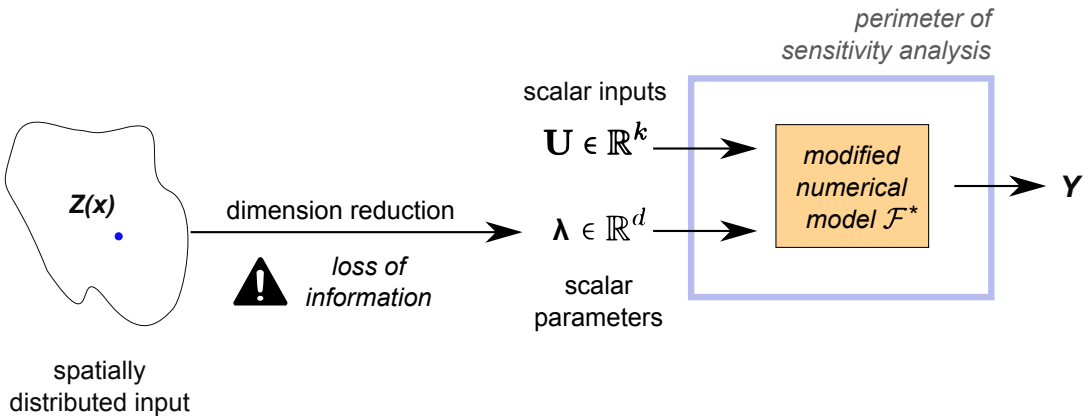

Figure 3.2: Dimension reduction method

Step 3 consider the new code $\mathcal{F}^\star$ (modified version of $\mathcal{F}$):

$$Y = \mathcal{F}^\star(U_1, \ldots, U_k, \lambda_1, \ldots, \lambda_d) \tag{3.5}$$

Step 4 create an input sample M_{tot} of inputs $(\lambda_j)_{j\in[\![1;d]\!]}$ along with other scalar model inputs U_i (following the pseudo-Monte Carlo procedure described in §1.2.2.5 on page 32) and evaluate code $\mathcal{F}^\star$ for each line of the sample;

Step 4 estimate first and total-order sensitivity indices for each scalar parameter λ_j (§1.2.2.5 on page 32).

Dimension reduction with grouping

Step 1 approximate spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ by a deterministic function ϕ of a small number of scalar inputs $\lambda_1, \ldots, \lambda_d$:

$$Z(\mathbf{x}) \approx \phi(\lambda_1, \ldots, \lambda_d, \mathbf{x}) \tag{3.6}$$

Step 2 identify a joint pdf p_λ for the set of parameters $(\lambda_j)_{j\in[\![1;d]\!]}$; if the parameters $(\lambda_j)_{j\in[\![1;d]\!]}$ are assumed to be independent, joint pdf p_λ is simply the product of the marginal pdfs p_{λ_i}; if not, the joint pdf p_λ will describe the correlated structure of these parameters;

Step 3 consider the new code $\mathcal{F}^\star$ (modified version of $\mathcal{F}$):

$$Y = \mathcal{F}^\star(U_1, \ldots, U_k, \lambda_1, \ldots, \lambda_d) \tag{3.7}$$

Step 4 create an input sample M_{tot} with the group of inputs $(\lambda_j)_{j\in[\![1;d]\!]}$ along with other scalar model inputs U_i (following the pseudo-Monte Carlo procedure described in §1.2.2.5 on page 32) and evaluate code $\mathcal{F}^\star$ for each line of the sample;

Step 5 compute the first and total-order sensitivity indices of the whole group of inputs $(\lambda_j)_{j\in[\![1;d]\!]}$ (§1.2.2.5 on page 32).

Here are some examples of works based on this approach. Volkova et al. (2008) carried out sensitivity analysis of a model of groundwater transport for radionuclide migration on a radwaste disposal site. To represent spatially distributed soil properties such as the hydraulic conductivity or infiltration, they divided up the spatial domain Ω into four zones ν_1 to ν_4 ($d = 4$); each soil property was then described by a single scalar random variable λ_i on each zone v_i, which was assumed to represent the average value of the soil property over the zone. All these random variables were assumed independent, and first and total-order sensitivity indices were estimated for each of them. Busby et al. (2010) also use a "*dimension reduction*" approach to perform the sensitivity analysis of an oil reservoir model in which one of the model inputs is a spatially distributed permeability field $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. This input field is first expanded on an orthogonal basis (Karhunen-Loève expansion); the main coefficients λ_i of this expansion are then considered as new, uncorrelated inputs, and sensitivity indices are estimated for each of them by a pseudo Monte Carlo procedure. A last example of the "*dimension reduction*" approach is given by Francos et al. (2003), who performed the variance-based sensitivity analysis of the SWAT computer model (Soil and Water Assessment Tool). In the original SWAT model, the study area Ω was divided into 11 topographical sub-basins and further into 156 hydrological response units, and 15 soil properties where given for each of these units. For sensitivity analysis, the spatial variability of soil properties was reduced and each spatially distributed input variable was considered homogeneous over supports of larger size (either sub-basins, or region composed of many hydrological response units grouped by soil types, land use types or growing crops). This dimension reduction resulted in a downsizing of the total number of input variables, from more than 1 000 in the initial model to only 82 in the reduced setting.

3.1.2.3 Switch or trigger input

The "*switch input*" or "*trigger input*" approach was first suggested by Crosetto and Tarantola (2001). In this approach, we assume that the uncertainty on the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ can be simulated by a stochastic process $\mathcal{P}$. The idea of the "*switch input*" method is to modify the original numerical model $\mathcal{F}$ by introducing a so-called "*trigger*" variable, denoted by ξ, which is assumed to be a boolean random variable. When $\xi = 0$, the spatially distributed input $Z(\mathbf{x})$ is kept fixed to its nominal value; when $\xi = 1$, the stochastic process $\mathcal{P}$ is used to generate a random realisation of $Z(\mathbf{x})$. The trigger variable ξ is a switch between two situations (Figure 3.3 on the next page). Using this stratagem, the sensitivity indices estimated for the trigger variable ξ are used as a measure of the sensitivity of the model output Y to the variability of $Z(\mathbf{x})$. The steps are as follows:

Trigger method

Step 1 choose a nominal value $Z^0(\mathbf{x})$ for the spatially distributed input;

Step 2 define a "*trigger input*", denoted by ξ, which is assumed to be a boolean random variable such that $\mathbb{P}(\xi = 0) = \mathbb{P}(\xi = 1) = 1/2$;

Step 3 sample trigger input ξ along with other scalar inputs U_i to generate an input matrix, following the sampling procedure described in §1.2.2.5 on page 32;

Step 4 evaluate model $\mathcal{F}$ for each line i of the input matrix:

- if $\xi^{(i)} = 0$, use nominal value $Z^0(\mathbf{x})$ to evaluate model $\mathcal{F}$;
- if $\xi^{(i)} = 1$, use stochastic process $\mathcal{P}$ to generate a random realisation of $Z(\mathbf{x})$ to evaluate model $\mathcal{F}$;

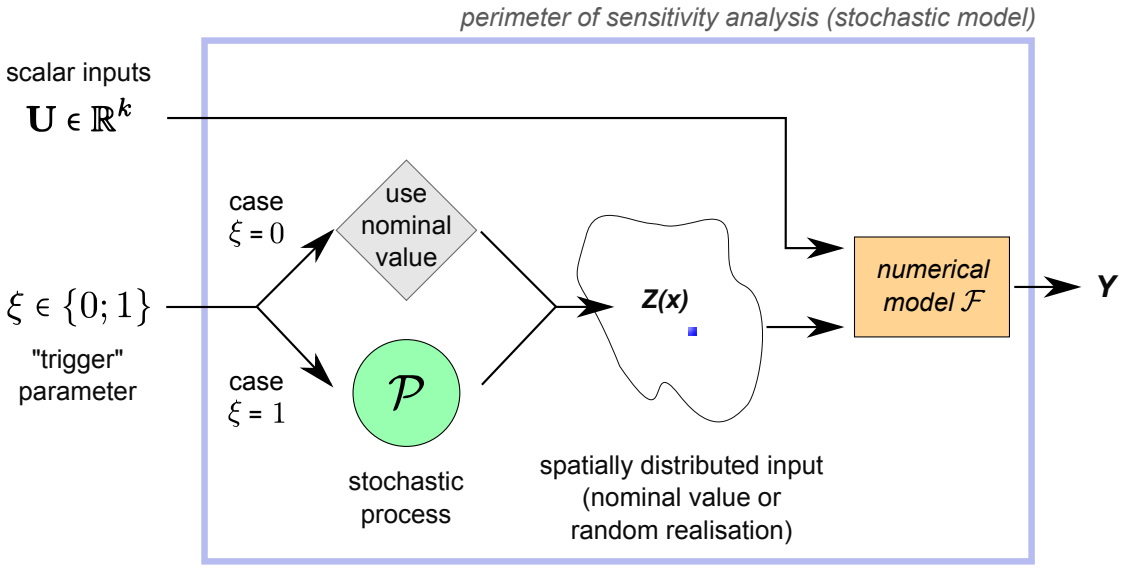

Figure 3.3: Trigger input method

Step 5 from the output vector, calculate the first and total-order sensitivity indices S_ξ and ST_ξ of trigger input ξ (§1.2.2.5 on page 32). These indices are taken as a measure of the influence of $Z(\mathbf{x})$ on the variance of model output Y.

Crosetto and Tarantola (2001) applied this "*trigger*" approach on a GIS-based hydrologic model that simulate flood discharges from forecast rainfall on a given area. Five spatially distributed inputs $Z_i(\mathbf{x})$ were considered: rainfall intensity maps, vector layer on flood-exposed assets, porosity maps, interception map and soil moisture map. These inputs include 2D and 3D data, quantitative raster data, categorical raster data and vector data. A stochastic process $\mathcal{P}_i$ was defined for each of them to generate random error maps to be added to the initial maps. For one of the inputs (the rainfall intensity map), the stochastic process $\mathcal{P}$ accounted for spatial auto-correlation in $Z_i(\mathbf{x})$: random realisations of an error Gaussian Random Field were generated using Cholesky decomposition technique. For the other inputs, stochastic processes $\mathcal{P}_i$ did not account for the spatial auto-correlation of uncertainty. Five trigger variables ξ_1 to ξ_5 were included in the sensitivity analysis and first and total-order sensitivity indices were estimated for each of them using the E-FAST procedure.

3.1.2.4 Map labelling

Lilburne and Tarantola (2009) suggest to represent the uncertainty of $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ using a set of n randomly generated realisations (n possibly large), generated by a stochastic process $\mathcal{P}$. Each realisation is then "*labelled*" by a single integer. The steps are as follows:

Map labelling method

Step 1 generate n random realisations of the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ using a stochastic process $\mathcal{P}$. Store them in some permanent memory space. These realisations are considered as equiprobable;

Step 2 label each realisation with a unique integer $l \in [\![1;n]\!]$; the random realisation associated with label l is denoted by $Z^{(l)}$;

Step 3 consider a random label $\mathbf{L}$, which is assumed to follow a discrete uniform pdf in $[\![1;n]\!]$;

Step 4 sample random label $\mathbf{L}$ along with other scalar inputs U_i to generate an input matrix, following the sampling procedure described in §1.2.2.5 on page 32;

Step 5 evaluate the model $\mathcal{F}$ for each line of the input matrix; spatially distributed input $Z(\mathbf{x})$ is replaced by a reference to the random label $\mathbf{L}$: on line i, the sampled value $l^{(i)}$ of the label indicates that random realisation $Z^{(l^{(i)})}$ of the spatially distributed input must be considered to evaluate model $\mathcal{F}$ for this sample line;

Step 6 from the output vector, calculate the first and total-order sensitivity indices $S_{\mathbf{L}}$ and $ST_{\mathbf{L}}$ of random label $\mathbf{L}$ (§1.2.2.5 on page 32). These indices $S_{\mathbf{L}}$ and $ST_{\mathbf{L}}$ are taken as a measure of the influence of $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ on the variance of model output Y.

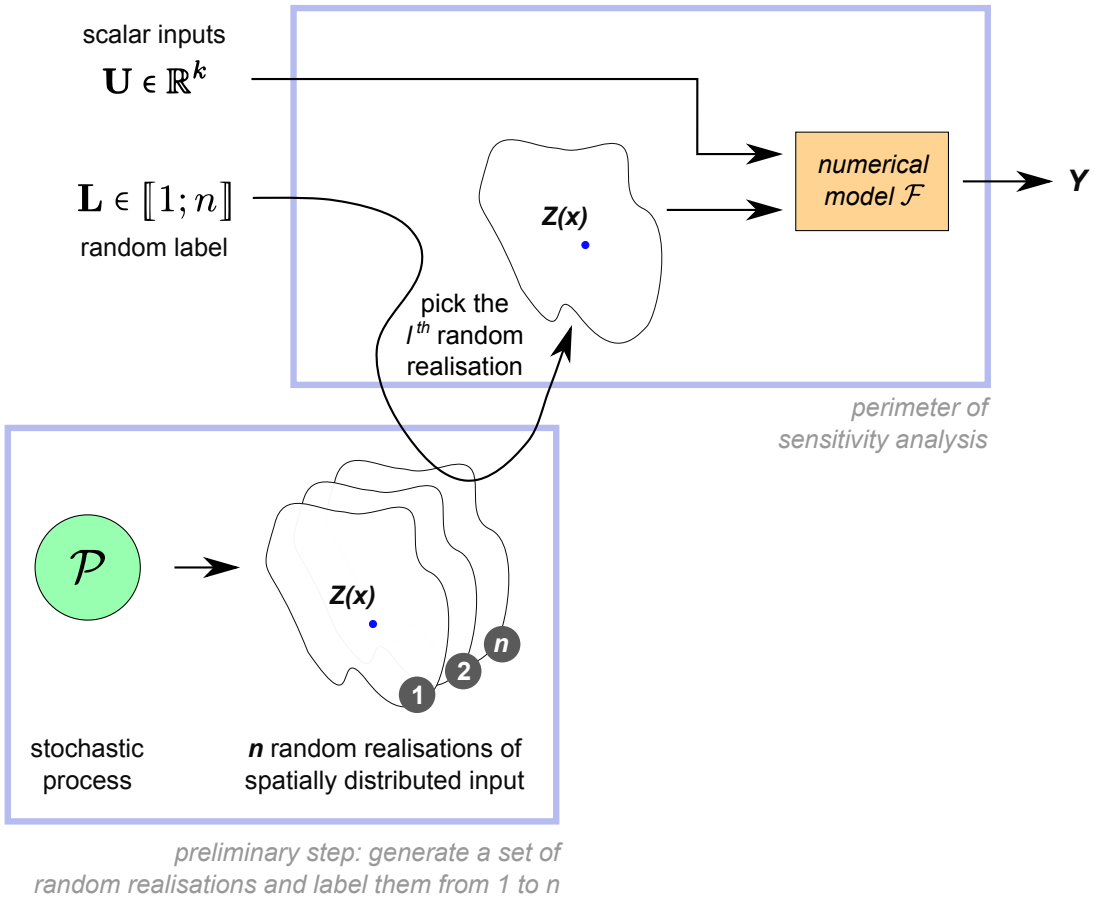

Figure 3.4: Map labelling method

Lilburne and Tarantola (2009) applied the "*map labelling*" approach (Figure 3.4) to a spatial model for simulating nitrate transport from paddock to groundwater (AquiferSim). Four spatially distributed inputs were considered (soil map, land use map, river recharge map and aquifer transmissivity map) and a small set of up to $n = 4$ random realisations was generated for each of them. The "*map labelling*" approach

was also used by Ruffo et al. (2006) to perform sensitivity analysis of a model for oil reservoir production forecasting. Two spatially distributed inputs were considered (basin geometry and heat flow map) and a small set of $n = 8$ (basin geometry) or $n = 4$ (heat flow) random realisations was generated for each of them.

3.1.2.5 Joint meta-models

The "*joint meta-models*" method was introduced by Iooss and Ribatet (2009). In this approach, we also assume that the uncertainty on the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ can be simulated by a stochastic process denoted by $\mathcal{P}$. Iooss and Ribatet (2009) suggest to consider $Z(\mathbf{x})$ as an "*uncontrollable*" input, that is, an input variable whose random values cannot be reproduced or fixed. Two different meta-models— generalized linear models, GLM—are used to approximate the conditional expectation $\mathbb{E}(Y \mid \mathbf{U})$ and the conditional variance $\mathrm{var}[Y \mid \mathbf{U}]$ of model output. These meta-models are linked and are built from a small set of model simulations, in which the value of the uncertain spatially distributed input $Z(\mathbf{x})$ is randomly sampled for each model run. We will not give a detailed description of the steps involved to build these two meta-models. First meta-model on the conditionnal expectation $\mathbb{E}(Y \mid \mathbf{U})$ ("*mean model*") is used to estimate first-order sensitivity indices of scalar inputs $U_1, \ldots, U_k$. The second meta-model on conditional variance $\mathrm{var}[Y \mid \mathbf{U}]$ ("*dispersion model*") is used to estimate total sensitivity index ST_Z of the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. Iooss and Ribatet (2009) apply this method on a model for simulation of nuclear fuel irradiation. Marrel et al. (2012) also apply it on an oil reservoir model, and use geostatistical simulation to obtain random realisations of a porosity map.

If there are several spatially distributed inputs $Z_{k+1}(\mathbf{x}), \ldots, Z_K(\mathbf{x})$, they can be considered as a whole as a single "*uncontrollable*" group of inputs. A couple of first-order and total-order sensitivity indices can then be calculated for the whole group, but not separately for each spatially distributed input $Z_i(\mathbf{x})$.

3.1.2.6 Second level uncertainty modelling

The last method, named "*second level uncertainty modelling*" or simply "*second level*", is not really comparable to the others, because it does not measure the same quantity of interest (see below). However, we think that it is necessary to discuss it, because: i) this method has been encountered in the literature on sensitivity analysis of spatial models; and ii) it is helpful to better understand the difference between this approach and the previous ones.

In this last method, we also assume that the uncertainty on spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ can be simulated by a stochastic process $\mathcal{P}$. More precisely, we consider the case in which this stochastic process $\mathcal{P}$ is controlled by some scalar parameters, denoted by θ_1 to θ_d. These parameters usually aim at controlling some features of the random realisations of $Z(\mathbf{x})$. We can use the term "*second level parameters*" to refer to these parameters θ_i, because they are used to model uncertainty on the first-level model input $Z(\mathbf{x})$. For example, if $Z(\mathbf{x})$ was modelled by a Gaussian Random Field, second level parameters θ_i could be the range, nugget and sill of the variogram. If $Z(\mathbf{x})$ was a vector layer in a GIS software, second level parameters θ_i could control a number of characteristics of the generated maps, such as the number, shape or surface area of the polygonal features. The idea of the "*second level*" approach is to consider that second level parameters θ_i are themselves uncertain, and to estimate their variance-based sensitivity indices with respect to the model output Y[a] (Figure 3.5 on the next page). These parameters

[a]This approach is for example suggested by Bonin (2006) when he says that "*a common practice [...] is to substitute to the input variables the parameters of the errors models designed for these variables.*"

can be grouped for sensitivity analysis or not (see §1.2.2.4 on page 32 for a discussion on grouping in VB-GSA). The steps are as follows:

Second level method

Step 1 identify a number of scalar parameters θ_i that control the stochastic process $\mathcal{P}$

Step 2 identify a pdf p_{θ_i} for each parameter θ_i, or a joint pdf p_θ for the whole group of parameters $(\theta_j)_{j\in[\![1;d]\!]}$

Step 3 sample random parameters θ_i (grouped or not grouped) along with other scalar inputs U_i to generate an input matrix, following the sampling procedure described in §1.2.2.5 on page 32;

Step 4 for each line i of the sample matrix:

- generate a new random realisation of $Z(\mathbf{x})$ with the stochastic process $\mathcal{P}$, using parameter values $(\theta_1^{(i)},\dots,\theta_d^{(i)})$;
- evaluate the code $\mathcal{F}$ using scalar inputs $U_1^{(i)},\dots,U_k^{(i)}$ and the new random realisation of $Z(\mathbf{x})$;

Step 5 (with grouping) from the output vector, calculate the first and total-order sensitivity indices S_θ and ST_θ of the group of parameters $(\theta_j)_{j\in[\![1;d]\!]}$ (§1.2.2.5 on page 32);

Step 5 (without grouping) from the output vector, calculate the first and total-order sensitivity indices S_{θ_i} and ST_{θ_i} for each parameter θ_j (§1.2.2.5 on page 32).

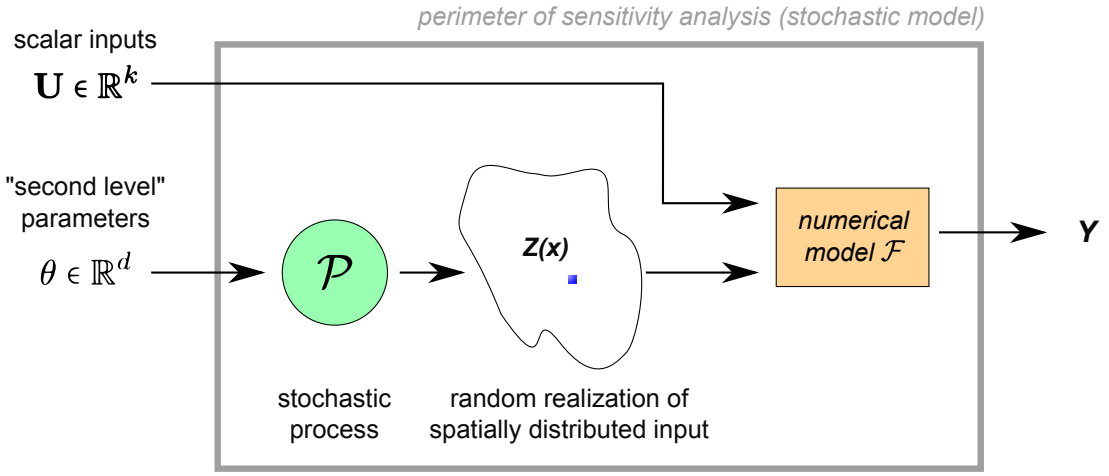

Figure 3.5: Second level method

In fact, it is not the effect of first-level uncertainty of $Z(\mathbf{x})$ on the variance of model output Y that is assessed here, but rather the effect of the uncertainty of the second level parameters θ_i. The model $\mathcal{F}$ under study is modified and becomes a stochastic numerical code: the stochastic process $\mathcal{P}$ which is used to generate random realisations of the spatially distributed input Z is now included in the initial deterministic model $\mathcal{F}$.

The "*second level*" approach is the one chosen by Gumiere (2009) to carry out sensitivity analysis of the MHYDAS water erosion model (Gumiere et al. 2011). They consider a single spatially distributed input Z, which is a map of the locations of vegetative filters that influence sedimentological connectivity,

represented by an oriented tree. To account for the uncertainty in the location of these vegetative filters, they use a stochastic process $\mathcal{P}$ to generate random realisations of Z. This stochastic process is controlled by three "*second level*" scalar parameters θ_1, θ_2 and θ_3 that influence the properties of random trees Z: upstream-dowstream gradient θ_1, density parameter θ_2, contrast parameter θ_3. Finally, they estimate variance-based sensitivity indices of second level parameters θ_1, θ_2 and θ_3 with respect to the MHYDAS-Erosion model outputs, using a procedure based on LHS sampling.

Another example of the "*second level*" approach is curently developed by Monod (2012). They investigate the behaviour of a spatially distributed model that simulates the dispersion of modified genes across plots of cultivated land. One of the model inputs is a map of the landscape which gives the position and geometry of each plot of cultivated land (vector data). To account for the uncertainty on this spatially distrbuted input, they developed an ad-hoc stochastic algorithm $\mathcal{P}$, which can generate any number of random realisations of the landscape, while controlling a number of important features θ_1 to θ_d, such as the level of aggregation of parcels growing the same crops, the empirical distribution of the parcels surface areas, or the shape of parcels (angles). Then, they compute variance-based sensitivity indices associated with each of these scalar parameters θ_i, based on a factorial design of experiments. These sensitivity indices are a measure of the contribution of each feature θ_i of the landscape to the variance of model output Y.

3.1.3 Numerical study

In order to illustrate the various methods presented in §3.1.2 on page 79, we compared them on three analytical test cases. We only give here a brief summary of the settings and results of this numerical study, which is detailed in Appendix §F on page 219.

3.1.3.1 Analytical test cases

Three different numerical models, denoted by $\mathcal{F}_1$ to $\mathcal{F}_3$, were studied. They all have two independent scalar inputs U_1 and U_2, a single spatially distributed input Z and a scalar output Y. We made a restrictive choice by considering only point-based and spatially additive models (see §1.1.4 to §1.1.5 on pages 22–23 for a definition). This choice is justified by the facts that: i) our case study model (NOE model) is point-based and spatially additive; and ii) most theoretical developments presented in this document are related to point-based and spatially additive models only.

The output of interest for all models $\mathcal{F}_1$ to $\mathcal{F}_3$ is the spatial average Y_v of a spatially distributed output of some local code $\mathcal{F}_{1,\text{loc}}$ to $\mathcal{F}_{3,\text{loc}}$ over a disk $v \subset \Omega$ of radius $r = 10$:

$$Y_v = \frac{1}{|v|} \int_{\mathbf{x} \in v} \mathcal{F}_{i,\text{loc}} \left[U_1, U_2, Z(\mathbf{x}) \right] d\mathbf{x} \tag{3.8}$$

U_1 and U_2 are independent scalar inputs with a different uniform pdf for each test case. $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is a 2D random field generated by a deterministic function denoted by $\phi_{\text{camp.}}$. This function was introduced by Campbell et al. (2006) and then modified by Marrel et al. (2011). We use this latter modified version

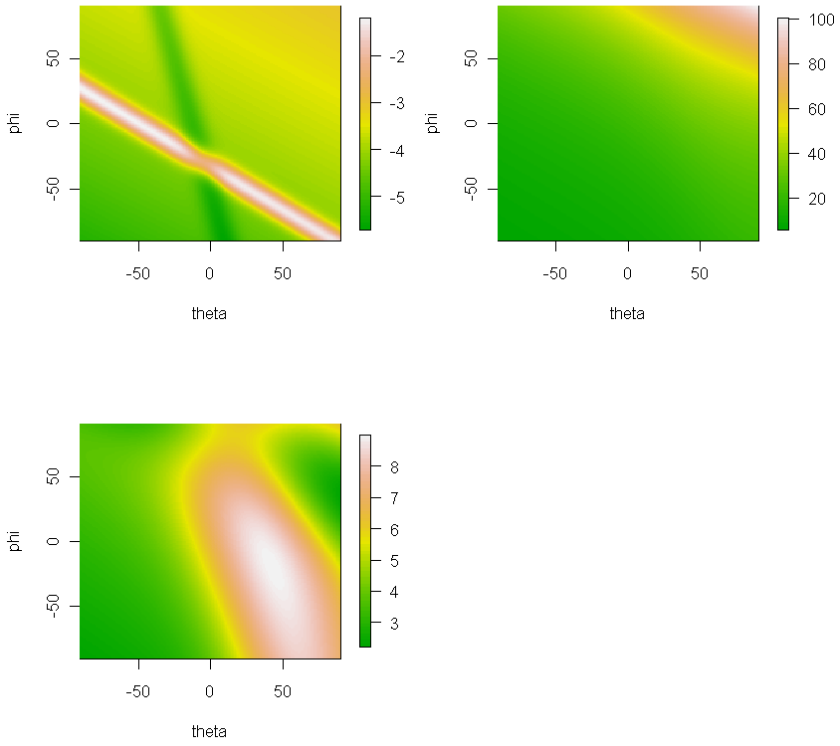

Figure 3.6: Three random realisations of the $\phi_{\text{camp.}}$ function

in our work [Eqn. (3.9)]:

$$\begin{aligned} Z(\mathbf{x}) &= \phi_{\text{camp.}}(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \mathbf{x}) \\ &= \lambda_1 \exp - \frac{(0.8x_\theta + 0.2x_\varphi - 10\lambda_2)^2}{60\lambda_1^2} + (\lambda_2 + \lambda_4) \exp \frac{(0.5x_\theta + 0.5x_\varphi)\lambda_1}{500} \\ &+ \lambda_5(\lambda_3 - 2) \exp \frac{-(0.4x_\theta + 0.6x_\varphi - 20\lambda_6)^2}{40\lambda_5^2} + (\lambda_6 + \lambda_8) \exp \frac{(0.3x_\theta + 0.7x_\varphi)\lambda_7}{250} \end{aligned} \tag{3.9}$$

where x_θ and x_φ are the polar coordinates of point $\mathbf{x} \in \Omega$. Parameters λ_1 to λ_8 are assumed to be i.i.d. random variables of uniform pdf $\mathcal{U}[-1, 5]$. Function $\phi_{\text{camp.}}$ was designed to produce a random field with strong spatial heterogeneity and spatial patterns depending on the values of λ_i parameters (Figure 3.6). In this test case, we can describe the stochastic process $\mathcal{P}$, which is used to generate random realisations of the uncertain spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, as a two step process: first step consists of drawing one random value for each scalar parameter λ_1 to λ_8 from its pdf; and second step is to use the function $\phi_{\text{camp.}}$ to generate a realisation of $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ from the sampled values of λ_i.

Finally, local code $\mathcal{F}_{i,\text{loc}}$ is different for each test case. Local code $\mathcal{F}_{1,\text{loc}}$ is the usual Ishigami function with parameters $A = 7$ and $B = 0.1$ (Homma and Saltelli 1996). $\mathcal{F}_{2,\text{loc}}$ is the usual Sobol G function with parameters $a_0 = 0$, $a_1 = 1$ and $a_2 = 4.5$ (Archer et al. 1997). Local code $\mathcal{F}_{3,\text{loc}}$ is a simple linear function: $\mathcal{F}_{3,\text{loc}}[u_1, u_2, z] = u_1 \cdot z + u_2$.

3.1.3.2 Methods under study

Sensitivity indices of model inputs U_1, U_2 and Z are computed for each test case $\mathcal{F}_1$ to $\mathcal{F}_3$, using the various methods discussed in §3.1.2 on page 79—except for the "*joint meta-modelling*" method. The same pseudo Monte Carlo procedure described in §1.2.2.5 on page 32 was used for each method, with a base sampling size $N = 2\,048$. We give here some explanations on the procedures, for more details see Appendix §F on page 219.

Dimension reduction/ macro-parameter (with or without grouping) In our test cases, the spatially distributed model input $Z(\mathbf{x})$ is generated with the deterministic function $\phi_{\text{camp.}}$. Here, dimension reduction is easy to perform: we just have to consider the eight scalar parameters λ_1 to λ_8, which are assumed to be independent. In this specific case, dimension reduction does not result in a reduction of the information contained in the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$: hence, dimension reduction approach is here similar to the macro-parameter approach. Both "*with grouping*" and "*without grouping*" options are implemented.

Map labelling A set of $n = 1\,000$ random realisations of the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ was generated using function $\phi_{\text{camp.}}$ on a random sample of scalar parameters λ_1 to λ_8. Sensitivity indices were estimated for test cases $\mathcal{F}_1$ to $\mathcal{F}_3$ following the method described in §3.1.2.4 on page 84.

Trigger method Trigger method was used to estimate sensitivity indices on test cases $\mathcal{F}_1$ to $\mathcal{F}_3$, with the following setting: i) when trigger input $\xi = 0$, then spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is fixed to its nominal value, obtained with the $\phi_{\text{camp.}}$ function for the following values of λ_i parameters: $(5, 3, 1, -1, 5, 3, 1, -1)$; ii) when trigger input $\xi = 1$, then spatially distributed input $Z(\mathbf{x})$ is generated with function $\phi_{\text{camp.}}$ from a random set of parameters $(\lambda_1, \ldots, \lambda_8)$.

Second level Here, the stochastic process $\mathcal{P}$ used to generate random realisations of $Z(\mathbf{x})$ is controlled by the pdf of scalar parameters λ_i, which are initially supposed to follow a uniform distribution $\mathcal{U}[-1, 5]$. To apply the "*second level*" approach described in §3.1.2.6 on page 86, we considered that the two boundaries of the interval $[-1, 5]$ were themselves uncertain. We denoted these two boundaries by θ_1 and θ_2, and assumed they were random variables with the following laws: $\theta_1 \sim \mathcal{U}[-2, 0]$ and $\theta_2 \sim \mathcal{U}[4, 6]$, respectively. We then computed first and total-order sensitivity indices for the group (θ_1, θ_2) as explained in §3.1.2.6 on page 86.

3.1.3.3 Results

Table 3.1 on the facing page gives sensitivity indices estimates of model inputs with respect to the outputs of models $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$. Notations S_Z and ST_Z denote the measure of sensitivity related to the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. This measure depends on the method considered:

- for "*dimension reduction with grouping*" method, S_Z and ST_Z are the sensitivity indices of the group of scalar inputs $(\lambda_1, \ldots, \lambda_8)$;

Table 3.1: Sensitivity indices estimates for each test case and each method (base sample size $N = 2\,048$)

	Méthode	S_{U_1}	S_{U_2}	S_Z	ST_{U_1}	ST_{U_2}	ST_Z
$\mathcal{F}_1$	Map labelling	0.85029021	-0.00240045	0.08615907	0.85400421	0.06653359	0.14921352
	Dim. red. (no groups)	0.83122767	-0.00732305	-0.02939551	0.82528089	0.07106520	0.7841338
	Dim. red. (groups)	0.83122767	-0.00732305	0.09104848	0.82528090	0.07106520	0.1714519
	Trigger	0.90419686	-0.00668187	0.00597127	0.98693707	0.09534858	0.09948201
	Second level	0.41921743	0.0412273	0.351084231	0.48971827	0.08128145	0.36249868
$\mathcal{F}_2$	Map labelling	0.39106100	0.16071710	0.26105662	0.54440171	0.2550799	0.4013758
	Dim. red. (no groups)	0.40854115	0.15463054	0.16372373	0.54443509	0.24548002	0.4202739
	Dim. red. (groups)	0.40854121	0.15463054	0.25502153	0.54443517	0.24548001	0.3775389
	Trigger	0.36150243	0.14331481	0.11839480	0.71237405	0.47894579	0.43113721
	Second level	0.21557175	0.08716610	0.53112684	0.36419231	0.19001217	0.62068430
$\mathcal{F}_3$	Map labelling	0.16185561	0.51015161	0.36420441	0.16981867	0.4657908	0.3632460
	Dim. red. (no groups)	0.15610543	0.46926415	0.31304476	0.17798195	0.47171151	0.4054459
	Dim. red. (groups)	0.15610543	0.46926415	0.34933227	0.17798203	0.47171151	0.3750211
	Trigger	0.11714831	0.55745845	0.11311281	0.35444583	0.77012715	0.32716394
	Second level	0.13164180	0.62611001	0.03006099	0.18131888	0.67992681	0.03550100

- for "*dimension reduction without grouping*" method, we obtain individual sensitivity indices of scalar inputs $\lambda_1, \ldots, \lambda_8$, which are given in Appendix §F on page 219; a naive way to build sensitivity indices S_Z and ST_Z is to compute the sum of first-order indices $S_{\lambda_1}, \ldots, S_{\lambda_8}$ of scalar inputs $\lambda_1, \ldots, \lambda_8$ as well as the sum of their total-order indices $ST_{\lambda_1}, \ldots, ST_{\lambda_8}$; we thus indicate these two sums, but we will explain in the discussion why they are misleading (§3.1.4.1 on the following page);
- for "*map labbeling*" method, S_Z and ST_Z are the sensitivity indices of random label input $\mathbf{L}$;
- for "*trigger*" method, S_Z and ST_Z are the sensitivity indices of trigger input ξ;
- for the sake of comparison and though we know that it does not measure the same quantity of interest as discussed in §3.1.2.6 on page 86, we also indicate the measure obtained with the "*second level*" method: in this case, S_Z and ST_Z are the sensitivity indices of the group of inputs (θ_1, θ_2).

In this specific case study, sensitivity indices estimates obtained with the "*dimension reduction with grouping*" approach can be taken as reference values. Indeed, for all test cases $\mathcal{F}_1$ to $\mathcal{F}_3$, the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is completely represented by the set of scalar inputs λ_1 to λ_8, and thus the

first and total-order sensitivity indices of the group $(\lambda_1, \ldots, \lambda_8)$ are a perfect measure of the contribution of $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ to the variance of model output Y.

It first appears in Table 3.1 on the preceding page that the "*map labelling*" method gives the same values of sensitivity indices as the "*dimension reduction with grouping*" approach, for all sensitivity indices and all test cases—max bias is lesser than $5 \cdot 10^{-2}$. We can interpret these results as an empirical validation of the "*map labelling*" method.

Next, it can be noted that "*dimension reduction without grouping*" and "*dimension reduction with grouping*" methods yield identical estimates for sensitivity indices S_{U_1}, S_{U_2}, ST_{U_1} and ST_{U_2}—max bias is lesser than 10^{-2}. On the contrary, the two methods differ in the information they bring on the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. The "*dimension reduction without grouping*" method gives sensitivity indices estimates for each scalar parameter λ_i. The sum of first-order sensitivity indices $\sum_{i=1}^{8} S_{\lambda_i}$ could be chosen as a measure of the main effect of spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ on the variance of model output Y. Yet, this sum does not account for the contribution of the interactions between scalar inputs λ_i to the variance of Y. Hence sum $\sum_{i=1}^{8} S_{\lambda_i}$ is always smaller than or equal to first-order sensitivity index S_Z. In a similar way, the sum of total-order sensitivity indices ST_{λ_1} to ST_{λ_8} could be seen as a good indication of the total contribution of spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ to the variance of model ouput Y. Nevertheless, in this sum, the interactions between scalar parameter λ_i and other parameters λ_j are counted multiples times (once for each parameter), while they are only counted once in total-order sensitivity index S_Z. Hence, the sum $\sum_{i=1}^{8} ST_{\lambda_i}$ is always larger than or equal to total-order sensitivity index ST_Z.

Finally, Table 3.1 on the previous page clearly suggests that the "*trigger*" method and the "*second level*" method do not yield correct estimates of sensitivity indices, as they are far from the reference values obtained with "*dimension reduction with grouping*". In particular, with the "*trigger*" method, first-order index S_Z is always over-estimated while total-order sensitivity indices ST_{U_1} and ST_{U_2} are under-estimated. This point will be further discussed in §3.1.4.1 on the facing page.

3.1.4 Discussion

We discuss here the pros and cons of the various methods displayed in §3.1.2 along with the results of the numerical study described in §3.1.3. The content of this discussion is summarized in Table 3.2 on page 97.

3.1.4.1 Methods do not produce the same information

All the methods displayed in §3.1.2 intend to define variance-based sensitivity indices that could measure the influence of an uncertain spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ on the variance of model output Y. However, they do not produce exactly the same information.

Influence of $Z(\mathbf{x})$ with fixed description of uncertainty A first group of a methods intend to measure the contribution of the spatially distributed input $Z(\mathbf{x})$ to the variance of model output Y, with a "*fixed*" description of uncertainty on $Z(\mathbf{x})$. This group includes the "*macro-parameter*", the "*dimension reduction*", the "*map labelling*" and the "*joint meta-models*" techniques.

The "*macro-parameter with grouping*", "*dimension reduction with grouping*", "*map labelling*" and "*joint meta-modelling*" methods all result in the estimation of variance-based sensitivity indices S_Z and ST_Z that describe the contribution of the whole spatially distributed model input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ to the variance of model output Y. It should be noted that in the "*joint meta-modelling*" method, only total-order sensitivity index ST_Z can be estimated but first-order index S_Z remains unknown. Besides, the "*dimension reduction with grouping*" method only yields an approximation of sensitivity indices of $Z(\mathbf{x})$, as the total information initially contained in $Z(\mathbf{x})$ is reduced to a small set of scalar parameters.

The information brought by the "*macro-parameter without grouping*" or "*dimension reduction without grouping*" methods is slightly different. They give separate estimates of the first and total-order sensitivity indices S_{λ_j}, ST_{λ_j} for each scalar parameter λ_j, $j \in \{1, \ldots, d\}$. As mentioned in §3.1.3 on page 88, the set of first-order indices $\left(S_{\lambda_j}\right)_{j\in[\![1;d]\!]}$ does not bring the same information as the first-order sensitivity index S_Z of the whole spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, because S_Z accounts for the role of the interactions between $(\lambda_j)_{j\in[\![1;d]\!]}$ while first-order indices S_{λ_j} do not. In the same way, the set of total-order indices $\left(ST_{\lambda_j}\right)_{j\in[\![1;d]\!]}$ does not bring the same information as the total-order sensitivity index ST_Z, because interactions between scalar parameters λ_j are counted multiple times in the set of indices $\left(ST_{\lambda_j}\right)_{j\in[\![1;d]\!]}$ while it is counted just once in ST_Z. Hence the set of sensitivity indices $\left(S_{\lambda_j}, ST_{\lambda_j}\right)_{j\in[\![1;d]\!]}$ does not yield a good measure of the main and total contributions of $Z(\mathbf{x})$ to the variance of model output Y.

Discarding the trigger method Results of the numerical study (Table 3.1 on page 91) also clearly suggest that the "*trigger*" method does not produce the same information as other methods displayed in §3.1.2. Sensitivity index of the "*trigger*" parameter ξ proved to be always smaller than the measure S_Z brought by the "*macro-parameter*" or "*map labelling*" methods. One possible explanation is that, in the "*trigger*" method, uncertainty in $Z(\mathbf{x})$ is taken into account only when the sampled value of the trigger input is equal to $\xi = 1$, that is, one out of every two model runs in average ($\mathbb{P}(\xi = 0) = \mathbb{P}(\xi = 1) = 1/2$). Hence, the effect of uncertain model input $Z(\mathbf{x})$ on the variance of model output Y is systematically under-estimated. This result leads us to believe that the "*trigger*" method is not appropriate to deal with spatially distributed inputs in VB-GSA.

Influence of second level uncertainty As mentionned in §3.1.2.6 on page 86 and corroborated by the numerical study (Table 3.1 on page 91), the "*second level*" method does not produce the same measures of importance as other techniques. In this approach, random realisations of the spatially distributed input $Z(\mathbf{x})$ are generated by a stochastic process $\mathcal{P}$. This stochastic process is controlled by a number of scalar parameters θ_i that are themselves considered as uncertain. What is measured in this method is the contribution of these uncertain "*second level*" parameters θ_i to the variance of the model output Y. This measure is by definition different from the contribution of the uncertain spatially distributed input $Z(\mathbf{x})$ to the variance of Y for a "*fixed*" stochastic process $\mathcal{P}$ (i.e., a stochastic process with constant values of parameters θ_i). Hence, the "*second level*" method cannot really be compared to other techniques, as they do not aim at measuring the same effects.

Influence of map attributes Finally, we can mention another possible situation, which is not handled by any of the techniques that were described in §3.1.2, and which does not measure the same quantity of interest either. Let consider the case in which a "*fixed*" stochastic process $\mathcal{P}$ is used to generate random realisations of the uncertain spatially distributed input $Z(\mathbf{x})$. As explained earlier, most of the methods described in §3.1.2 measure the entire contribution of $Z(\mathbf{x})$ to the variance of the model output Y. However, if $Z(\mathbf{x})$ appears to be an influential model input, then the modeller may want to know

more precisely which "*map attributes*" contribute the most to the variability of the model output. A "*map attribute*" can be any scalar descriptor of the spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. For example, if $Z(\mathbf{x})$ is modelled as a Stationary Random Field, some scalar descriptors are the value $Z(\mathbf{x}^*)$ at a specific location, the average value of $Z(\mathbf{x})$ over some spatial support, the proportion of the spatial domain Ω for which $Z(\mathbf{x})$ exceeds a certain threshold, etc. If the spatially distributed input is a map of plots of cultivated land (vector data), then the following scalar descriptors could be chosen: number of plots of land, average surface area of the plots, indicator of aggregation, etc. Two key characteristics of these "*map attributes*" must be highlighted. First, they cannot be sampled in a controlled way: when the "*fixed*" stochastic process $\mathcal{P}$ is used to generate random realisations of $Z(\mathbf{x})$, the map attributes associated to each random realisation of $Z(\mathbf{x})$ will usually have different and uncontrolled values. These "*map attributes*" are thus different from the "*second level*" parameters θ_i presented earlier. Second, if the modeller considers several different "*map attributes*" at the same time, they will most often be correlated. To assess how much the selected scalar map attributes of $Z(\mathbf{x})$ contribute to the variance of model output Y, a possible way is to proceed in two steps: i) first, carry out VB-GSA of the model using for example the "*map labelling*" approach to estimate the sensitivity index S_Z associated to $Z(\mathbf{x})$; ii) then, using the same set of model runs, perform a complementary graphical analysis to discuss qualitatively the influence of the different map attributes. This graphical analysis may for example make use of scatterplots or Contribution to the Sample Mean plot (Bolado-Lavin et al. 2009). Up to our knowledge, this approach has never been used in the literature.

3.1.4.2 Computational cost

The methods displayed in §3.1.2 also differ in their computational cost. The cost of a method depends on: i) the number of model simulations it requires, and ii) the number of random realisations of spatially distributed input $Z(\mathbf{x})$ is needed. We compare these costs with the hypothesis that the same sampling procedure (described in §1.2.2.5 on page 32) is used for all methods in which sensitivity indices are estimated from a pseudo Monte Carlo sample, with a base sample size N and a total sample size $N_{\text{tot}} = (K + 2) \cdot N$ where K is the number of model inputs or groups of model inputs. Computational costs associated with each method are given in Table 3.2 on page 97. It appears that the least CPU-intensive approaches are the "*map labelling*" and the "*joint meta-modelling*" methods.

3.1.4.3 Spatial structure of uncertainty

Some methods displayed in §3.1.2 enable a complex correlated description of the spatial structure of uncertainty in $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, others do not. In the "*map labelling*", "*trigger*" and "*joint meta-modelling*" methods, the random realisations of $Z(\mathbf{x})$ can be generated using any stochastic process $\mathcal{P}$, based on any algorithm and any software, allowing complex spatial descriptions of variability to be simulated. Spatial auto-correlation can be taken into account and modelled (e.g., using Random Fields and geostatistical simulation). On the contrary, in the "*macro-parameter*" or "*dimension reduction*" methods, modelling the spatial structure of uncertainty in $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is more complicated: it requires to be able to characterise the joint pdf p_λ of the scalar parameters $(\lambda_j)_{j \in \{1,\dots d\}}$ that represent spatially distributed input $Z(\mathbf{x})$, which is often impossible.

3.1.4.4 Coupling with a meta-model

When the CPU cost of a single run of model $\mathcal{F}$ is too high (typically more than one minute), estimating sensitivity indices with an intensive sampling-based approach becomes computationally intractable. In this case, all the methods described in §3.1.2 are inapropriate, except for the "*joint meta-modelling*" method which only requires a small set of model runs. Nevertheless, some of the other methods reviewed can be coupled with meta-modelling to estimate sensitivity indices at a lower CPU cost. This is the case for the "*dimension reduction*" method: a meta-model can be built from a small set of runs of the modified numerical model $\mathcal{F}^\star$ to estimate model output Y from the set of scalar inputs U_i and the set of parameters $(\lambda_j)_{j\in[\![1;d]\!]}$. Sensitivity indices can then be estimated at a low cost from the meta-model through any intensive sampling-based procedure. It is also the case for the "*second level*" approach, in which only scalar inputs are considered (scalar inputs U_i and second level parameters θ_i). On the contrary, it does not seem possible to build meta-models coupled with the "*macro-parameter*", "*trigger*" or "*map labelling*" methods. In the "*macro-parameter*" approach, the number of scalar inputs λ_j that should be included in the meta-model is usually too large ($> 1\,000$). In the "*trigger*" and "*meta-modelling*" methods, the trouble comes from the nature of model inputs: trigger variable ξ is a boolean variable, "*map label*" input $\mathbf{L}$ is a random integer. These inputs can not be included, up to our knowledge, in the usual meta-models (Gaussian Processes, MARS, etc.) that only cope with scalar random variables with continuous pdf.

3.1.4.5 Decision tree to choose the appropriate method

For a given model $\mathcal{F}$ with spatially distributed inputs $Z_i(\mathbf{x})$, one may look at the following criteria to choose from the methods for VB-GSA displayed in §3.1.2: i) the number of spatially distributed inputs $Z_i(\mathbf{x})$; ii) the dimension of spatially distributed inputs $Z_i(\mathbf{x})$ (number of cells or number of spatial objects); iii) the possibility to describe $Z_i(\mathbf{x})$ with a small number of scalar parameters; and iv) the cost of the generation and storage of each random realisation of $Z_i(\mathbf{x})$. We suggest to use a decision tree (Figure 3.7 on the following page) to choose among the various methods based on these criteria.

It must be noted that this decision tree is suitable for the case of a model $\mathcal{F}$ with a low CPU cost, for which intensive simulation is possible. When the cost of one run of model $\mathcal{F}$ is too high, then meta-modelling must be used to lower the computational burden of the analysis. As discussed in §3.1.4.4, some of the methods for VB-GSA with spatial inputs can be coupled with a meta-model: they are identified by a dashed box in Figure 3.7. Others cannot: in particular, there is no available solution to apply VB-GSA to a model with high CPU-cost and several spatially distributed inputs $Z_i(\mathbf{x})$ when these inputs cannot be reduced to a small set of scalar parameters.

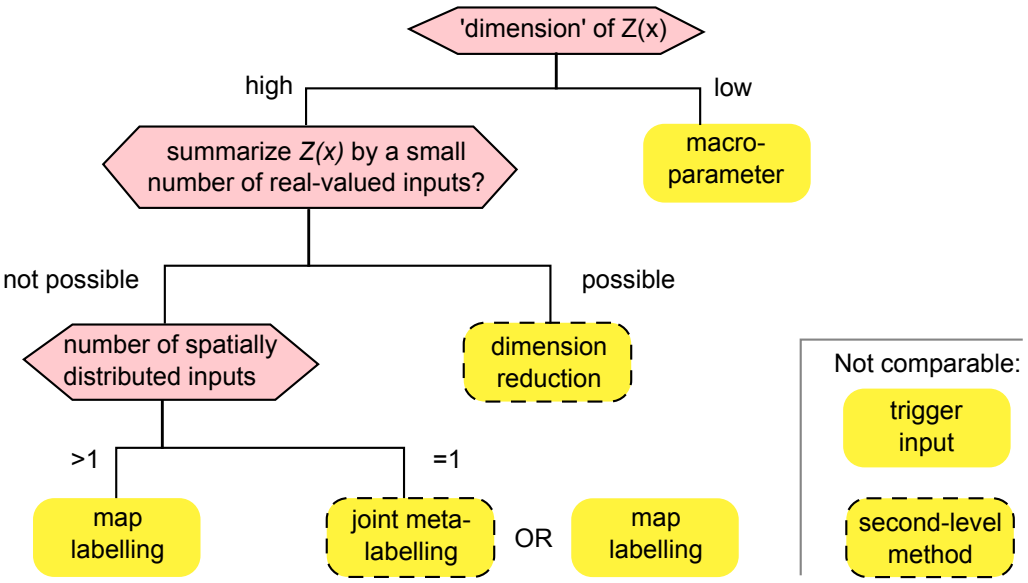

Figure 3.7: Decision tree for VB-GSA with spatially distributed inputs; dashed boxes indicate that the method can be coupled with meta-modelling

Table 3.2: Methods for VB-GSA with spatially distributed inputs

Method	# of map realisations needed	# of model runs needed	Coupling with a meta-model is possible	Possibility to cope with several spatial inputs	Sensitivity indices associated with $Z(\mathbf{x})$	Possibility to model spatial auto-correlation in $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$	Comments
Macro-parameter	N	$N \cdot (k+1+2)$	No	Yes	indices of the group of inputs $(\lambda_j)_{j\in\{1,\dots d\}}$	No	Possibly intractable for large data
Dimension reduction (with groups)	N	$N \cdot (k+1+2)$	No	Yes	indices of the group of inputs $(\lambda_j)_{j\in\{1,\dots d\}}$	No	Simplification of the input data
Dimension reduction (without groups)	N	$N \cdot (k+d+2)$	Yes	Yes	indices of each input λ_i, $i = 1 \dots d$	No	Simplification of the input data
Trigger	$\sim N/2$	$N \cdot (k+1+2)$	No	Yes	indices of the trigger input ξ	Yes	Sensitivity indices do not have the same meaning
Map labelling	$n < N$	$N \cdot (k+1+2)$	No	Yes	indices of the label input $\mathbf{L}$	Yes	Spatially distributed input $Z(\mathbf{x})$ is under-sampled compared to other inputs
Joint GAM metamodel	$\ll N$	$\ll N$	Yes	No	total-order index ST_Z only	Yes	Use of a metamodel
Second level	N	$N \cdot (k+d+2)$	Yes	Yes	indices of second level parameters θ_i	Yes	Sensitivity indices do not have the same meaning

3.2 Sampling issues

3.2.1 Introduction

We displayed in §3.1 on page 79 a number of methods to compute variance-based sensitivity indices for a spatially distributed model input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$. Most of these methods require sampling a number of random realisations of $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, that can be generated by any stochastic process $\mathcal{P}$. The size and quality of this sampling will influence the precision and accuracy of sensitivity indices estimates, but also the CPU cost of the analysis. With time consuming models and large input data, using an effective sampling scheme is necessary to get the most accurate sensitivity indices with the fewest model runs and/or the fewest random realisations of $Z(\mathbf{x})$. This issue of input sampling has been extensively discussed when model inputs are scalar random variables U_i; some papers (Kucherenko et al. 2011) are devoted to the comparison of various sampling techniques such as Latin Hypercube Sampling (LHS) or $LP - \tau$ sequences and discuss their influence on the quality of sensitivity indices estimates (accuracy, precision and convergence).

In our work, we focused on the "*map labelling*" approach to estimate sensitivity indices of spatially distributed inputs in VB-GSA (§3.1.2.4 on page 84). The "*map labelling*" approach uses a number n of random realisations of the spatially distributed input $Z(\mathbf{x})$ that may be smaller than the base sample size N of the samples used to estimate sensitivity indices. The choice of size n is usually driven by constraints of time (generating random realisations of $Z(\mathbf{x})$ using stochastic process $\mathcal{P}$ may be CPU-intensive) and constraints of disk space (storing a too large number of spatially distributed data with high spatial resolution may be intractable). Efficient spatial sampling techniques are thus needed to generate random realisations of spatially distributed inputs $Z_i(\mathbf{x})$ and get accurate sensitivity indices estimates at low cost.

We start this section in §3.2.2 by a brief review on the various strategies to model uncertainty on a spatially distributed input $Z(\mathbf{x})$, describing the main stochastic processes $\mathcal{P}$ that can be used to generate random realisations of $Z(\mathbf{x})$. Next, we investigate the issue of spatial sampling for VB-GSA with the "*map labelling*" method in two ways. In §3.2.3 on page 100, we observe the convergence of sensitivity indices estimates provided by the "*map labelling*" method on analytical test cases $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ for increasing number n of random realisations of $Z(\mathbf{x})$. Then, in a work presented at the Accuracy 2010 conference (Saint-Geours et al. 2010) and reproduced here in §3.2.4 on page 101, we study a simple point-based and spatially additive analytical test case in which spatially distributed input $Z(\mathbf{x})$ is modelled as a Gaussian Random Field, and we compare two geostatistical simulation algorithms to generate random realisations of $Z(\mathbf{x})$: Simple Random Sampling (SRS) and spatial Latin Hypercube Sampling (LHS). The purpose of this work is to assess whether spatial LHS yield better sensitivity indices estimates (better accuracy, precision or convergence with increasing n) than SRS. We finally briefly discuss the outcomes of both numerical studies in §3.2.5 on page 110.

3.2.2 Modelling uncertainty on spatially distributed inputs: a review

In this section, we very briefly survey the various methods by which uncertainty can be modeled and simulated for spatially distributed data. Our motivation for this survey is the following: in §3.1, we displayed methods to handle a spatially distributed input $Z(\mathbf{x})$ in variance-based global sensitivity analysis; in most of these methods[b], we assume that the uncertainty on $Z(\mathbf{x})$ can be simulated using a stochastic

[b] Namely, the map labelling, trigger, joint meta-model and second level methods.

process or any ad-hoc algorithm, which is denoted by $\mathcal{P}$. This stochastic process $\mathcal{P}$ is supposed to be able to generate any number of random realisations of $Z(\mathbf{x})$. Our goal here is to give an quick overview of the main algorithms $\mathcal{P}$ that can be used for that purpose, depending on the nature of the spatially distributed input $Z(\mathbf{x})$. We know that the issue of modelling uncertainty for geospatial data has been extensively discussed in the literature: it is by no means the scope of this section to display an exhaustive review on this question. The interested reader will find more material in Fisher et al. (2005) or Hunsaker et al. (2001).

We offer to distinguish between uncertainty modelling for raster spatial data and vector spatial data.

3.2.2.1 Uncertainty models for raster data

Up to our knowledge, the vast majority of methods given in the literature to model uncertainty for spatial raster data is based on geostatistics (Chilès and Delfiner 1999).

Quantitative raster data For quantitative raster data (e.g., a Digital Terrain Model), the usual models found in the literature consist of perturbing the nominal raster data with some spatially distributed error field, which is most often modelled as a Stationary Random Field (Bonin 2006). A wide range of geostatistical simulation algorithms $\mathcal{P}$ are available to simulate random realisations of the error field, such as Sequential Gaussian Simulation, Turning Bands or Cholesky decomposition (Lantuéjoul 2002). These simulations algorithms may be conditional if they take into account some known data—e.g., the value of the error field at some specific locations, the average value over a block, etc.—or unconditional if they don't. They may be exact or approximated. A number of open-source packages and softwares are available to easily implement these simulation algorithms, such as SGeMS (Remy et al. 2009) or the `RandomFields` package on R statistical software (Schlather 2001).

Aerts et al. (2003) give an example of how to use geostatistical simulation to account for uncertainty on the raster input of a spatial model. They consider a model which aims at finding an optimal location for a ski run, using a slope map which is derived from a Digital Terrain Model of the study area. They model the uncertainty on the input DTM with a stationary error random field, whose characteristics—i.e., variogram parameters—are determined from a set of ground control points. Then, they use Sequential Gaussian Simulation to generate a set of 500 random realisations of the error field which is added to the inital DTM. The whole procedure results in 500 equally probable DEMs, which are subsequently used to propagate uncertainty through the ski run allocation model.

Categorical raster data For categorical raster data (e.g., land use determined by classification on a satellite image), the usual uncertainty models are based on the indicator kriging theory. Hession et al. (2006) display a review of the various geostatistical simulation algorithms $\mathcal{P}$ that have been developed in this framework. Besides, another option to simulate uncertainty on categorical data is based on the information given by a "*confusion matrix*". The confusion matrix is mostly used to assess the accuracy of land use classifications: each value $p_{i,j}$ of the confusion matrix is the proportion of raster cells classified as members of the land use class i but actually belonging to land use class j according to ground control data. From the information contained in the confusion matrix, it is possible to simulate random confusions between land use labels, and thus to generate random realisations of land use maps from a nominal land use raster data. This method named "*confusion frequency simulation*" was initially suggested by Fisher (1991). Finally, there are also some non-standard methods to model uncertainty in categorical raster data, such as neural networks or Markov chains, which are briefly reviewed by Hession et al. (2006).

3.2.2.2 Uncertainty models for vector data

As clearly stated by Bonin (2006), "*things are much more complicated when dealing with vector data*". Indeed, modelling uncertainty for GIS vector layers is quite difficult and this issue has been paid less attention in the literature than questions related to uncertainty in raster data. Brown et al. (2005) suggest to divide uncertainty in spatial vector data into "*positional uncertainty*" and "*attribute uncertainty*".

Positional uncertainty "*Positional uncertainty*" refers to the error on the position and shape of the objects in space. The basic geometric components of a GIS vector layer are points, lines and polygons. Girres and Julien (2010) exhaustively list the various sources of positional uncertainty in a GIS vector layer, including projections, georeferencing, generalization of object boundaries, digitalization, etc. Brown et al. (2005) further suggest to distinguish point objects, rigid objects and deformable objects which behave differently with respect to positional uncertainty. However, despite extensive research on the field, Bonin (2006) observes that very few statistical models exist to describe uncertainties in the position and shape of GIS geometrical features, and even fewer stochastic processes $\mathcal{P}$ have ever been designed to simulate such uncertainties. This is an open research question, which is in particular investigated in the field of image segmentation evaluation (Neubert et al. 2008).

Attribute uncertainty "*Attribute uncertainty*" refers to the errors on the various descriptive data associated with each object in a GIS vector layer. These attribute data can be quantitative (e.g., the density of a country), categorical (e.g., the crop cultivated on a plot of land) or even textual (e.g., the name of a city). Many typologies of attribute data uncertainty are offered in the literature (Radoux et al. 2011). For example, MacEachren et al. (2005) build a typology based on the following nine components: accuracy/error, completeness, consistency, credibility, currency, interrelatedness, lineage, precision/resolution, and subjectivity. However, these typologies of uncertainty are almost never associated with any stochastic process $\mathcal{P}$ or ad-hoc algorithm that could simulate random realisations of uncertain attribute data. One notable exception is vector landuse maps—in which land use class is a categorical attribute—for wich the "*confusion frequency simulation*" described for raster data in §3.2.2.1 on the preceding page can be applied (Lilburne et al. 2006). This latter method will be used to model uncertainty on the assets map in the NOE code.

3.2.3 Convergence of sensitivity indices estimates with map labelling method

3.2.3.1 Introduction

This subsection is devoted to a numerical study of the convergence of sensitivity indices estimates with the "*map labelling*" approach described in §3.1.2.4 on page 84. The "*map labelling*" approach requires to generate a set of n random realisations of spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ before launching sensitivity analysis. The accuracy of sensitivity indices estimates depends on the number n of these random realisations. We assess the effect of n on sensitivity indices estimates for the three analytical test cases which we already studied in §3.1.3 on page 88.

3.2.3.2 Method

We considere the test cases $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ that were already studied in §3.1.3 on page 88. In these test cases, $Z(\mathbf{x})$ is a random field generated by the function $\phi_{\text{camp.}}$. For each test case, we compute sensitivity

indices estimates with the "*map labelling*" approach for an increasing number n of random realisations of spatially distributed input $Z(\mathbf{x})$. The following procedure is used for $n = 3, 10, 25, 50, 100, 500$ and $1\,000$:

Step 1 generate a set of n random realisations of spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$, using the stochastic process $\mathcal{P}$ described in §3.1.3.1 on page 88;

Step 2 estimate sensitivity indices from this set of random realisations with the "*map labelling*" approach, for a base sample size $N = 2\,048$ (§3.1.2.4 on page 84);

Step 3 repeat the steps 1 and 2 a hundred times;

Step 4 compute for each sensitivity index its empirical mean and its empiral standard deviation over the 100 replicas.

In order to assess the accuracy of the sensitivity indices estimates, we compare them to the reference values obtained with the "*dimension reduction with grouping*" approach, given in Table 3.1 on page 91.

3.2.3.3 Results

Results (Figure 3.8 on the next page) show that, in the "*map labelling*" approach, the accuracy of sensitivity indices estimates increases with the number n of random realisations of spatially distributed input $Z(\mathbf{x})$. For all analytical test cases $\mathcal{F}_1$ to $\mathcal{F}_3$ and all sensitivity indices, the standard deviation of estimates over 100 replicas is less than 0.05 when $n \geq 50$. Besides, it can be noted that when n is to small ($n < 50$), sensitivity indices estimates are biased: there seems to be a trend of under-estimation of S_Z and ST_Z, while S_{U_1}, ST_{U_1}, S_{U_2} and ST_{U_2} are over-estimated, compared to reference values obtained with the "*dimension reduction with grouping*" approach. This bias gets smaller for larger values of n.

3.2.4 Comparing SRS and LHS to generate random realisations of $Z(\mathbf{x})$

In this subsection, we study a simple point-based and spatially additive analytical test case in which spatially distributed input $Z(\mathbf{x})$ is modelled as a Gaussian Random Field. We compare two geostatistical simulation algorithms $\mathcal{P}_1$ and $\mathcal{P}_2$ to generate random realisations of $Z(\mathbf{x})$: Simple Random Sampling (SRS) and spatial Latin Hypercube Sampling (LHS). Our purpose is to assess whether spatial LHS yield better sensitivity indices estimates (better accuracy, precision or convergence with increasing n) than SRS.

➤ **Note to the reader:** This section was published in the proceedings of the ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Saint-Geours et al. 2010) with the following title: "*Latin Hypercube Sampling of Gaussian random field for Sobol' global sensitivity analysis of models with spatial inputs and scalar output*". Some extra explanations were added to the published material: they are identified by a grey box. Moreover, in order to keep a general consistency of notations throughout this document, the original notations given in the published paper were changed.

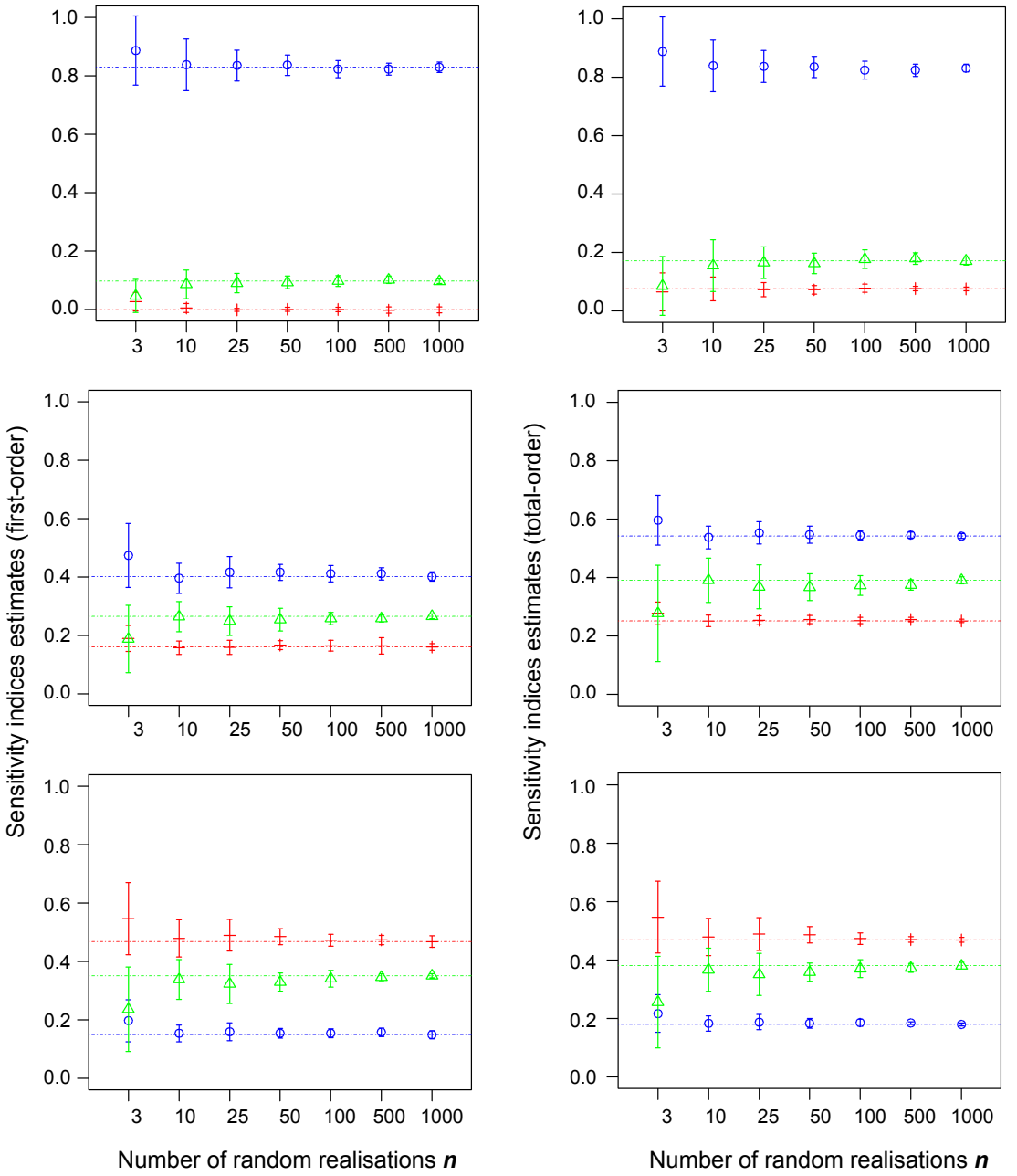

Figure 3.8: Map labelling approach: convergence of sensitivity indices estimates for U_1 ($\circ$), U_2 ($+$) and Z ($\triangle$) with increasing n. Test cases $\mathcal{F}_1$ (top), $\mathcal{F}_2$ (center) and $\mathcal{F}_3$ (bottom)

3.2.4.1 Introduction

Sensitivity analysis (SA) techniques are increasingly recognized as useful tools for the modeller: they allow robustness of model predictions to be checked and help identifying the input factors that account for most of model output variability (Saltelli et al. 2008). Among the various available SA techniques (see Helton and Davis (2006) for a review), variance-based Sobol' global sensitivity analysis (VB-GSA) has several advantages: it explores widely the space of uncertain input factors and is suitable for complex models with non-linear effects and interactions. It can be applied to models with spatial inputs by associating randomly generated map realisations to scalar values (Lilburne and Tarantola 2009). This allows complex description of spatial uncertainty to be used: when model inputs are continuous 2D fields (e.g., a digital elevation model built from some limited terrain points), random map realisations can be generated through geostatistical simulation (Chilès and Delfiner 1999). Yet VB-GSA approach needs a large number of model runs to compute sensitivity indices. With time consuming models, using an effective sampling scheme is necessary to get the most accurate sensitivity indices with the fewest model runs. This issue has been widely discussed in the case of models with scalar inputs. But in many models used for environmental risk assessment (e.g., a flood damage model), inputs are maps (e.g., a water level map) rather than scalars. In such a case, it is also of great importance to generate a relatively small set of map realisations capturing most of the variability of the spatial inputs. Latin Hypercube Sampling (LHS) of Gaussian random fields (Pebesma and Heuvelink 1999) may be a way to reach better efficiency in the computation of sensitivity indices on spatial models. The purpose of this paper is thus to discuss the influence of LHS sampling design when used to generate geostatistical simulations for VB-GSA of models with spatial inputs and a single scalar output. Sensitivity indices are estimated on an artificial model (a simplification of a real flood damage model) with a 2D spatial input (a Gaussian random field), with a) two different sampling designs of geostatistical simulations (Simple Random Sampling and LHS), b) increasing sample size. Relative bias and standard deviation are used to compare exactness and precision of sensitivity indices estimates.

3.2.4.2 A simple spatial model

Flood damage model description Let $Y = \mathcal{F}(\mathbf{U}, Z)$ be a spatial flood damage model with two inputs. Z is a map of the maximal water levels reached during a flood event on a study area. The water levels $Z(\mathbf{x})$ are given at each location $\mathbf{x}$ of the area, represented by a regular grid $\mathcal{G}$ as shown in Figure 3.9 on the following page (total number of pixels: $G = 2\,500$). $\mathbf{U} = (U_1, U_2)$ is a vector of $\mathbb{R}^2$. It describes a linear damage function $\mathcal{F}_{\text{loc}}$: the surfacic damage due to submersion under a water level z is $\mathcal{F}_{\text{loc}}(\mathbf{U}, z) = U_1 \cdot z + U_2$. The model output Y is the total damage due to the flood on the study area [Eqn. (3.10)].

$$Y = \sum_{\mathbf{x} \in \mathcal{G}} \mathcal{F}_{\text{loc}}(\mathbf{U}, Z(\mathbf{x})) \tag{3.10}$$

➤ **Extra comment:** according to the definitions given in §1.1 on page 18, the analytical test case under study is a point-based, spatially additive and linear model.

Uncertainties in input factors The two input factors of model $\mathcal{F}$ are considered uncertain. Map Z is described as a Gaussian random field of mean $\mu = 7$ and variance $\sigma^2 = 121$. Spatial correlation follows an exponential covariance model $C(h)$, with range parameter $a = 10$ and a nugget effect parameter

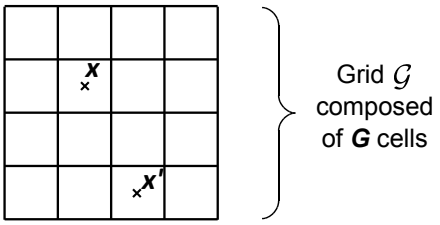

Figure 3.9: Study area represented by a discrete grid $\mathcal{G}$

$\eta = 0.3$ (h being the Euclidian distance between two points) [Eqn. (3.11)]. Parameter values were chosen arbitrarily.

$$C(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ (1-\eta)\cdot\sigma^2 & \forall h > 0 \end{cases} \tag{3.11}$$

The two components U_1 and U_2 of vector $\mathbf{U}$ are independent and follow normal distributions, of respective means $\mu_1 = 6$, $\mu_2 = 1$, and variances $\sigma_1^2 = 16$, $\sigma_2^2 = 1$. The two input factors Z and $\mathbf{U}$ are independent.

3.2.4.3 Spatial global sensitivity analysis: method

Through Sobol' global sensitivity analysis (VB-GSA), we can discuss the relative influence of uncertainty in map Z and uncertainty in $\mathbf{U}$ on the model output variability. The simple form of Y makes it possible to give analytical expression of Sobol' sensitivity indices for each input factor. These exact values are then compared with estimates, which are computed with a sampling-based method, using a set of random geostatistical simulations of map Z.

Analytical expression of sensitivity indices

➤ **Note to the reader:** A proper definition of variance-based sensitivity indices can be found in §1.2.2.3 on page 30.

Sobol' sensitivity indices are based on the decomposition of the output variance in conditional variances. First-order sensitivity index of input factor $\mathbf{U}$ is defined as $\text{var}[\mathbb{E}(Y \mid \mathbf{U})]/\text{var}(Y)$. It represents the main effect contribution of input factor $\mathbf{U}$ to the variance of output Y. For more details on VB-GSA basics, see Saltelli et al. (2008). Let Λ be the average of water levels over the study area, and Γ the average of local damage function $\mathcal{F}_{\text{loc}}$ over the grid:

$$\Lambda = \frac{1}{G}\sum_{\mathbf{x}\in\mathcal{G}} Z(\mathbf{x}) \quad \text{and} \quad \Gamma = \frac{1}{G}\sum_{\mathbf{x}\in\mathcal{G}} \mathcal{F}_{\text{loc}}(\mathbf{U}, Z(\mathbf{x})) \tag{3.12}$$

Λ and Γ are random variables depending on Z and $\mathbf{U}$. Let σ_Λ^2 and σ_Γ^2 be the respective variances of Λ and Γ. Total variance of model output Y is $\text{var}(Y) = G^2\cdot\sigma_\Gamma^2$ where σ_Γ^2 is given by Eqn. (3.13):

$$\sigma_\Gamma^2 = [\mu_1^2\cdot\sigma_\Lambda^2] + [\mu^2\cdot\sigma_1^2 + \sigma_2^2] + [\sigma_1^2\cdot\sigma_\Lambda^2] \tag{3.13}$$

Table 3.3: Sensitivity indices exact values

Sensitivity index	Notation	Value
First-order index of map Z	S_Z	0.309
First-order index of $\mathbf{U}$	$S_\mathbf{U}$	0.554
Total-order index of map Z	ST_Z	0.446
Total-order index of $\mathbf{U}$	$ST_\mathbf{U}$	0.691
Second-order index	$S_{Z,\mathbf{U}}$	0.137

Variance σ^2_Λ depends on the model $C(h)$ of spatial correlation in map Z [Eqn. (3.14)]. $h_{\mathbf{x},\mathbf{x}'}$ is the Euclidian distance between two points $\mathbf{x}$ and $\mathbf{x}'$ on grid $\mathcal{G}$.

$$\sigma^2_\Lambda = \frac{1}{G^2} \sum_{(\mathbf{x},\mathbf{x}')\in\mathcal{G}^2} C(h_{\mathbf{x},\mathbf{x}'}) \tag{3.14}$$

The conditional expectation $\mathbb{E}(Y \mid Z)$s is given by:

$$\sum_{\mathbf{x}\in\mathcal{G}} \mathbb{E}\left[U_1 \cdot Z(\mathbf{x}) + U_2 \mid Z(\mathbf{x})\right] = G \cdot (\mu_1 \cdot \Lambda + \mu_2) \tag{3.15}$$

First-order sensitivity index of Z is then:

$$S_Z = \mu_1^2 \cdot \frac{\sigma^2_\Lambda}{\sigma^2_\Gamma} \tag{3.16}$$

The conditional expectation $\mathbb{E}(Y|\mathbf{U})$ is given by:

$$\sum_{\mathbf{x}\in\mathcal{G}} \mathbb{E}\left[U_1 \cdot Z(\mathbf{x}) + U_2 \mid U_1, U_2\right] = G \cdot (U_1 \cdot \mu + U_2) \tag{3.17}$$

First-order sensitivity index of $\mathbf{U}$ is then:

$$S_\mathbf{U} = \frac{\mu^2 \cdot \sigma_1^2 + \sigma_2^2}{\sigma^2_\Gamma} \tag{3.18}$$

Interactions between input factors Z and $\mathbf{U}$ are accounted for by the second order sensitivity index $S_{\mathbf{U},Z}$:

$$S_{\mathbf{U},Z} = 1 - S_Z - S_\mathbf{U} = \sigma_1^2 \cdot \frac{\sigma^2_\Lambda}{\sigma^2_\Gamma} \tag{3.19}$$

Total-order sensitivity indices account for the total contribution to Y variation due to an input factor. In this case of a model with two input factors, total-order sensitivity index ST_Z is simply the sum of first-order index S_Z and second order index $S_{\mathbf{U},Z}$ (and accordingly for $ST_\mathbf{U}$). Table 3.3 gives the exact values for first-order, second order and total-order sensitivity indices, derived from Eqn. (3.11) on the facing page, Eqn. (3.13) on the preceding page, Eqn. (3.14), Eqn. (3.16), Eqn. (3.18) and Eqn. (3.19).

Generating map realisations In order to estimate sensitivity indices, a set of n random realisations of Gaussian field Z must be sampled. Two methods are considered to generate this set: Simple Random Sampling ("*SRS set*") and Latin Hypercube Sampling ("*LHS set*"). First a SRS set is generated using LU decomposition of the covariance matrix (Chilès and Delfiner 1999). From this set, following the procedure described in Pebesma and Heuvelink (1999), a LHS set of maps is drawn (Figure 3.10 on the following page). This procedure works by repeating the following steps at each location $\mathbf{x}$ of grid $\mathcal{G}$:

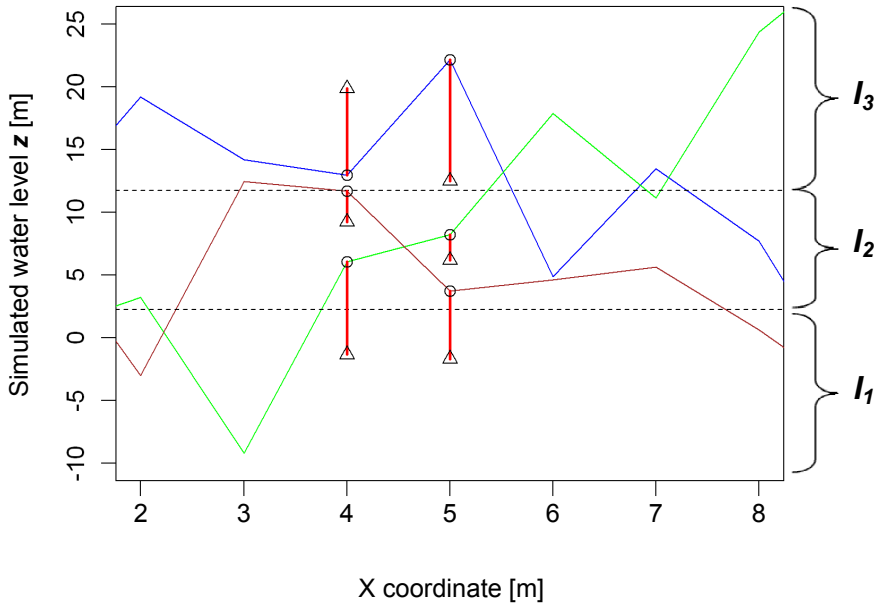

Figure 3.10: Three simulations of Gaussian random field Z by Simple Random Sampling along a line of grid $\mathcal{G}$. Water levels simulated by Simple Random Sampling (∘) and Latin Hypercube Sampling (△) are given for two locations $\mathbf{x}$ et $\mathbf{x}'$. Vertical lines indicate the shift for individual sample elements; dotted horizontal lines indicate stratum boundaries.

- Let $s(\mathbf{x})$ be the vector of the n sampled values at location $\mathbf{x}$ from the SRS set: $s(\mathbf{x}) = (s_1(\mathbf{x}), \ldots, s_n(\mathbf{x}))$.
- Let $r(\mathbf{x})$ be the vector with the ranks of $s(\mathbf{x})$: $r_i(\mathbf{x})$ is the rank of $s_i(\mathbf{x})$ in the ordered list of $(s_j(\mathbf{x}))$. $r(\mathbf{x})$ is a permutation of $\{1; \ldots; n\}$.
- Let F^{-1} be the inverse marginal distribution of $\mathcal{N}(\mu, \sigma)$. Divide the pdf of $Z(\mathbf{x})$ into n equally probable strata I_i according to Eqn. (3.20):

$$I_i = \left[F^{-1}\left(\frac{i-1}{n}\right); F^{-1}\left(\frac{i}{n}\right) \right] \text{ for } i \in \{1, \ldots, n\} \tag{3.20}$$

- The new value $z_i(\mathbf{x})$ of the i^{th} simulation at point $\mathbf{x}$ is obtained by randomly sampling a value in $\mathcal{U}\left(I_{r_i(\mathbf{x})}\right)$.

At each location $\mathbf{x}$, the ranking of the n simulations from the SRS set is preserved in the LHS set: a spatial correlation is thus maintained in each realisation of map Z (Figure 3.10).

Estimating sensitivity indices First-order and total-order sensitivity indices are estimated using "*map labelling*" approach (Lilburne and Tarantola 2009), which is a generalisation of the methods of Sobol' and Saltelli to spatially dependent models. It uses two quasi-random samples of size N, combined through several permutations, to explore the uncertainty domain of input factors Z and $\mathbf{U}$ and estimate sensitivity indices. Spatial input Z is handled by sampling $2 \cdot N$ scalar values from a discrete uniform distribution in $\{1; \ldots; n\}$: each discrete level is associated with a single simulation of Z from the set of n maps previously generated. Input factor $\mathbf{U}$ is treated as a "*group of factors*"; components U_1 and U_2 are sampled

Table 3.4: Exactness (relative bias in %) and precision (± standard deviation) of sensitivity indices, for different sizes n of the set of generated maps

	n = 10 generated maps		n = 250 generated maps	
Index	SRS	LHS	SRS	LHS
S_Z	-7.71% ± 0.097	-11.51% ± 0.080	+1.12% ± 0.047	+1.05% ± 0.042
$S_{\mathbf{U}}$	+9.26% ± 0.126	+11.85% ± 0.107	+2.12% ± 0.043	+1.80% ± 0.038
$S_{Z,\mathbf{U}}$	-17.13% ± 0.062	-19.30% ± 0.061	-8.06% ± 0.049	-6.60% ± 0.049
ST_Z	-9.81% ± 0.124	-13.10% ± 0.105	-1.23% ± 0.042	-0.70% ± 0.038
$ST_{\mathbf{U}}$	+4.02% ± 0.096	+5.66% ± 0.079	+0.10% ± 0.047	+0.13% ± 0.042

independently from their pdf, but sensitivity indices are estimated globally for the group $\mathbf{U} = (U_1, U_2)$. Total number of model runs is $N_{\text{tot}} = 2 \cdot N \cdot (K+1)$ where N is the size of the quasi-random samples and K is the number of (groups of) input factors. Here N is fixed (N = 512), K = 2 and N_{tot} = 3 072.

Sensitivity indices estimates are computed using either SRS or LHS set of maps, for an increasing number n of generated maps. The whole process is replicated 1 000 times. Mean values with ± standard deviation bars for each estimate are shown on Figure 3.11 on the next page. while Table 3.4. gives for each estimate its standard deviation over the 1 000 replicas, and the relative bias compared to its analytical value, for n = 10 and n = 250.

3.2.4.4 Results and discussion

Exactness and precision of sensitivity indices estimates Figure 3.11 on the next page and Table 3.4 show that for small n, S_Z estimate has a negative bias, while $S_{\mathbf{U}}$ has a positive one. When n is low, the small set of map simulations fails to capture the overall variability of Gaussian random field Z: thus the influence of Z variability on model output Y is underestimated; conversely the influence of $\mathbf{U}$ is overestimated. This bias decreases when n increases. LHS sampling of Gaussian random field Z doesn't bring improvement to estimates bias. For small n ($n \leq 10$), LHS estimates have an additional bias which will be discussed in §3.2.4.4. For larger sets of simulated maps ($n \geq 25$), LHS procedure yields estimates whose relative biases are not significantly different from SRS estimates (significance tested with a Welch's t-test for each value of n).

Table 3.4 also shows that LHS estimates have a slightly smaller standard deviation than SRS estimates: this gain is significant for many (S_i, n) couples (Levene's test). This finding is consistent with more general results on variance reduction associated with LHS, in the case of sampling of scalar random variables (Helton and Davis 2003).

Disturbance of spatial correlation by LHS For small number n of generated maps (n = 3, 10), estimates computed with an LHS set of maps have an additional bias (underestimation of S_Z, overestimation of $S_{\mathbf{U}}$) compared to SRS estimates. This additional bias can be explained by the fact that spatial correlation is disturbed by the LHS procedure. Figure 3.12 on page 109 shows that maps from a LHS set have smaller spatial correlations than those from a SRS set, as discussed in Pebesma and Heuvelink (1999). But spatial correlation in map Z influences the value of σ^2_Λ and thus the values of sensitivity indices. Eqn. (3.11) on page 104 and Eqn. (3.14) on page 105 show that σ^2_Λ decreases when spatial cor-

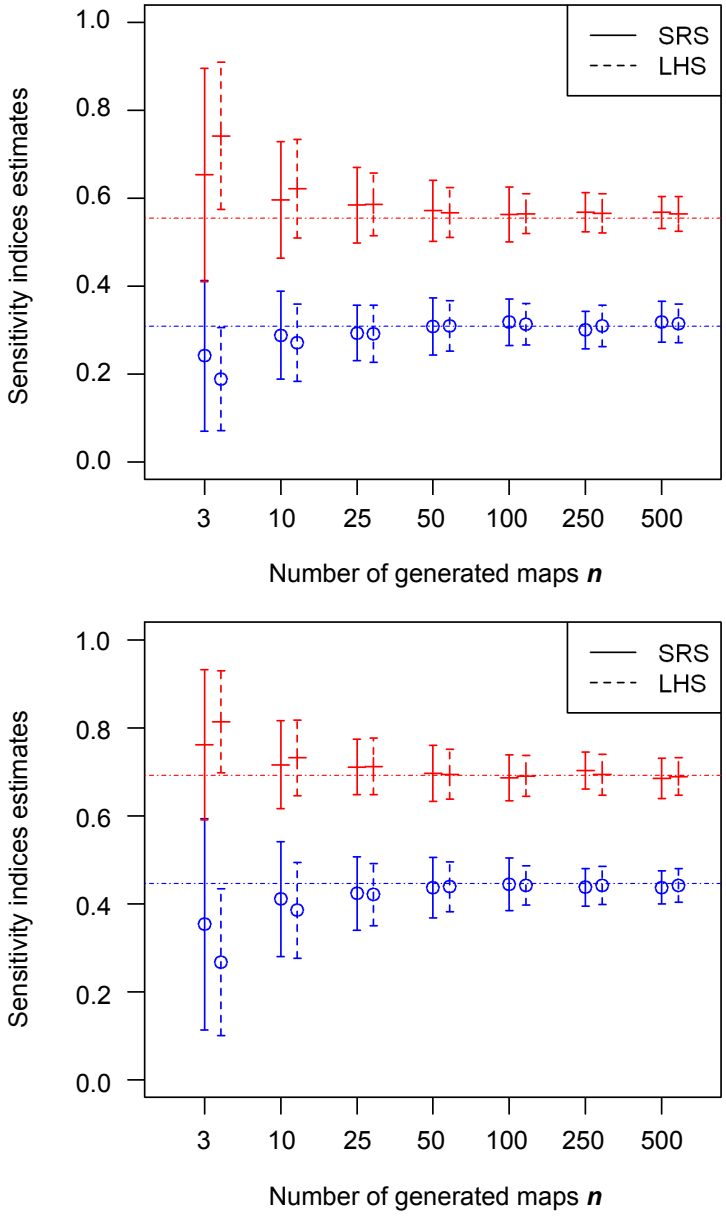

Figure 3.11: First-order (top) and total-order (bottom) sensitivity indices of input factors Z (∘) and $\mathbf{U}$ (+), depending on number n of generated maps and sampling strategy (SRS, LHS). Mean values with ±s.d interval over 1 000 replicas. The dashed lines show the analytical results from Table 3.3 on page 105.

Table 3.5: Exactness (relative bias in %) of the estimates of E_Λ (expectation of Λ) and σ^2_Λ, using either LHS or SRS simulation strategy, for different number n of generated maps

	n = 10		n = 100		n = 200	
Bias	SRS	LHS	SRS	LHS	SRS	LHS
$\epsilon(\hat{E}_\Lambda)$	-0.20[a]	-0.014	-0.54	-2.10^{-3}	-0.13	-4.10^{-5}
$\epsilon(\hat{\sigma}^2_\Lambda)$	-0.74	-0.28	-1.10	0.46	-0.59	-0.21

[a]. Mean value of relative bias over 100 replicas

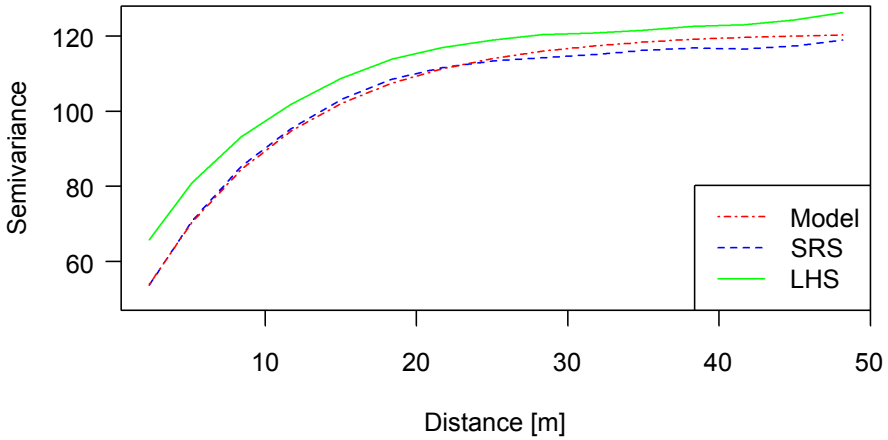

Figure 3.12: Average semivariograms for exponential model, SRS set of maps and LHS set of maps, n = 10 maps.

relation in map Z decreases (smaller range parameter a). This results in an additional underestimation of sensitivity index S_Z when estimated with a LHS set of maps.

Discussion LHS sampling of Gaussian random field Z yields some improvement to the variability of sensitivity indices estimates but no significant improvement to estimates bias. These results can be explained by a general property of LHS: the more the target quantity (here the sensitivity indices) is additive in the variables sampled, the more LHS improves on SRS (Pebesma and Heuvelink 1999). Here the values of sensitivity indices depend heavily on the variance σ^2_Λ of Λ, the average water level over the study area [Eqn. (3.14) on page 105, Eqn. (3.16) on page 105 and Eqn. (3.18) on page 105]. But σ^2_Λ is not additive in sampled water levels $Z(\mathbf{x})$: as a result, the efficiency gain brought by LHS procedure is small.

As an illustration, Table 3.5. gives relative bias of estimates of expectation E_Λ and variance σ^2_Λ, computed on SRS and LHS sets of maps. LHS brings a tremendous gain in estimate bias for the expectation E_Λ (additive in sampled water levels).

3.2.4.5 Conclusion

Sobol' sensitivity indices were estimated on an artificial spatial model (derived from a complex model for flood risk economic assessment) with a 2D spatial input (a Gaussian random field), and compared to their analytical values. Two sampling strategies were used to generate realisations of input Gaussian random field: Simple Random Sampling and Latin Hypercube Sampling (higher CPU cost). Results show that

(1) LHS sensitivity indices estimates have a significantly smaller variance (2) LHS sampling brings no significant improvement in estimates bias (3) for small sample size, disturbance of spatial correlation by LHS procedure yields an additional bias in estimates. The poor improvement brought by LHS sampling comes from sensitivity indices not being additive in the variables sampled. Other ways should be sought to select input map realisations to perform sensitivity analysis of spatial models. These conclusions would be different if SA was computed "*locally*", i.e. with a spatially distributed output (e.g., a map of damage). In this latter case, the spatial dimension of the problem would be reduced, and we could expect LHS to bring the same improvement as in a nonspatial context.

➤ **Note to the reader:** End of the section that was published in the proceedings of the ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Saint-Geours et al. 2010) with the following title: "*Latin Hypercube Sampling of Gaussian random field for Sobol' global sensitivity analysis of models with spatial inputs and scalar output*".

3.2.5 Discussion

This section aimed at exploring the issue of sampling random realisations of a spatially distributed input $Z(\mathbf{x})$ for VB-GSA.

First, the numerical study performed on the analytical tests cases $\mathcal{F}_1$ to $\mathcal{F}_3$ bring some insights on the impact of sampling size n on the convergence of sensitivity indices estimates in the "*map labelling*" approach. On these test cases, where the spatially distributed input is generated by function $\phi_{\text{camp.}}$, we found that a number $n \geq 50$ of random realisations of $Z(\mathbf{x})$ is enough to get accurate estimates of sensitivity indices. We also observed a trend for under-estimation of sensitivity index S_Z for small values of n. A possible explanation is the following: for small values of n the set of random realisations of $Z(\mathbf{x})$ is too small to represent reasonably well the overall uncertainty on the spatially distributed input, and its influence on the variance of model output Y is under-estimated. These results are corroborated by the numerical study performed for the *Accuracy2010* conference, in which the spatially distributed input $Z(\mathbf{x})$ is a Gaussian Random Field. Unfortunately, these conclusions cannot be generalised immediately to other test cases or real-world numerical models. Hence, figures showing the convergence of sensitivity indices estimates with increasing number n of random realisations of $Z(\mathbf{x})$ (e.g., Figure 3.11 on page 108) should be drawn for each case study to determine the minimum sampling size n.

We also partly investigated the issue of choosing an efficient stochastic process $\mathcal{P}$ to generate random realisations of the spatially distributed input $Z(\mathbf{x})$ in the "*map labelling*" approach. In the specific case in which $Z(\mathbf{x})$ is a Gaussian Random Field, we compared the performances of two different sampling techniques, LHS and SRS. The results show that even if LHS brings small improvement on sensitivity indices estimates, it also disturbs spatial correlation for small values of n, resulting in an additional bias in sensitivity indices estimates. This study could have been completed by considering other sampling methods to generate random realisations of Gaussian Random Fields, such as "*sequential spatial simulation using LHS*" developed by Kyriakidis (2005). Another path of research is opened by the work of Scheidt and Caers (2009) who suggest to draw optimal samples of random realisations of $Z(\mathbf{x})$ based on the definition of a distance between realisations. Unfortunately, time was too short to explore these points more deeply.

3.3 Application to the NOE code on the Orb Delta case study

In this section, we carry out a first variance-based sensitivity analysis of the NOE code on the Orb Delta case study. As mentioned in Chapter 2 (§2.3.2.3 on page 64), two groups of spatially distributed inputs must be handled: the assets map (2D vector data) and the hazard maps (2D raster data). To cope with these inputs in VB-GSA of the NOE code, we chose a method based on the NOE characteristics described in §2.4 on page 72 and on the decision-tree given in Figure 3.7 on page 96. First, a run of the NOE code on the Orb Delta case study only lasts 30 seconds on a computer of average performance: intensive simulation is thus possible. Second, the computational burden related to the simulation and storage of spatially distributed model inputs is rather high but not limitative—the weight of a set of water depth maps $\mathcal{H}(e_j)$ is approximately 500 Mo and that of the assets map is 50 Mo. Third, both the number of objects in the assets map ($\sim 20\,000$) and the dimension of water depth maps ($3\,500 \times 3\,500$) are large: hence, the "*macro-parameter*" approach appeared to be intractable. Fourth, reducing the total information contained in the hazard and assets map was impossible, so the "*dimension reduction*" method was discarded. Fifth, there are more than one single spatially distributed input in the NOE modelling framework, which makes the "*joint meta-modelling*" approach developed by Iooss and Ribatet (2009) inappropriate. Finally, we chose to use the "*map labelling*" approach to perform VB-GSA of the NOE modelling framework.

We only paid attention in this analysis to the NOE scalar outputs aggregated over the entire floodplain: total flood damages $D(e_j)$, AAD, AAD', ΔAAD, and NPV indicators. Spatially distributed outputs will be considered in Chapter 4.

➤ **Note to the reader:** The following paragraphs §3.3.1 to §3.3.6 on pages 111–130 are extracted from a draft paper submitted to the *Journal of Flood Risk Management* in July 2012 with the following title: "*Ranking sources of uncertainty in flood damage modelling: a case study on the cost-benefit analysis of a flood mitigation project in the Orb Delta, France*". Some extra explanations were added to the submitted material: they are identified by a grey box. In addition, some sections of this paper are developed into more details in Chapter 1 or Chapter 2 and can safely be skipped by the reader: a specific note is added on top of them.

3.3.1 Introduction

Flooding is recognised as one of the most damaging natural hazards, responsible for approximately one-third of the total economic losses due to natural hazards in Europe (EEA et al. 2008). Following the approval of the EU flood directive (2007/60/EC) in 2007, EU member states now have to establish flood risk management plans focused on prevention, protection and preparedness in all flood-prone river basins and coastal areas. To evaluate these plans, flood risk managers are advised to use cost-benefit analysis (CBA) as part of the appraisal (European Commission 2008). In France since 2011, using CBA is mandatory for local managers who claim national subsidies. Despite known limitations (European Commission 2009a), CBA is a useful tool that provides significant rational information to the decision makers. To consider the expected benefits related to flood management, CBA requires an accurate estimate of the amount of flood damage that will be reduced yearly by the appraised measures. We will use the term CBA-AD to refer to this implementation of CBA based on avoided damage. This estimate relies on a complex modelling chain, including hydrological, hydraulic and economic modelling as well as GIS-based spatial analysis (Messner et al. 2007). Two output indicators are commonly produced in such studies: the

reduction of the expected annual flood damage costs (Arnell 1989) and the net present value of the appraised measures (Erdlenbruch et al. 2008). These two indicators may also take into account the benefits and costs that are not directly related to flood management, such as environmental impacts or landscape modification, but this issue is not in the scope of the present paper.

Meanwhile, there is a growing consensus (Apel et al. 2004) that flood damage assessments are fraught with uncertainties, which arise from innacurate or missing data, model assumptions, measurement errors, incomplete knowledge, etc.—see Refsgaard et al. (2007) and Walker et al. (2003) for an enlightening discussion on the nature of uncertainty. Uncertainty analysis is thus required to identify and quantify the impacts of uncertainties in the modelling chain to i) increase the reliability of flood damage assessments and related CBA-AD (Mostert and Junier 2009) and ii) inform relevant stakeholders with the best information possible for decision making (Ascough et al. 2008). Over the last few years, various methods of quantitative uncertainty analysis have been used in flood damage assessment research (Pappenberger et al. 2006; Pappenberger and Beven 2006). Nevertheless, many authors first focused on the uncertainty in a single component of the flood damage assessment chain: hydraulic modelling (e.g., Bernardara et al. (2010); Gouldby et al. (2010); de Rocquigny (2010), inundation mapping (Bales and Wagner 2009; Stephens et al. 2012), damage functions (Kutschera 2009; Merz et al. 2004; Merz and Thieken 2009; Merz et al. 2010) or land use (Te Linde et al. 2011). To go further, a number of recent studies investigated how combinations of these uncertainty sources interact and propagate through flood damage assessments. They differ by the components under study (extreme value statistics, hydraulic model, potential dyke breach, inundation mapping, exposure assessment, damage functions) and by the uncertainty analysis method used. In some studies (Koivumäki et al. 2010; Merz and Thieken 2009; de Moel and Aerts 2011; Briant 2001), the various components of the modelling chain were varied manually in a "*one-factor-at-a-time*" (OAT) approach (Saltelli et al. 2008) to estimate the confidence bounds around the flood damage estimates. Other authors described uncertainty sources in a probabilistic setting and explored the space of input uncertainty within a Monte Carlo framework (Helton and Davis 2006), which requires a large number of model evaluations (Apel et al. 2008; de Kort and Booij 2007; de Moel et al. 2012; Weichel et al. 2007).

Another related but distinct issue is to identify, in the flood damage assessment process, the main sources of uncertainty that contribute the most to the variability of damage estimates and CBA-AD outcomes, which is the role of sensitivity analysis methods (SA). These methods aim to study how the uncertainty of a model output can be apportioned to different sources of uncertainty in its inputs (Saltelli et al. 2008). SA is recognised as an essential component of model building (European Commission 2009a; CREM 2009) and is widely used in different fields (Cariboni et al. 2007; Tarantola et al. 2002). Ranking uncertainty sources, usually by so-called "*sensitivity indices*" or "*importance measures*" is useful to orientate further research, collect additional data on most influential inputs but also simplify the model under study by fixing non-influential inputs. While many quantitative SA approaches are available, most studies in the field of flood damage assessment used a "*one-at-a-time*" and qualitative SA approach, manually comparing the separate effects of each uncertainty source on the damage estimates. They used the large number of model evaluations produced from uncertainty analysis either in a probabilistic setting or using various versions of input data (Apel et al. 2004; Koivumäki et al. 2010; de Moel and Aerts 2011; Pappenberger et al. 2008). To our knowledge, only the work of de Moel et al. (2012) was based on a quantitative global sensitivity analysis method (GSA), in which i) quantitative sensitivity indices are estimated for each uncertainty source and ii) all uncertain model inputs are varied at the same time, which allows the effect of their interactions on the overall output variability to be discussed.

Nevertheless, to date, only the uncertainty on the flood damage assessments have been studied without questioning how this uncertainty may impact the robustness of the CBA-AD of flood management policies nor how to improve this robustness. Our paper is an attempt in this direction. We try to answer

the following questions: how does uncertainty propagate through the CBA-AD of a flood management policy? What is the ranking of the uncertainty sources in such a CBA-AD?

We discuss these questions through a case study of a CBA-AD applied to a flood risk management plan on the Orb River delta, France, where only structural flood-control measures are considered. Our goal is to check the robustness of the CBA-AD results and assess the contribution of uncertainty sources to the overall output variability. A modelling chain named NOE (§3.3.2 on the following page) is used to estimate the potential flood damages at the scale of individual assets and perform a cost-benefit analysis of the flood control measures. Uncertainty sources are then described in a probabilistic framework, propagated through the NOE model with pseudo Monte Carlo simulations, and variance-based sensitivity indices are computed for each of them (§3.3.3 on page 117). Only epistemic uncertainty (Refsgaard et al. 2007) is considered here because aleatory uncertainty is already accounted for in the definition of output CBA-AD indicators (average annual flood damages). The description of uncertainty sources is based on the literature, expert opinion or measurements using univariate or bivariate probability density functions or more complex models for spatially distributed uncertainty. The results (§3.3.4 on page 125) include confidence bounds and empirical pdf of CBA-AD outputs and a ranking of the uncertainty sources based on their sensitivity indices. We discuss the outcomes of our approach and its limits in §3.3.5 on page 128.

3.3.1.1 Study site

✄ **Note to the reader:** The Orb Delta study site is already presented in §2.3.2.1 on page 59 and the following paragraph can safely be skipped.

As a study area, we selected the Lower Orb River fluvial plain, known as the Orb Delta, located in the south of France. We focused on a 15 km reach from Béziers to the Mediterranean sea that is bounded by an area of 63 sq. km and includes the cities of Béziers, Portiragnes, Sauvian, Sérignan, Valras-Plage and Villeneuve-lès-Béziers (Figure 2.15 on page 59). The Orb catchment has a typical Mediterranean subhumid regime. The annual maximum discharge in Béziers (Tabarka gauge) varies from year to year between 100 and 1 500 m^3/s (BCEOM 2000). The flood prone area in the Orb delta is home to approximately 16 290 permanent people (total population of the six localities: 90 000 people), 774 companies and 30 seaside campgrounds (which attract up to 100 000 tourists in summertime). Approximately one-third of the area is devoted to agriculture. The flood of December 1995 - January 1996, with a peak discharge of 1 700 m^3/s at the Tabarka gauge, caused a total amount of damage of 53 M€ (SMVOL 2011).

In 2001, local authorities launched a flood risk management project, mainly based on various structural mitigation measures, including dyke strengthening around urban areas, restoration of sea outfalls and channel improvement. In 2011, to claim national subsidies, they completed a cost-benefit analysis of their project (Grelot et al. 2012).

This study site was mainly chosen because it was a "*real*" case study, with a flood risk management plan under construction and a cost-benefit analysis produced by the local authorities. Moreover, the area was already documented with numerous available data. These data included aerial photographs, a high-resolution Digital Terrain Model (DTM) built from photogrammetry, the annual maximum flow series from 1967 to 2009 at the Tabarka gauge, and various spatial datasets on buildings, agricultural land and economic activities in the area (Erdlenbruch et al. 2008).

3.3.2 Description of the NOE modelling framework

✂ **Note to the reader:** The flowchart of the NOE modelling framework and its application on the Orb Delta case study are presented into more details in Chapter 2. This section is a simplified and shorten version of §2.2 (for the purpose of publication) and it can safely be skipped.

Cost-benefit analysis based on avoided damages (CBA-AD) was used to evaluate the flood risk management project launched on the Orb River. In the literature, flood damage assessments and related CBA-AD vary in their scope and scale as well as the data used and their outputs. Here, a complex modelling chain named NOE (Erdlenbruch et al. 2008) combines hydrological, hydraulic, GIS and economic modelling to estimate the flood damages on individual assets and compute two output indicators: i) the Average Annual Avoided Damage (ΔAAD [€/year]) over the study area, which is defined as the amount of annual expected damage costs that are reduced due to the flood mitigation measures; and ii) the Net Present Value (NPV [€]) of the flood mitigation measures. The modelling chain consists of seven steps that are further described in this section (Figure 3.13).

3.3.2.1 Flood scenarios

The calculation of the Average Annual Avoided Damage requires damage estimation for a number of relevant flood scenarios with different characteristics to represent the aleatory uncertainty associated with flood hazard in the study area. The first step of the NOE modelling chain is thus to choose a range of potential flood events of various magnitudes. Six flood scenarios e_1 to e_6 were selected, characterised by a maximum discharge q_i at Tabarka gauge (Table 2.1 on page 61). e_1 is supposed to be the smallest flood event that induces damage (q_1 = 1 000 m^3/s).

Scenarios e_2 to e_5 include historical floods and design floods. Scenario e_6 is an extreme flood, which would result in an over-topping of all existing flood-control dykes.

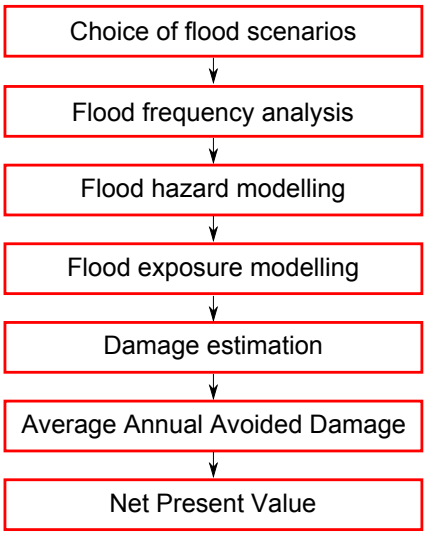

Figure 3.13: Simplified flowchart of the NOE model

3.3.2.2 Flood frequency analysis

The return intervals T_1 to T_6 and exceedance frequencies f_1 to f_6 associated with flood scenarios e_1 to e_6 were deduced (Table 2.1 on page 61) from the discharge-frequency Gumbel curve ($Q - f$), which was fitted on the annual maximum flow series at the Tabarka gauging station available from 1967 to 2009 (AMFS 1967-2009).

3.3.2.3 Flood hazard modelling

For each flood scenario, a typical flood hydrograph was first generated based on expert opinion (BCEOM 2000). The 1D, step-backwater hydraulic ISIS Flow model (unsteady flow) was then used to propagate the hydrographs within the floodplain. The ISIS Flow model solves the full 2D, depth-averaged momentum and continuity equations for free-surface flow (ISIS 2012). Two different floodplain flow simulations were produced for each flood scenario e_2 to e_6: one describing the present situation and the other describing future situation with enforced mitigation measures. The floodplain flow simulations were then combined with a high-resolution DTM, to produce two sets of raster data with a 5 m cell size, giving spatially explicit maximum water depths (WD) for each flood event over the study area: $\mathcal{H}(e_1)$ to $\mathcal{H}(e_6)$ (present situation) and $\mathcal{H}(e'_1)$ to $\mathcal{H}(e'_6)$ (future situation).

3.3.2.4 Flood exposure modelling

Four economic sectors were considered in the exposure analysis: private housing, agricultural land, campgrounds and other economic activities. Flood exposure was assessed at the scale of small individual assets (buildings, plots of cultivated land, etc.). Data from various sources were collected to build a land use geo-database (LU-GDB) over the study area, including digital cadastral maps, a dataset of the regional Chamber of Commerce and Industry (2009), and the national agricultural land use statistics (RPG dataset, 2009). An extensive field survey was also conducted to collect additional data on assets, such as ground floor elevation of buildings. In the end, the LU-GDB dataset describes private housing units (individual buildings), plots of cultivated land, campgrounds and other economic activities by individual polygonal features in a single GIS vector layer (Table 2.2 on page 62). Plots of cultivated land were further characterised by a subtype (wheat, vineyard, etc.), while economic activities were classified into sixty categories following the French classification of economic activities NAF2008 (INSEE 2008).

The flood exposure of assets was then assessed by confronting the LU-GDB dataset with water depth maps $\mathcal{H}(e_1)$ to $\mathcal{H}(e_6)$ and $\mathcal{H}(e'_1)$ to $\mathcal{H}(e'_6)$. For each exposed object (represented by a polygonal feature in LU-GDB dataset) and each inundation map, the average water depth over the object was extracted as an attribute column by a simple overlay analysis. To compute meaningful average water depths for very large objects (e.g., large plots of cultivated land), we first divided all features into pieces of $40\,000$ sq. m max by intersecting features of the LU-GDB dataset with a regular square grid of 200 m cell size.

3.3.2.5 Damage estimation

The following module of the NOE modelling chain estimates the total damage costs (D) within the study area for each flood scenario e_1 to e_6, for the present ($D(e_1)$ to $D(e_6)$) and future ($D(e'_1)$ to $D(e'_6)$) situations. We will denote by ΔD_1 to ΔD_6 the damage reduction brought by the mitigation measures for each flood scenario: $\Delta D_i = D(e_i) - D(e'_i)$. As scenario e_1 was defined as the "*flood event where damage*

to property begins", the damage estimates $D(e_1)$ and $D(e'_1)$ are both assumed to be equal to zero. For scenarios e_2 to e_6, as a coarse estimation, only direct and tangible monetary losses were considered—Merz et al. (2010) present the other types of damages that should be estimated for a more complete analysis.

Damage functions were used (Table 2.3 on page 65), which depend mainly on the following parameters: type and floor surface area of the exposed object, average water depth, and season of occurrence (campgrounds and agriculture). Flood velocity and flood duration were considered to be homogeneous. These damage functions were taken from the recommendations of the French State (MEDDTL 2011). For a complete description, see the original study (Grelot et al. 2012). In the end, a total of 94 depth-damage relationships were used, one for each land use type and subtype.

3.3.2.6 Average Annual Avoided Damages

The average annual damage cost from flooding (AAD [€/year]) is a common performance indicator used to measure potential flood damages over a given territory (Arnell 1989; Messner et al. 2007). It is equal to the area under the damage-frequency curve, which is the graph of damage D against exceedance frequency $f = 1/T$:

$$AAD = \int_0^1 D(f)df \tag{3.21}$$

To assess the benefits of the flood risk management project launched on the Orb River in 2001, we computed the potential reduction of the average annual damage costs brought by the mitigation measures, i.e., the variation $\Delta AAD = AAD - AAD'$ from the present to the future situation. This Average Annual Avoided Damage (ΔAAD [€/year]) is also equal to:

$$\Delta AAD = \int_0^1 \Delta D(f)df \tag{3.22}$$

It can be computed from the range of flood scenarios e_1 to e_6 and corresponding avoided damages ΔD_1 to ΔD_6 by estimating the integral [Eqn. (3.22)] with a simple trapezoidal rule (Figure 3.14 on the facing page).

➤ **Note to the reader:** For a more general definition of the Average Annual Avoided Damages, see §2.2.5 on page 44.

3.3.2.7 Net Present Value of the mitigation measures

The last step of the NOE modelling chain is the cost-benefit analysis, which evaluates the efficiency of the flood mitigation measures by comparing their costs with their expected benefits. The costs of the mitigation measures include the initial investment (CI = 35.2M€) and maintenance costs (CM = 1.6M€/year). The benefits of the project are measured by the ΔAAD indicator. The Net Present Value (NPV [€]) of the flood mitigation measures is then calculated by comparing the discounted costs and benefits over a period of R = 30 years [Eqn. (3.23)].

$$NPV = -CI + \sum_{i=0}^{R} (\Delta AAD - CM) \cdot \tau_i \tag{3.23}$$

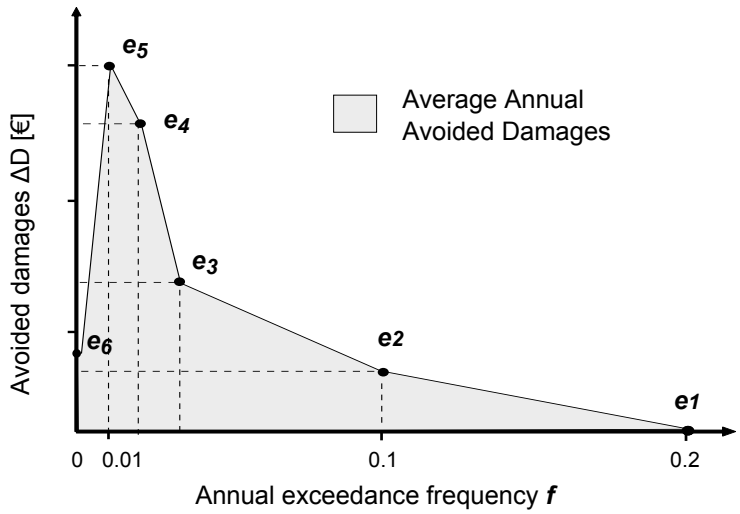

Figure 3.14: Computation of the Average Annual Avoided Damages

where τ_i is the discount coefficient for year i. A positive NPV value means that the benefits generated by the flood risk management project outweigh its costs. The larger the NPV value, the more efficient the flood mitigation measures are.

3.3.3 Uncertainty and sensitivity analysis

Uncertainty and sensitivity analyses of the NOE modelling chain (Figure 3.15 on page 119) were performed using variance-based global sensitivity analysis approach (Saltelli et al. 2008). In the first step of the analysis, sources of uncertainty in the NOE modelling chain were identified and modelled in a probabilistic framework, and a set of random realisations was sampled for each uncertain modelled input. Next, pseudo-Monte Carlo simulations were used to explore the space of input uncertainty and assess the resulting variance of model outputs (ΔAAD and NPV indicators). Finally, sensitivity indices were computed to rank the sources of uncertainty, depending on their contribution to the variance of outputs.

➤ **Sensitivity indices with respect to the NPV indicator:** In addition to the complete sensitivity analysis of the NOE modelling framework that we want to perform, we observed that it was possible to calculate variance-based sensitivity indices analytically in Eqn. (3.23) on the facing page that defines the Net Present Value indicator. The exact formulae of these sensitivity indices are given in Appendix §B on page 203. Besides, we also proved that there is a relation between sensitivity indices with respect to the ΔAAD indicator and the sensitivity indices with respect to the NPV indicator. This point is developed in Appendix §B on page 203.

3.3.3.1 Modelling sources of uncertainty

Table 3.6 on the next page lists the epistemic uncertainty sources that we took into account in the uncertainty and sensitivity analyses of the NOE modelling chain. Each source of uncertainty was modelled in a probabilistic framework using measurement or expert opinion.

Table 3.6: Sources of uncertainty in the NOE modelling chain

Model input	Nature of uncertainty	Modelling uncertainty	Sample size
Hazard maps $\mathcal{H}(e_2)$ to $\mathcal{H}(e_6)$ and $\mathcal{H}(e'_2)$ to $\mathcal{H}(e'_6)$	Errors in hydraulic modelling; errors in DTM	Only DTM error is considered: Gaussian noise without spatial correlation (mean = 0 cm, s.d. 17 cm).	100

LU-GDB dataset	◦ Misclassification of land use types and subtypes
	◦ Variability of ground floor elevation of buildings (lack of data)
	◦ Shape and surface of polygonal features (measurement errors)
		◦ Confusion matrix for land use misclassification.
		◦ Empirical distribution of groundfloor elevation of assets from field survey:
		  ◦ private housing: empirical histograms on five homogeneous zones
		  ◦ campgrounds: U[-0.3m; +0.3m]
		  ◦ plots of cultivated land: U[-0.2m; +0.2m]
		  ◦ other economic activities: U[0.4m; 0.6m]
		◦ Random multiplying coefficient for surface area of buildings: U[0.75; 0.85]
			1 000

Return intervals T_1 to T_6	Lack of data (short time series of annual maximum discharge).	Confidence interval on the fitted $Q - f$ relation	1 000
Depth-damage curves	Lack of data	For each depth-damage curve, random multiplying coefficient ϵ_i drawn from $\mathcal{U}[0.5; 1.5]$. ϵ_i are independent.	1 000
Project costs CI and CM	Lack of knowledge	Based on expert opinion. Triangular distributions with parameters [M€]: $CI(min)$ = 28.6; $CI(mode)$ = 38.6; $CI(max)$ = 35.2; $CM(min)$ = 1.3; $CM(mode)$ = 1.6; $CM(max)$ = 1.8	4 096

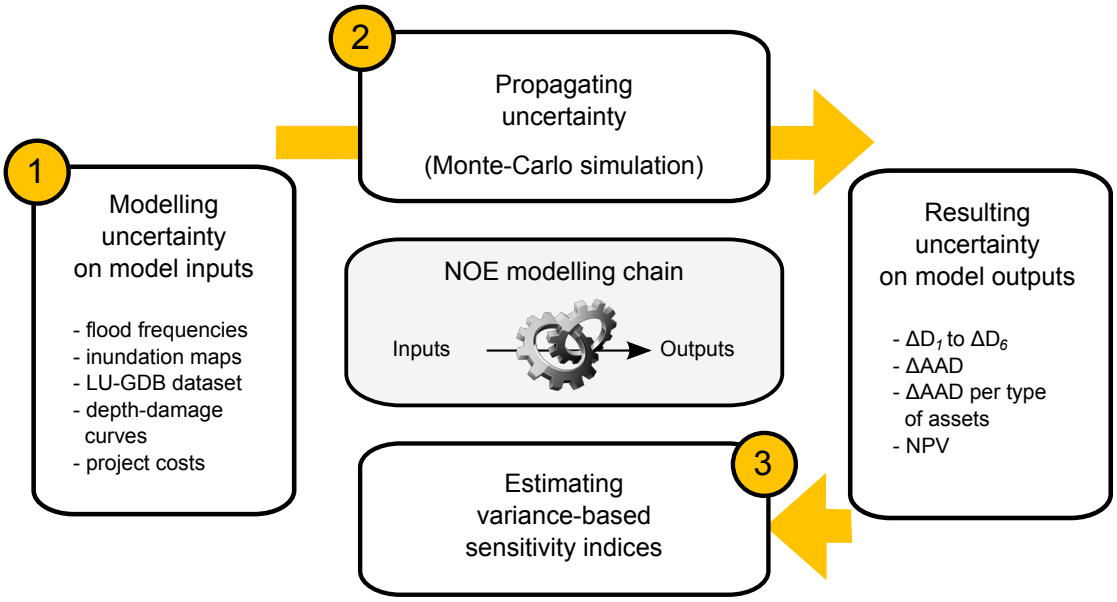

Figure 3.15: Flowchart for uncertainty and sensitivity analysis

Flood frequency analysis Uncertainty in the flood frequency analysis may arise from i) stream gauge measurement errors (Neppel et al. 2010); ii) possible non-stationarity of the series due to climate change (Khaliq et al. 2006) and iii) uncertain fitting of a discharge-frequency ($Q - f$) relationship to the AMFS 1967-2009 dataset (Countryman and Tustison 2008). Here, only the latter uncertainty was modelled and simulated. After a log transformation leading to the usual linear regression context, standard joint distributions (Maidment 1993) for the parameters of the fitted Gumbel curve were calculated (Figure 3.16 on the following page). A set of $n_1 = 10^3$ Gumbel curves was then randomly sampled from the parameter joint distribution. From this set of curves, 10^3 exceedance frequencies f_i and according return intervals $T_i = 1/f_i$ (Figure 3.17 on the next page) were generated for each discharge value $q_i (i = 1 \text{ to } 6)$.

Flood hazard modelling Another major source of uncertainty in the NOE modelling chain is the inundation mapping process, which includes hydraulic modelling and combination with a high resolution DTM, to derive water depth maps $\mathcal{H}(e_1)$ to $\mathcal{H}(e_6)$ and $\mathcal{H}(e'_1)$ to $\mathcal{H}(e'_6)$. For the sake of simplicity, a restrictive choice was assumed in considering the error on the high-resolution Digital Terrain Model as the single uncertainty source in water depth maps. Including more detailed descriptions of the hydraulic uncertainties in this study was impossible as the ISIS hydraulic model used for initial flow simulations was not available to us. This choice may be partly justified by the findings of both Bales and Wagner (2009) and Koivumäki et al. (2010), who investigated the various sources of error encountered in this process and conclude that high-resolution topographic data is the most important factor required for accurate inundation maps. The DTM—a raster data of 5 m cell size—was initially built by stereo-photogrammetry. Both measurement errors and interpolation errors affect the quality of this input data (Wechsler 2007). These errors were modelled by a Gaussian noise without spatial correlation, whose characteristics were determined from a set of 500 control field points (mean = 0 cm, s.d. = 17 cm). A set of $n_2 = 100$ random realisations of the Gaussian random error field was generated and added as "*noise*" to the initial water depth maps $\mathcal{H}(e_1)$ to $\mathcal{H}(e_6)$ and $\mathcal{H}(e'_1)$ to $\mathcal{H}(e'_6)$. We may note that this procedure induces independent variations of the water levels for each exposed assets; it differs from the study of de Moel and Aerts (2011), who described uncertainty in the water levels with a spatially uniform bias.

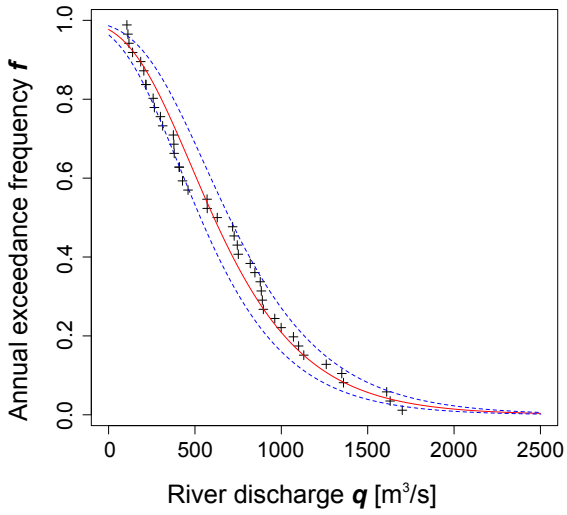

Figure 3.16: Annual maximum flow series AMFS 1967-2009 and fitted Gumbel distribution (red solid line) with 95% confidence bounds (dotted blue lines)

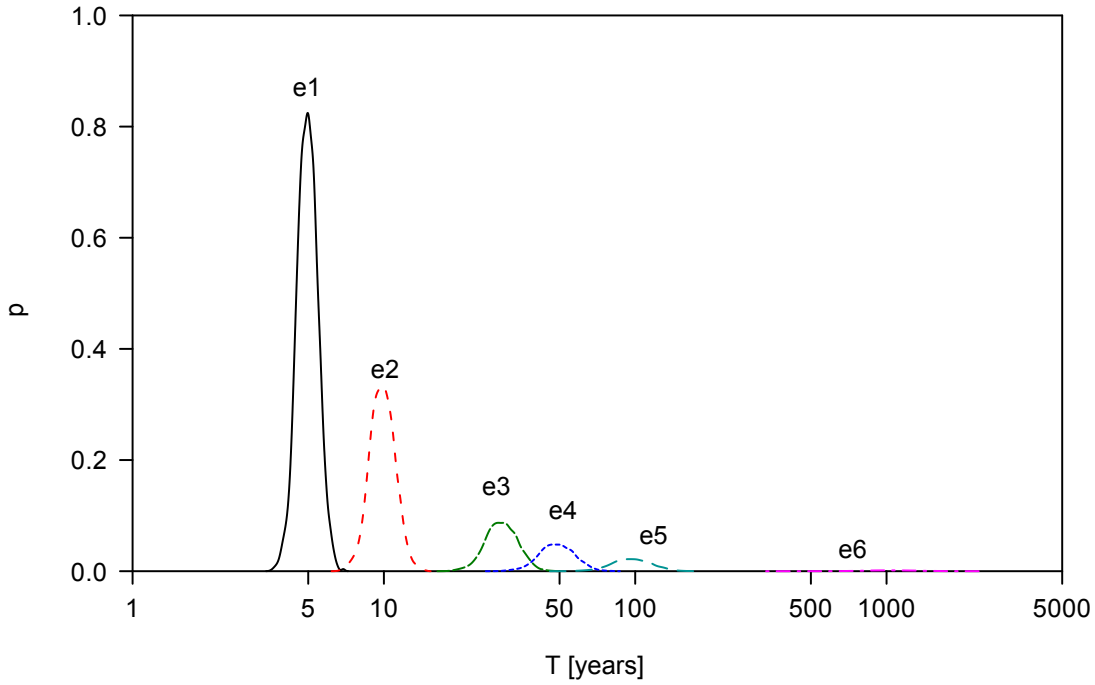

Figure 3.17: Empirical distributions of return intervals T for flood scenarios e_1 to e_6

➤ **Extra comment:** We also investigated how to model and simulate the auto-correlated structure of uncertainty in the DTM. In a previous sensitivity analysis of the NOE modelling chain on the Orb Delta, published in Journal de la Société Française de Statistiques (Saint-Geours et al. 2011a reproduced in Appendices), we modelled the measurement errors and interpolation errors on the initial DTM by a Gaussian Random Field with an exponential variogram model, whose characteristics were determined from a set of 500 control field points.

Flood exposure modelling The third source of uncertainty is the location and attribute data errors in the LU-GDB dataset (Koivumäki et al. 2010). The error in the land use GIS layer may stem from the following: i) misclassification of polygonal features representing assets; ii) error on the ground floor elevation of buildings; and iii) error on the surface area of features. Other sources of uncertainty were identified: geometric errors (Bonin 2006; Girres and Julien 2010), errors in the asset locations, and the evolution of land use over time (Te Linde et al. 2011). Although in some studies (de Moel and Aerts 2011) the uncertainty of the land use data was represented by using a small number of different datasets, here each uncertainty source was modelled in a probabilistic setting. To describe the misclassification of polygonal features, which may arise in the process of photo-interpretation, a confusion matrix (Fisher 1991) was built based on expert opinion, giving a confusion probability $p_{i,j}$ for each pair of land use types (i, j) (Table 3.7). Then, to model the variability of the ground floor elevation of buildings, measurements were taken during a field survey on a sample of 100 buildings. The study area was divided into five homogeneous zones; in each zone, the distribution of ground floor elevation was described by an empirical histogram (Figure 3.18 on page 123). Random ground floor elevation was also attributed to campgrounds, plots of agricultural land and other economic activities (Table 3.6 on page 118). Next, the surface area of the buildings was also randomised, as the features area extracted from cadastral maps differ from the effective surface area of buildings that should be taken into account for flood damage estimation (e.g., wall width should be subtracted). To cope with this issue, the nominal surface of each building was multiplied by a corrective random coefficient drawn independently in a uniform pdf in [0.75; 0.85], considering a digitalising error of 0.3 mm at the map scale (Hengl 2006). Finally, from this probabilistic description of the uncertainty in the LU-GDB dataset (confusion matrix, empirical distribution of ground floor elevations, corrective coefficient for surface areas), a set of $n_3 = 1\,000$ random LU-GDB datasets was sampled.

Table 3.7: Confusion matrix of LU-GDB dataset

Land use type	Number of sub-types	Probability of confusion between sub-types
Private housing	1	No confusion.
Agricultural land	15	25% chance of confusion between durum wheat and bread wheat; 10% chance of confusion between colza, maize, barley and sunflower; 25% chance of confusion between permanent and temporary grassland.
Campgrounds	18	No confusion.
Other economic activities	60	0.17% chance of belonging to each other class of economic activities.

➤ **Extra comment:** As mentioned in §2.2.8.2 on page 56, there are various ways to assess the exposure of assets to a flooding event, for example using the maximum values or the mean values of flood caracteristics (water height, velocity, flood duration) over the object. This modelling choice results in an additional uncertainty, which is not related to the inaccuracy of model inputs. It was not considered in the sensitivity analysis of the NOE modelling framework on the Orb Delta case study. This issue is further discussed in Appendix §D on page 212.

➤ **Extra comment:** The evolution of land use over time was identified as an extra source of uncertainty in flood exposure assessment, but was not taken into account in this study. In the future, a possible way to explore this issue would be to simulate land use changes (with uncertainty) over the next 30 years on the study area. Flood exposure and flood damages would be computed at each time step, and the average ammount of damages over time would be considered. One likely problem would be that the spatial scale of models that simulate land use change (e.g., MOLAND or CLUE-s) may not be commensurate with the scale needed for exposure assessment in flood damage modelling.

Damage estimation The fourth uncertain model input is the set of 94 depth-damage curves (one for each land use type and subtype) used for the damage estimation. Uncertainty about the damage functions has been extensively discussed in previous studies (Koivumäki et al. 2010; Kutschera 2009; Merz et al. 2004, 2010; Merz and Thieken 2009; de Moel and Aerts 2011). In these papers, uncertainty was mainly represented by using two or three different sets of damage functions coming from various studies. Only de Moel and Aerts (2011) used a parametric uncertainty model (beta pdf) derived from Egorova et al. (2008). Here, to treat all uncertainty sources in a probabilistic framework, we made the choice to use a single set of depth-damage curves and represent their uncertainty by a uniform pdf, defining a -50% to $+50\%$ uncertainty range around nominal curves (Figure 3.19 on the next page). Depth-damage curves associated with each land use type and subtype were assumed to vary independently—contrary to de Moel and Aerts (2011), where they were sampled collectively from a single p-value. A total of $n_4 = 1\,000$ random sets of depth-damage relations was sampled this way.

➤ **Extra comment:** The $[-50\%, +50\%]$ uncertainty range was chosen based on expert opinion. Torterotot (1993) studied the uncertainty on depth-damage curves for private building and displayed coefficients of variation associated with average annual avoided damages around 40%. Other authors could choose a much larger range (Merz et al. 2010)—in particular, Merz et al. (2004) quantified the uncertainty which is associated with damage estimates to buildings using statistical information and found that damages could differ by more than one order of magnitude for similar flood intensity parameters.

Project costs Finally, the last source of uncertainty in the NOE modelling chain is related to the costs of the flood risk management project. Based on expert opinion, investment costs CI and maintenance costs CM were assumed to follow a triangular pdf with the parameters shown in Table 3.6 on page 118.

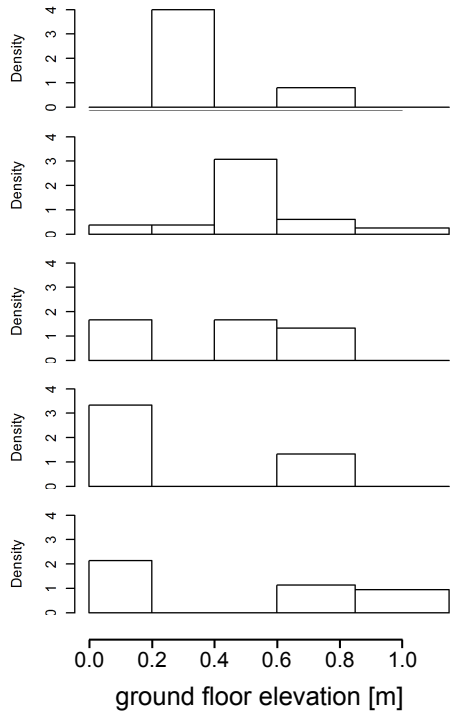

Figure 3.18: Empirical distribution of the elevation of ground floor of buildings over zones A (top) to E (bottom)

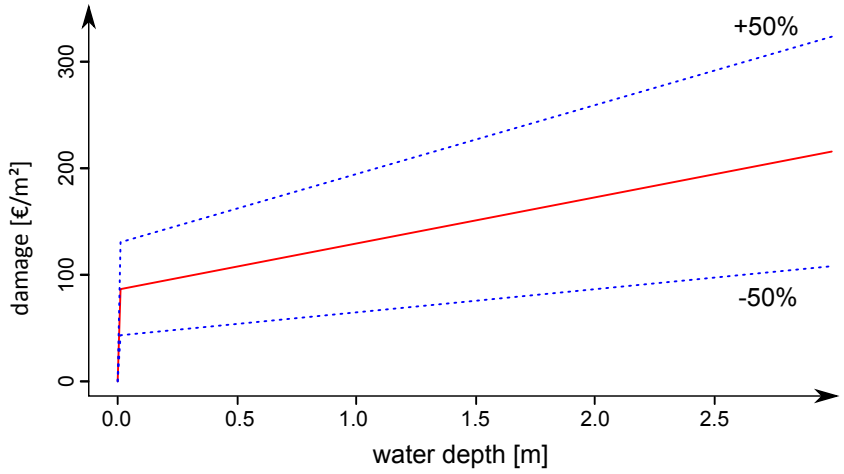

Figure 3.19: Nominal depth-damage curve for private housing (red solid line) with a $[-50\%; +50\%]$ uncertainty range (blue dotted lines)

➤ **Extra comment on discount rate and time horizon:** Two parameters were considered constant is this analysis: the discount rates τ_i and the length of time R over which the Net Present Value of the flood mitigation project is calculated. Both have conventional values that are used in most CBA-AD studies, hence it did not seem appropriate to consider them as uncertain. Yet these conventional values are constantly discussed in the literature devoted to CBA theory—for example, Almansa and Martínez-Paz (2011) discusses the value of discount rates τ_i for the CBA of projects with long term environmental impacts.

3.3.3.2 Propagating uncertainty

✂ **Note to the reader:** The computational aspects of uncertainty propragation in VB-GSA are presented into more details in §1.2.2.5 on page 32 and the following paragraph can safely be skipped.

Once uncertainty had been modelled for each source of uncertainty, it was propagated through the NOE modelling chain using Monte Carlo simulation and a specific sampling scheme following Lilburne and Tarantola (2009). It uses two independent quasi-random LP-τ matrices M_1 and M_2 (Sobol' 1967), here of length $N = 4096$—this sampling size was chosen to fill the necessary conditions for the LP-τ samples (N must be a power of 2) and large enough to obtain a satisfactory level of accuracy for the sensitivity indices estimates. These two matrices were combined through several permutations to explore the uncertainty domain of the five model inputs considered, respectively: exceedance frequencies, inundation maps, LU-GDB dataset, depth-damage curves, and project costs. The i^{th} line of sample M_1 or M_2 is a set $(l_1^{(i)}, l_2^{(i)}, l_3^{(i)}, l_4^{(i)}, l_5^{(i)})$ where each $l_i^{(i)}$ is a random integer label sampled from $1, \ldots, n_j$ associated with a single random realisation of the j^{th} model input (from the set of n_j random realisations that was previously generated). One can note that the number n_j of random realisations is not the same for each model input: these numbers were chosen under the constraints of CPU time and storage space. Next, the NOE modelling chain was run for each line of samples M_1, M_2 and a number of combinations of M_1 and M_2—more details on the procedure can be found in Lilburne and Tarantola (2009). The total number of model runs was $N_{\text{tot}} = 28\,672$ (this number depends on the base sample size N and the number of uncertain model inputs considered) for a total CPU time of 24 hours on a 6-nodes cluster computer.

3.3.3.3 Variance-based sensitivity indices

✂ **Note to the reader:** The mathematical basis and computational aspects of VB-GSA are presented into more details in §1.2.2 on page 29 and the following paragraph can safely be skipped.

Uncertainty propagation results in a set of $N_{\text{tot}} = 28\,672$ values for the following outputs of interest:

- avoided flood damages per scenario (ΔD_1 to ΔD_6)
- Average Annual Avoided Damages (ΔAAD)
- Average Annual Avoided Damages per type of assets
- Net Present Value of mitigation measures (NPV)

Then, the variance-based total-order sensitivity indices of each source of uncertainty with respect to each output of interest were estimated using the expressions given by Lilburne and Tarantola (2009). These sensitivity indices, denoted by ST_i, measure the contribution of a given source of uncertainty, denoted by U_i, and all its interactions with other sources of uncertainty, denoted by $U_{\sim i}$, to the variance of a given model output, denoted by Y:

$$ST_i = \frac{\mathbb{E}\left[\text{var}\left(Y \mid U_{\sim i}\right)\right]}{\text{var}(Y)} \tag{3.24}$$

ST_i is the expected part of output variance var(Y) that would remain if all sources of uncertainty but U_i were fixed. Please refer to Saltelli et al. (2008) for more details on global sensitivity analysis and the estimation of sensitivity indices.

➤ **Extra comment:** Only total-order sensitivity indices were considered in this case study, because estimates of first-order indices displayed too large confidence intervals—these confidence intervals were computed by bootstrap using 100 replicas.

3.3.4 Results

3.3.4.1 Uncertainty analysis

Table 3.8 on the next page summarises the outcome of the uncertainty analysis. For each output of interest, it gives descriptive statistics over the N_{tot} = 28 672 model runs. It displays mean values of avoided damages per flood scenario, ranging from 9.593 M€ to 111.5 M€. The largest avoided damage is reached for scenario e_5 (100-year design flood), with a total reduction ΔD_5 = 111.5 M€, whereas the mitigation measures performed worst for the extreme flood scenario e_6, with a mean avoided damage ΔD_6 = 9.593 M€ and a negative minimum value of -4.695 M€, meaning that the damage costs might increase from the present to the future situation for this scenario. The Average Annual Avoided Damage indicator shows a mean value of ΔAAD = 5.459 M€/year. Table 3.8 on the following page clearly suggests that the contribution of the four types of assets to this total indicator is uneven: while the economic activities and private housing account respectively for 64% and 34% of the total ΔAAD, the share of campgrounds and cultivated land is only equal to 1.3% and 0.7%, respectively. Finally the effect of the flood mitigation measures appears to be heavily dependent on the type of assets considered.

Despite all input uncertainties, the ΔAAD indicator on private housing and economic activities (other than agriculture and campgrounds) proves to be always positive in this uncertainty analysis. In contrast, the mitigation measures will most likely result in an increase in the average annual damage for agricultural land as ΔAAD is negative in this sector for all model runs. Regarding campgrounds, no conclusion can be drawn from the study as ΔAAD is positive in this sector for only 72.6% of the model runs. It can also be noted that all flood damage indicators display a coefficient of variation ranging from 11.76% to 44.80%.

Finally, Figure 3.20 on the next page shows the empirical distribution of the Net Present Value of the flood mitigation measures over N_{tot} = 28 672 simulations. With a mean value of +34.29 M€, the flood risk management project seems to be a sound investment. The NPV indicator also appears to be positive for 96% of model runs, which we interpret to mean that despite all the uncertainty sources that were considered in the NOE modelling chain, the benefits of the flood mitigation measures still prove to almost certainly outweigh their costs.

3.3.4.2 Sensitivity analysis

The total-order variance-based sensitivity indices were computed for each uncertain model input with respect to each output of interest (Table 3.9 on page 128). First, the variance of the total ΔAAD indicator can be observed to be almost equally explained by the uncertainty in the exceedance frequencies, water depth maps, depth-damage curves and LU-GDB dataset, with sensitivity indices ranging from 0.18 to

Table 3.8: Descriptive statistics for each output of interest over N_{tot} = 28 672 simulations

Output	mean	s.d.	2.5% perc.	97.5% perc.	c.var.
Avoided damage per flood scenario (ΔD [M€])					
ΔD_1	0	0	0	0	—
ΔD_2	20.21	7.796	11.164	40.6443	38.57%
ΔD_3	48.84	8.972	29.307	65.927	18.37%
ΔD_4	58.28	10.63	39.491	84.334	18.24%
ΔD_5	111.5	13.11	84.410	136.405	11.76%
ΔD_6	9.593	4.298	1.967	20.149	44.80%
Average Annual Avoided Damage (ΔAAD [M€/year])					
Total	5.459	1.11	3.6109	7.9816	20.33%
Eco. activities	3.506	0.9406	2.0995	5.77331	26.83%
Private housing	1.887	0.4465	1.1002	2.8317	23.66%
Campgrounds	0.071	0.129	-0.213	0.3116	18.23%
Agricultural land	$-5.274.10^{-3}$	$1.912.10^{-3}$	$-9.25.10^{-3}$	$-1.965.10^{-3}$	36.25%
Net Present Value (NPV [M€])					
NPV	34.29	21.01	-40.83	106.5	61.27%

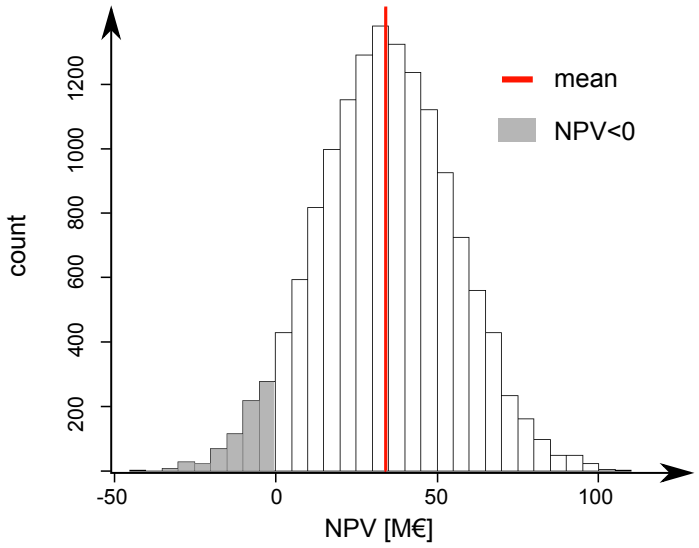

Figure 3.20: Empirical distribution of the NPV indicator over N_{tot} = 28 672 simulations and mean value (solid line)

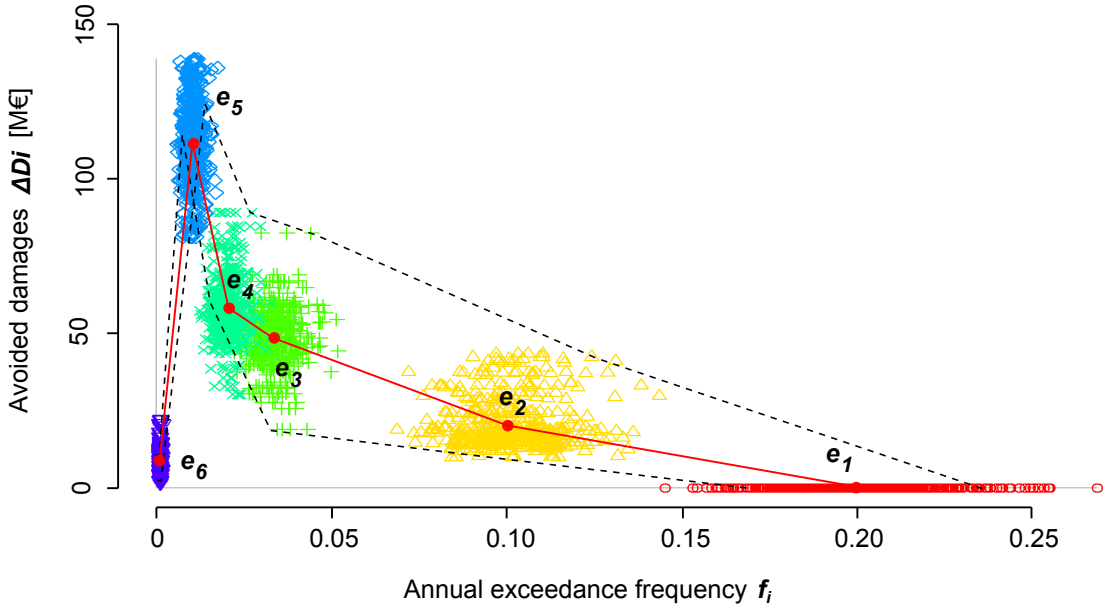

Figure 3.21: Avoided damages against exceedance frequency: nominal curve (solid line), first 500 simulations (dots), curves associated with minimum and maximum ΔAAD (dashed lines)

0.33. No main source of uncertainty can be identified, meaning that they are in a sense "*well-balanced*". However, sensitivity indices with respect to the partial ΔAAD indicator for each economic sector give a very different picture. The variance of ΔAAD on private housing appears to be mainly explained by the uncertainty of the depth-damage curves (sensitivity index: 0.78). For campgrounds and agricultural land, the depth-damage curves also prove to be the most important source of uncertainty (sensitivity index: 0.6 and 0.4, respectively), followed by the uncertainty of the LU-GDB dataset (sensitivity index: 0.38 for both sectors). In addition, for private housing, agricultural land and campgrounds, the uncertainty in the water depth maps is almost non-influential (sensitivity index < 0.02), while it is the second most important source of uncertainty for other economic activities (sensitivity index: 0.38). Finally, Table 3.9 on the next page also indicates that the variance of the Net Present Value of flood mitigation project is mainly due to the uncertainty on its benefits (measured by total ΔAAD) rather than the uncertainty on its costs, which contribute to only 12% of the NPV variance.

To further identify the factors that explain the variance of ΔAAD indicator, sensitivity analysis can also be based on a graphical analysis. Figure 3.21 displays the sampled flood exceedance frequencies f_1 to f_6 against the estimated avoided damages ΔD_1 to ΔD_6 for all N_{tot} = 28 672 model runs, along with the curves associated with the nominal, minimum and maximum value of the ΔAAD indicator, which is equal to the area under the curve. It may be noted that as the damage and frequency estimates of each flood scenario are correlated, the extremum values of the ΔAAD indicator do not always correspond to the extremum points in the damage-frequency graph for all flood scenarios. Figure 3.20 on the preceding page supports the conclusion that the variance of the ΔAAD indicator is mainly due to the uncertain position of scenarios e_1 (first flood event where the damage to property begins) and e_2 (10-year design flood) on this damage-frequency graph. Flood scenarios e_3 to e_5 show a larger dispersion of estimated avoided damages, but their position on the x-axis (exceedance frequency) is less spread; hence their contribution to the total variance of the ΔAAD indicator is small.

Table 3.9: Total-order sensitivity indices with respect to the different outputs of interest. Grey cases indicate the most important sources of uncertainty

	Total-order sensitivity index of:				
	Exceed. Prob.	Water depths	Depth damage curves	LU-GDB dataset	Project costs
Average Annual Avoided Damage (per economic sector)					
Total	0.33	0.18	0.29	0.21	0
Eco. activities	0.4	0.38	0.2	0.3	0
Private housing	0.22	0.005	0.78	0.03	0
Campgrounds	0.2	0.01	0.6	0.38	0
Agricultural land	0.27	0.02	0.4	0.38	0
NPV	0.25	0.18	0.23	0.21	0.12

3.3.5 Discussion

Assessing robustness of a flood risk CBA study Our first goal was to assess the robustness of the cost-benefit analysis of the structural flood mitigation measures on the Orb River through an uncertainty analysis of the NOE modelling chain. Our approach was strongly motivated by prior publications in which uncertainty analysis was used to evaluate the robustness of flood damage assessments. We completed these works by propagating uncertainty up to the cost-benefit analysis of the flood mitigation measures, which had not been performed before. We obtained empirical descriptive statistics for the two CBA-AD outcomes, the Average Annual Avoided Damages and the Net Present Value of flood mitigation measures, displaying quite large coefficients of variation of 20.33% and 61.27%, respectively. The visualisation of the Monte Carlo simulations on a damage-frequency graph (Figure 3.21 on the preceding page) gave us a new insight on the robustness of the ΔAAD indicator, proving that its variance is mainly explained by the uncertain characterisation (in terms of exceedance frequency and estimated damages) of flood scenarios with small return intervals (5 and 10-year floods). We also found (Figure 3.20 on page 126) that the probability of the project costs outweighing the project benefits appears to be lower than 5%. We are convinced that these results may prove useful to provide water managers and stakeholders with a more complete picture on the cost-benefit analysis and the associated uncertainty, even though we know that they are usually untrained in coping with the uncertainty related to scientific information in flood risk studies (Morss et al. 2005).

3.3.5.1 Improving the NOE modelling framework

Our research also sought to identify the main sources of uncertainty in the NOE modelling chain to find ways to improve it. Although variance-based sensitivity analysis is widely used for that purpose in many disciplines, it had never been applied, to our knowledge, to a CBA of flood control measures. We demonstrated the use of Sobol' sensitivity indices (Table 3.9) to rank the sources of uncertainty, depending on their contribution to the variance of the ΔAAD and NPV indicators. A first conclusion is that, on a global scale, sources of uncertainty are well-balanced, meaning that they all explain a significant share of the variance of the ΔAAD and NPV indicators. Yet, more useful lessons can be learnt from looking at each economic sector separately. The uncertainty on the depth-damage curves proved to be the key factor that explains the variance of the average annual damages on private housing, campgrounds and

agricultural land. This result is in line with those of Apel et al. (2008) and de Moel and Aerts (2011), who found that in damage assessment of a single flood event, the choice of damage functions is a much more important factor than the choice of a hazard model. It thus supports the conclusion that improving the depth-damage curves and more generally the damage functions are priorities to make more robust flood damage assessments in these sectors. Our results also indicate that there is almost a third of the variance of the ΔAAD and NPV indicators that cannot be reduced as it stems from the variance of the flood return interval estimates. This observation corroborates the conclusions of Apel et al. (2004), who stated that reliable extreme value statistics were crucially important for reducing the uncertainty of the risk assessment. Unfortunately, reducing this input uncertainty would require longer time series of maximum discharges at Tabarka gauging station, which are not available. Finally, the uncertainty on the floor elevation of buildings proved to have a negligible contribution to the variability of the annual damage estimates for private housing. This finding fits well with the results of Koivumäki et al. (2010), who showed that adding a single elevation value per building was inadequate to obtain more accurate damage estimates. Of course, these findings are specific to the Orb River study site: in a different case study, ranking of the various uncertainty sources may be significantly different.

3.3.5.2 Averaging-out effects

The sensitivity analysis also provided an interesting insight on how the uncertainty on inundation maps influence the variance of damage estimation. Our results offer evidence that improving water depth estimation would be of almost no use in reducing variance of ΔAAD estimates for campgrounds, agricultural land and private housing, while it is the second most important source of uncertainty for other economic activities. This finding is in apparent conflict with the conclusions of Apel et al. (2004, 2008) or de Moel and Aerts (2011), who reported that uncertainty in the water depths is less important than other uncertainty sources without distinction of the economic sector considered. This discrepancy may be explained by two different "*averaging-out effects*": one based on the surface area of assets and the other based on the number of assets. On one hand, campgrounds and agricultural land have a large surface area compared with other types of assets (33 000 and 9 000 sq. m. in average, Table 2.2 on page 62): as a result, the error on water depths, if unbiased, is reduced when it is averaged over the large surface area of these assets. Hence, for both sectors, the contribution of the water depth maps to the variance of the ΔAAD indicator is low. On the other hand, the polygonal features classified as *private housing* or *economic activities* have a rather small surface area (83 sq. m. and 904 sq. m. on average, respectively), and the uncertainty in the average water depth for each individual asset is thus large. Nevertheless, the number of features classified as *private housing* (16436) is much larger than the number of assets classified as *other economic activities* (691), which results in a "*number averaging-out effect*": the dispersion of water depth errors is averaged over the large number of housing polygons scattered across the study area. A similar "*number averaging-out effect*" may partly explain why the uncertainty on depth-damage curves appears to be more influential on the private housing sector, which is described with only one depth-damage curve, than for the other economic activities, which are described by 60 damage curves that are assumed to vary independently. These findings support the conclusion that various averaging-out effects (related not only to the surface area of assets and the number of assets but also to the number of land use types considered, the number of damage functions used, etc.) control the ranking of the uncertainty sources in the NOE modelling chain. Saint-Geours et al. (2012) discussed this issue from a theoretical point of view and showed that the ranking of the uncertainty sources is closely related to the spatial support (and thus to the scale) of the model output. This result is in agreement with the call of Koivumäki et al. (2010) for further research on what constitutes a reasonable scale-accuracy relationship in flood damage assessments: our results suggest that the scale-accuracy-sensitivity relationship must be further investigated.

3.3.5.3 Limits

It should be noted that our work is based on hypotheses that may limit the strength of some of its results. First, some sources of uncertainty were identified in the NOE modelling chain but not taken into account in the uncertainty and sensitivity analysis: the evolution of land use over the next thirty years, errors arising from uncertainty on friction coefficients in hydraulic modelling, errors on the location and shape of polygonal features in the LU-GDB dataset, etc. Why were they ignored? Because the data required to rigorously characterise them in a probabilistic framework were not available. Even if we can assume that some of these uncertainty sources would prove to be negligible in the NOE modelling chain, (e.g., errors on the location of assets in the LU-GDB dataset), others are definitely not (e.g., biased errors in hydraulic modelling). Moreover, even for those sources of uncertainty that were included in the study, the models of uncertainty are sometimes only based on expert opinion. The results of our study heavily depend on such uncertainty parameters. To cope with this issue, a second level uncertainty analysis could be performed by exploring how the output uncertainty and the sensitivity indices change for given sets of uncertainty parameters.

3.3.6 Conclusion

This work was performed with a view toward promoting the use of Monte Carlo uncertainty analysis and variance-based sensitivity analysis in flood damage assessments and related CBA-AD through a case study on the Orb Delta, France. For this case study, we derive the following main conclusions:

1. Monte Carlo uncertainty analysis allows empirical pdf and confidence bounds on the economic indicators of a cost-benefit analysis applied to flood mitigation measures to be computed.
2. The variance of the Average Annual Avoided Damages is mainly due to the uncertain characterisation of flood scenarios with small return intervals.
3. Approximately one-third of the variance of the ΔAAD and NPV indicators cannot be reduced as it stems from a flood frequency analysis based on short time series.
4. The ranking of uncertainty sources depend on the economic sector considered (private housing, agricultural land, economic activities)
5. Uncertainty in the depth-damage curves is a prominent factor for computing the ΔAAD for private housing and agricultural land.
6. The ranking of uncertainty sources is influenced by various averaging-out effects that depend on the number and surface area of the assets considered, the number of land use types, the number of damage functions, etc.

Further research is now needed to extend the reach of our study by trying to reduce the uncertainty in the input data identified as being influential in the study and including in the analysis uncertainty sources that were ignored so far.

➤ **Note to the reader:** End of the draft paper submitted to the *Journal of Flood Risk Management*.

3.3.7 Extra comments: results on other case studies

This subsection aims at giving some extra comments to enrich the results and the discussion presented in the submitted paper.

An important question is what general conclusions can be drawn on the NOE modelling framework from the single Orb Delta case study. Would these results hold for other study sites as well? To try and answer this question, we carried out a sensitivity analysis of the NOE modelling framework on two other study sites that were presented in Chapter 2 (§2.3.3 on page 70):

- the Vilaine floodplain (total extent of 17 sq. km.)
- the ZEC project on the Rhône river (650 sq. km.)

On these two case studies, we followed the same general sensitivity analysis flowchart as on the Orb Delta. The models chosen to describe uncertainty on each model input were similar, even though there were some small differences from one case study to one another. We will not present into details the results obtained on these two study sites, but here is a summary of the most important observations.

Ranking of model inputs First, it can be stated that the ranking of uncertainty sources obtained on the Orb Delta does not necessarily hold for other case studies. For example, here are the first-order sensitivity indices computed on the Redon floodplain with respect to the total AAD indicator:

depth-damage curves	0.15
flood exceedance probabilities	0.17
assets map	0.29
water depths	0.36

It appears that the most influential inputs with respect to the variance of the total average annual damages on this floodplain are the assets map and the water depths, whereas the same inputs were the less influential on the Orb Delta case study (Table 3.9 on page 128). We will discuss one possible explanation for this difference in the Chapter conclusion on page 134.

Influence of economic sector Next, these two extra studies corroborate the observation that the sensitivity indices of the NOE model inputs depend on the economic sector considered (private housing, agriculture, industry, etc.). For example, here are the first-order sensitivity indices obtained on the ZEC study site with respect to the total AAD indicator for: i) cultivated land only; ii) business and industrial assets only; and iii) private housing only:

	agriculture	*industry*	*private housing*
depth-damage curves	0.66	0.2	0.74
flood return intervals	0.08	0.13	0.06
assets map	0.00	0.00	0.00
hazard maps	0.16	0.54	0.01

This table suggests in particular that depth-damage curves are much more influential for agricultural assets and private housing assets than they are for business or industrial assets. This observation is in line with the results obtained on the Orb Delta (Table 3.9 on page 128).

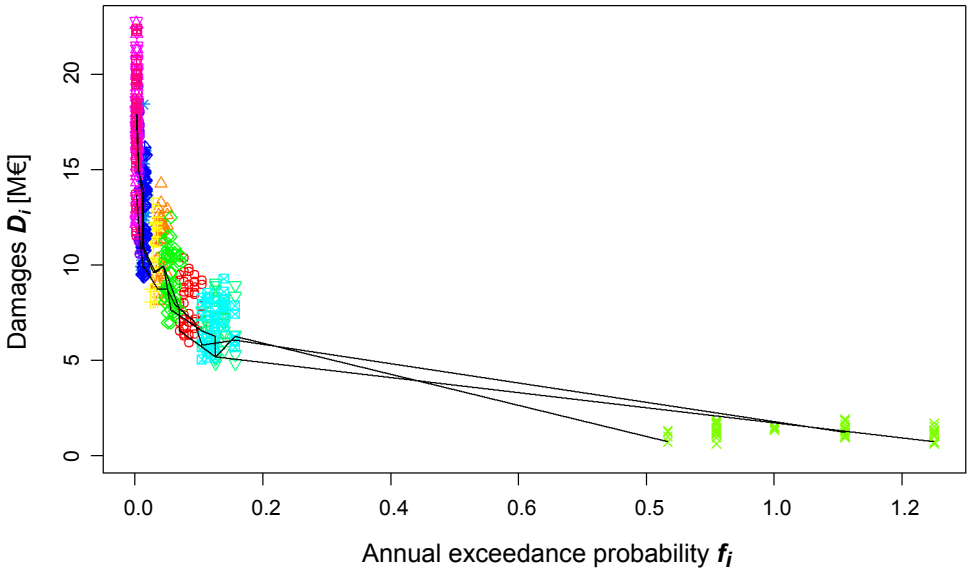

Figure 3.22: ZEC case study. Damages against annual exceedance probabilities: first 100 simulations (dots), three random simulations (solid lines)

Influence of flood scenarios with small return intervals Finally, we also observe on these two extra study sites that the flood scenarios with small return intervals contribute more to the variance of the AAD and ΔAAD indicators than scenarios with large return intervals, which is line with the results obtained on the Orb Delta. This observation is not based on the computation of sensitivity indices, but on the same graphical analysis as the one realized on the Orb Delta (§3.3.4.2 on page 127), in which uncertain damage estimates for each flood scenario are plotted against its uncertain exceedance probability— Figure 3.22 displays a similar plot for the ZEC study site. However, it is important to underline that this observation is most probably linked to the fact that, for all the case studies, uncertainty on flood exceedance probabilities was modelled in the same way, using empirical confidence bounds on the discharge-frequency curve (§3.3.3.1 on page 119). If another uncertainty model had been chosen for flood exceedance probabilities, then we may not have found that flood scenarios with small return intervals are the most influential.

We will try to draw some lessons from these comparative observations in the Chapter conclusion on page 134.

3.4 Chapter conclusion

3.4.1 On VB-GSA with spatially distributed inputs

The goal of this chapter was to investigate how VB-GSA could be extended to models with one or several spatially distributed inputs $Z_i(\mathbf{x})$.

State of the art of the available methods We established a state of the art of the available techniques to handle a spatially distributed model input $Z(\mathbf{x})$ in VB-GSA. We applied these techniques to a number of analytical test cases and discussed the outcomes of this numerical study.

A first key point is to clarify what measure of importance the modeller is interested in. The "*macro-parameter*", "*dimension reduction*","*map labelling*" and "*joint meta-models*" aim at measuring the contribution of the entire uncertain spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ to the variance of the model output Y. In these methods, the characterisation of the uncertainty on $Z(\mathbf{x})$ is fixed—fixed pdf of scalar parameters λ_i or fixed stochastic process $\mathcal{P}$ that generates random realisations of $Z(\mathbf{x})$. What is measured is the contribution of the overall uncertainty of $Z(\mathbf{x})$ to the variance of model output. On the contrary, in the "*second level*" approach, the description of uncertainty in $Z(\mathbf{x})$ is itself uncertain. What is measured is the contribution to var(Y) of the uncertain "*second level*" parameters θ_i that control the stochastic process $\mathcal{P}$. Finally, in the "*trigger*" method, the sensitivity indices that are computed do not properly measure the contribution of $Z(\mathbf{x})$ to the variance of model output Y: we recommend no to use this approach. In our research, we focused on the first group of techniques, that measure the contribution of the entire uncertain spatially distributed input $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ to the variance of the model output Y, with a fixed description of uncertainty.

Next, we systematically described the available techniques for a number of criteria and suggested to use a decision-tree (Figure 3.7 on page 96) to choose among them for a given spatial model. In our opinion, the most appropriate technique to handle spatially distributed inputs in VB-GSA of a numerical model $\mathcal{F}$ with low CPU cost is the "*map labelling*" approach. It is a convenient approach, easy to implement and to explain to model end-users. An important feature is that it allows a complex description of the spatial structure of uncertainty in $Z(\mathbf{x})$: random realisations of the spatially distributed input can be generated by any stochastic process $\mathcal{P}$ and ad-hoc algorithm. For all these reasons, we chose to use the "*map labelling*" technique to carry out VB-GSA of the NOE modelling framework on the Orb Delta case study. On the other hand, for time consuming models, the "*joint meta-model*" appears to be a good solution, except that it can only handle a single spatially distributed input.

Sampling issues in the "*map labelling*" approach As developed in §3.2 on page 98, the "*map-labelling*" approach requires sampling a number n of random realisations of the spatially distributed input $Z(\mathbf{x})$ to estimate its variance-based sensitivity indices. The choice of size n is usually driven by constraints of time and disk space. We carried out two numerical studies to investigate how sampling technique and size may influence the estimation of sensitivity indices: i) one to study the convergence of sensitivity indices estimates for increasing number n of random realisations of $Z(\mathbf{x})$; ii) the other one to compare two geostatistical simulation algorithms (SRS and LHS) to generate random realisations of $Z(\mathbf{x})$ when it is modelled as a Gaussian Random Field. Due to their exploratory nature, we cannot draw any firm conclusion from these experiments, neither on the minimum sampling size n in a general case, nor on the optimal sampling technique when $Z(\mathbf{x})$ is a Gaussian Random Field. However, there are two humble lessons to be learnt from these numerical studies: i) for small values of n the set of random realisations of $Z(\mathbf{x})$ is too small to represent reasonably well the overall uncertainty on the spatially

distributed input; hence, figures showing the convergence of sensitivity indices estimates with increasing number n of random realisations of $Z(\mathbf{x})$ should be drawn for each case study to determine the minimum sampling size n; ii) when $Z(\mathbf{x})$ is modelled as a Gaussian Random Field, spatial LHS does not yield much better sensitivity indices estimates than SRS, because it disturbs spatial correlation for small values of n.

Further research needed We have shown in this chapter that the available techniques to handle spatially distributed inputs in VB-GSA are in some respects limited. Further research is thus needed to improve these techniques, and design new methods to deal with unresolved situations. We will discuss these directions for future work in the general conclusion on page 174. We will also display in this general conclusion some ideas to extend the reach of our study.

3.4.2 On the NOE modelling framework and CBA-AD studies

Sensitivity analysis of the NOE modelling framework on the Orb Delta case study We successfully used the "*map labelling*" approach to perform VB-GSA of the NOE code on the Orb Delta case study. This approach allowed us to consider the uncertainty on two spatially distributed inputs: the assets and the hazard maps. Results show that the uncertainty on flood scenario annual exceedance probabilities is the main contributor to the variance of the ΔAAD and NPV output indicators. Unfortunately, this source of uncertainty cannot be easily reduced, because improving flood frequency estimates would require a longer time serie of annual maximum flows at the reference gauging station. In addition, we highlighted the major role of uncertain annual exceedance probabilities associated to the flood scenarios with small return intervals.

Next, we found that the ranking of uncertainty sources with respect to their contribution to the variance of the ΔAAD indicator depends on the economic sector considered (private housing, agriculture, industry, etc.). In particular, uncertainty in the depth-damage curves is a prominent source of uncertainty when computing the ΔAAD indicator for private housing and agricultural land.

An underlying explanation is that the structure of input data heavily depends on the economic sector considered: number of assets, average surface area of assets, nomenclature of assets, number of damage functions considered, shape of damage functions (linear/non linear, with/without threshold, etc.). For all these caracteristics, an "*averaging-out*" effect may occur, that is, many independent uncertainty sources may compensate each other when they are combined. This point will be further discussed in Chapter 4.

Generalizing the results obtained on the Orb Delta case study As mentioned earlier in §3.3.7 on page 131, an important question is what general conclusions can be drawn on the NOE modelling framework from the single Orb Delta case study. To adress this issue, we carried out sensitivity analysis on two other study sites: the Vilaine floodplain and the ZEC project along the Rhône river.

A first key result is that the specific ranking of uncertainty sources obtained on the Orb Delta does not hold for other case studies. One possible explanation lies in the difference of spatial extent between these study sites: the Orb Delta study site has a total extent of 63 sq. km., whereas the other two sites we investigated have a total extent of 17 sq. km. (Vilaine floodplain) and 650 sq. km. (ZEC project), respectively. The outcomes of sensitivity analysis may depend on a "*spatial averaging-out effect*" related to the total surface area of the floodplain. This point will be further investigated in Chapter 4.

Besides, this comparative analysis shows that the other conclusions drawn from the Orb Delta case study will hold for any other floodplain as well: i) the ranking of uncertainty sources depends on the economic sector considered; and ii) if the uncertainty on flood exceedance probabilities is simulated based on an empirical confidence bound around the discharge-frequency curve at a reference gauging station, then the uncertainty on flood scenarios with small return intervals contributes more to the variance of the AAD and ΔAAD indicators than flood scenarios with large return intervals.

Spliting groups of model inputs As mentioned in §2.3.2.3 on page 64, the numerous inputs of the NOE code were gathered into five groups to make the sensitivity analysis computationally more tractable. The outcomes of the analysis heavily depend on the initial choices that were made to compose these groups. For example, we cannot compute in this setting the variance-based sensitivity indices associated to the uncertain exceedance probability of a single flood scenario—but we have access to the sensitivity indices related to the group of exceedance probabilities for all scenarios. We cannot either discuss which of the 94 damage functions used in the Orb Delta case study contribute the most to the variance of the ΔAAD indicator. To go one step further, some of the five groups considered in this first analysis should be split into smaller groups or individual inputs for a second analysis.

Nevertheless, a group can be split only if its components (individual inputs or smaller groups) are assumed to be statistically independent. In the NOE modelling framework, at least two out of five groups cannot be divided: i) the group of scenario weights $\omega(e_j)$ (related to flood annual exceedance probabilities), which are obtained from a unique time serie of annual maximum flow at the reference gauge and are thus correlated; ii) the hazard maps $\mathcal{H}(e_j)$, which are produced from the same hydraulic model and the same uncertain DTM and thus cannot be considered as uncorrelated inputs. On the contrary, we have assumed in our first analysis that the uncertainty on each damage curve is independent from the others: hence, it could be possible to split the group of damage functions and assess the contribution of each damage function to the variance of the NOE outputs. In the same way, the group of scalar inputs describing the costs of the flood risk management plan can be split, if we assume that the uncertainty on the initial investment costs CI is independent from the uncertainty on the maintenance costs CM. The case of the last model input, the assets map, is more complex: this model input contains a number of uncertain attribute data, such as the groundfloor elevation of buildings, the type of assets, the surface area of assets, etc. One may be interested in calculating the individual contribution of one these attributes to the variance of model outputs. This could be performed by spliting the assets map into a number of simple maps with a single uncertain attribute each. Unfortunately, due to lack of time none of these complementary analyses was carried out.

Choice of the uncertainty models Last but not least, a key limit of our sensitivity analysis of the NOE code on the Orb Delta case study is the choice of the uncertainty sources that were considered and their modelling. As explained in §3.3.5.3 on page 130, some sources of uncertainty were identified in the NOE modelling framework but not taken into account. Besides, even for those sources of uncertainty that were included in the study, the choice of a model of uncertainty was sometimes supported by few or even no data (e.g., damage functions). This limit will be carefully discussed in the general conclusion.

Chapter 4

Scale issues in variance-based global sensitivity analysis

THIS chapter is devoted to the exploration of "*scale issues*" in VB-GSA of spatial models. Our interest in this question arised when comparing the results of sensitivity analysis of the NOE modelling framework obtained on different case studies (§3.3.7 on page 131). We observed that the ranking of the uncertainty sources was different on each investigated case study, for which the extent of the study area ranged from approximately 10 km^2 to more than 100 km^2. We made the hypothesis that these variations could be partly explained by some "*averaging-out effect*" of local uncertainties related to the scale of each study site. To clarify our question, we will use the notion of the "*scale triplet*" introduced by Blöschl and Sivapalan (1995) and presented in Chapter 1 (§1.1.3 on page 20). We offer to study how the results of VB-GSA of a spatial model depend on two components of the scale triplet: the "*support*" and the "*spacing*". The third component ("*extent*") will only be discussed in the Chapter conclusion.

We limit our study to the specific class of point-based and spatially additive models, as defined in Chapter 1. We consider a model $\mathcal{F}$ with a single spatially distributed input $Z(\mathbf{x})$, which is modelled as a Stationary Random Field (SRF). The output of interest is the average value Y_v of $Y(\mathbf{x})$ over a given support $v \subset \Omega$. This limitative framework is partly justified by the fact that our case study model (the NOE code) is itself point-based and spatially additive. The other reason is that a number of nice properties can be obtained analytically for this class of models, which can approximate a number of "*real-world*" applications.

This chapter is composed of three sections. The first section §4.1 on page 140 was published in *Mathematical Geosciences* with the following title: "*Change of support in spatial variance-based sensitivity analysis*" (Saint-Geours et al. 2012). In this publication, we first survey the few existing papers that discuss scale issues in sensitivity analysis. Next, we define "*site*" and "*block*" sensitivity indices to account for the role of spatial support v over which the model output is aggregated. We then explain how block sensitivity indices depend not only on the size of support v, but also on the covariance structure of the spatially distributed input $Z(\mathbf{x})$. These theoretical developments on change of support are illustrated by a simple analytical test case. Then, in section §4.2 on page 153, we investigate how the results of VB-GSA of model $\mathcal{F}$ may be influenced by another component of the scale triplet: the spacing. We consider the case in which the aggregated output Y_Ω is approximated by a weighted sum of the output values $Y(\mathbf{x}_i)$ computed at a finite number of points $\mathbf{x}_i \in \Omega$. We give an expression of the resulting approximation error on sensitivity indices, and show that this error depends on the spacing of the set of points $\mathbf{x}_i$. Next, section §4.3 on page 156 illustrates the effect of spatial support on VB-GSA results for the NOE modelling framework on the Orb Delta case study. Sensitivity indices of the NOE model inputs are computed with respect to the aggregated values of the ΔAAD indicator over varying spatial supports v, to investigate the relationship between the ranking of uncertainty sources and the area over which the ΔAAD indicator is summed up. Finally, we close this chapter with some concluding remarks in §4.4 on page 169.

4.1 Influence of support 140
4.1.1 Introduction 140
4.1.2 VB-GSA for a point-based and spatially additive model 141
4.1.2.1 Description of spatial model $\mathcal{F}$ 141
4.1.2.2 Site sensitivity indices and block sensitivity indices 142
4.1.2.3 Illustrative example 143
4.1.3 Change of support effect on block sensitivity indices 144
4.1.3.1 Relation between site sensitivity indices and block sensitivity indices 144
4.1.3.2 Change of support effect 145
4.1.3.3 Link between covariance function and block sensitivity indices 146
4.1.3.4 Illustrative example 147
4.1.4 Discussion 147
4.1.5 Conclusion 150
4.1.6 Appendix A: Proof of the relation between site sensitivity indices and block sensitivity indices 151
4.1.7 Appendix B: Hermitian expansion of random field $\mathbb{E}_Z Y(\mathbf{x})$ 152
4.2 Influence of spacing 153
4.2.1 Introduction 153
4.2.2 Model description 153
4.2.3 Approximation on a grid of points 154
4.2.4 Expression of the approximation error and convergence 154
4.2.5 Conclusion 155
4.3 Application to the NOE code on the Orb Delta case study 156
4.3.1 Introduction 156
4.3.2 Methods 156
4.3.2.1 Overview 156
4.3.2.2 Setting A: ΔAAD indicator on individual supports 157
4.3.2.3 Setting B: raster maps of the ΔAAD indicator 157
4.3.3 Results 160
4.3.3.1 Setting A: ΔAAD indicator on individual supports 160
4.3.3.2 Setting B: raster maps of the ΔAAD indicator 162
4.3.4 Discussion 166
4.4 Chapter conclusion 169
4.4.1 On scale issues in VB-GSA of spatial models 169
4.4.2 On the NOE modelling framework 171

4.1 Influence of support

The purpose of this section is to analyse how the sensitivity indices in a point-based and spatially additive model depend on the support v over which the model output Y_v is aggregated. As mentioned in the chapter introduction, our interest in this question stems from the comparative analyses of the NOE code on different case studies (§3.3.7 on page 131). We observed that the ranking of the uncertainty sources was different on each investigated case study, for which the extent of the study area ranged from approximately 10 km^2 to more than 100 km^2. We made the hypothesis that these variations could be related to the surface area of each study site. Intuitively, one possible explanation is the folllowing: the "*local*" uncertainty of spatially distributed inputs (i.e., the assets map and the hazard maps) may compensate each other and finally average out when the output of interest $Y(\mathbf{x})$ is aggregated over a large area. Hence, the contribution of the spatially distributed inputs to the variance of the aggregated model output is smaller on large areas, and larger on small areas. In this section, we intend to give a theoretical presentation of this effect.

➤ **Note to the reader:** This section is a reproduction of a paper published in *Mathematical Geosciences*, with the following title: "*Change of support in spatial variance-based sensitivity analysis*" (Saint-Geours et al. 2012). Some parts of this paper are redundant with explanations that were given in previous chapters: they can safely be skipped by the reader. Besides, some extra comments were added to the published paper and its appendices: they are identified by a grey box. Finally, in order to keep a general consistency of notations throughout this document, some of the original notations given in the published paper were changed.

4.1.1 Introduction

Variance-based global sensitivity analysis (VB-GSA) is used to study how the variance of the output of a model can be apportioned to different sources of uncertainty in its inputs. Here, the term "*model*" denotes any computer code in which a response variable is calculated as a deterministic function of input variables. Originally developed in the 1990s (Sobol' 1993), VB-GSA is now recognized as an essential component of model building (European Commission 2009a; CREM 2009) and is widely used in different fields (Cariboni et al. 2007; Tarantola et al. 2002). VB-GSA is based on the decomposition of a model output variance into conditional variances. So-called "*first-order sensitivity indices*" measure the main effect contribution of each uncertain model input to the model output variance. Based on these sensitivity indices, ranking the model inputs helps to identify inputs that should be better scrutinized first. Reducing the uncertainty on the inputs with the largest sensitivity indices (e.g., by collecting additional data or changing the geographical pattern of data locations) will often result in a reduction in the variance of the model output. More generally, VB-GSA helps to explore the response surface of a "*black box*" computer code and to prioritize the possibly numerous processes that are involved in it.

Although VB-GSA was initially designed for models where both inputs and output can be described as real valued random variables, some recent work has extended VB-GSA to environmental models for which both the inputs and output are spatially distributed over a two-dimensional domain and can be described as random fields (Lilburne and Tarantola 2009 for a review). In these works, the computer code under study uses maps derived from field data (e.g., digital elevation models and land use maps). These maps are uncertain due to measurement errors, lack of knowledge or aleatory variability (Refsgaard et al. 2007; Brown and Heuvelink 2007). The uncertainty of these spatial inputs is usually modeled

using random fields. Model output is also spatially distributed (e.g., a flood map or a pollution map). Authors use geostatistical simulation to incorporate spatially distributed model inputs into the VB-GSA approach (Ruffo et al. 2006; Saint-Geours et al. 2010) and they display estimation procedures to compute sensitivity indices in a spatial context, either with respect to the spatial average of the model output (Lilburne and Tarantola 2009) or with respect to the values of the model output at each site of a study area (Marrel et al. 2011; Pettit and Wilson 2010; Heuvelink et al. 2010b).

Nevertheless, to date, none of these studies has reported on a key issue: the link between uncertainty propagation and model upscaling/downscaling. "*Model upscaling*" is the problem of translating knowledge from smaller scales to larger (Heuvelink 1998). In many environmental models, the physical quantities considered are spatially additive (e.g., porosity or evapotranspiration), i.e., their large-scale properties derive from small-scale properties by simple averaging (Chilès and Delfiner 1999 p.593). In this case, the model end user is usually interested in the spatial linear average or the sum of spatial output over a given spatial unit (e.g., the average porosity of a block or the total evapotranspiration over a plot of land) and model upscaling is thus reduced to a "*change of support*" problem (namely, a change of support of the end user's output of interest). Heuvelink (1998) observed that under a change of spatial support of the model output, the relative contribution of uncertain model inputs to the variance of the aggregated model output may change. Exploring how sensitivity analysis results interact with such a change of support is thus of great importance. It would allow the modeller to check the robustness of model-based environmental impact assessment studies and better assess the confidence of their results. Knowledge of this interaction would also allow the modeller to answer the following questions: What are the model inputs that explain the largest fraction of the variance of the output over a given spatial support? For which output support size does a given spatially distributed model input contribute to the largest fraction of the variance of the model output? How does the contribution of a spatially distributed input to the variance of the model output depend on the parameters of its covariance function?

The change of support effect has been extensively discussed in geostatistics in the context of regularization theory (Journel and Huijbregts 1978 p.77). Hence, we attempt in this paper to integrate regularization theory with VB-GSA framework. Our idea is to define "*site sensitivity indices*" and "*block sensitivity indices*" to i) provide a simple formalism that extends VB-GSA to spatial models when the modeller's interest is in the spatial average or the sum of model output over a given spatial support (§4.1.2) and ii) discuss how the relative contribution of uncertain model inputs to the variance of model output changes under model upscaling (§4.1.3 on page 144). We limit our study to point-based models, i.e., models for which the computation of the model output at some location uses the values of spatial inputs at that same location only (Heuvelink et al. 2010a). An example is used throughout this paper to illustrate formal definitions and properties. Finally, we discuss the limits of our approach and its connections to related works in §4.1.4 on page 147.

4.1.2 VB-GSA for a point-based and spatially additive model

4.1.2.1 Description of spatial model $\mathcal{F}$

We want to study a computer code $\mathcal{F}$ whose output is a map and whose inputs are a map and a set of k real valued variables. Both inputs and output are "*uncertain*" and are described as random variables or random fields. More precisely, we use the following notations: let $\Omega \subset \mathbb{R}^2$ denote a 2D spatial domain, $\mathbf{x} \in \Omega$ a site, $\mathbf{h}$ the lag vector between two sites $\mathbf{x}$ and $\mathbf{x}'$, and $v \subset \Omega$ some spatial support (block) of area $|v|$. We consider the model $Y = \mathcal{F}(\mathbf{U}, Z)$ where $\mathbf{U} = (U_1, \ldots, U_k)$ is a random vector and $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is a second-order stationary random field (SRF) — that we will often simply denote by $Z(\mathbf{x})$. $\mathbf{U}$ and

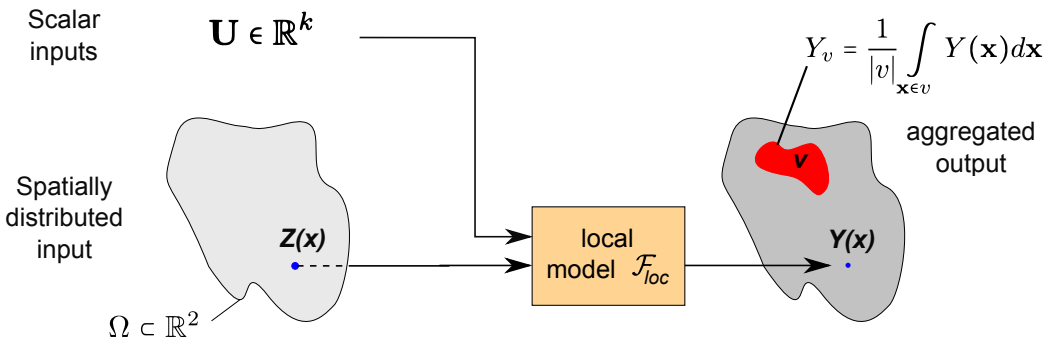

Figure 4.1: Point-based and spatially additive model with uncertain inputs $\mathbf{U}$ and $Z(\mathbf{x})$ and spatial output $Y(\mathbf{x})$. The modeller is interested in the block average of $Y(\mathbf{x})$ over some spatial unit v

$Z(\mathbf{x})$ are supposed to be independent. Covariance function $C(\cdot)$ of $Z(\mathbf{x})$ is assumed to be isotropic, characterized by correlation length $a \in \mathbb{R}$, nugget parameter $\eta \in [0;1[$ and of the form:

$$C(\mathbf{h}) = \begin{cases} C(0) & \text{if } \mathbf{h} = \mathbf{0} \\ (1-\eta)\cdot C(0)\cdot \rho_a(\|\mathbf{h}\|) & \text{if } \mathbf{h} \neq \mathbf{0} \end{cases} \tag{4.1}$$

where $\rho_a(\cdot)$ is some valid correlogram (Cressie 1993 p.67). The model output is a 2D random field $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$ that we will simply denote by $Y(\mathbf{x})$. We assume that the first two moments of $Y(\mathbf{x})$ exist. Finally, as discussed in the introduction, we limit our study to "*point-based*" models; hence, we assume that there exists a mapping $\mathcal{F}_{\text{loc}} : \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$ such that:

$$\forall \mathbf{x} \in \Omega\,, \quad Y(\mathbf{x}) = \mathcal{F}_{\text{loc}}\left[\mathbf{U}, Z(\mathbf{x})\right] \tag{4.2}$$

A sensitivity analysis of the model $\mathcal{F}$ must be performed with respect to a scalar quantity of interest derived from spatially distributed model output $Y(\mathbf{x})$. Here, we consider two different outputs of interest: the value $Y(\mathbf{x}^*)$ at some specific site $\mathbf{x}^* \in \Omega$ and the aggregated value $Y_v = 1/|v| \int_v Y(\mathbf{x})\, d\mathbf{x}$ over support v. Because model inputs $\mathbf{U}$ and $Z(\mathbf{x})$ are uncertain, $Y(\mathbf{x}^*)$ and Y_v are both random variables; the sensitivity analysis will describe the relative contribution of uncertain model inputs $\mathbf{U}$ and $Z(\mathbf{x})$ to their respective variances.

➤ **Extra comment:** according to the definitions given in §1.1 on page 18, the analytical test case under study is a point-based and spatially additive model (Figure 4.1). A complete analytical study has been written for the specific case in which model $\mathcal{F}$ is linear: it is given in Appendix §I on page 237.

4.1.2.2 Site sensitivity indices and block sensitivity indices

Before defining sensitivity indices for spatial model $\mathcal{F}$, we briefly review the mathematical basis of VB-GSA. Let us consider a model $Y = f(U_1, \ldots, U_k)$, where U_i are independent random variables and where the first two moments of Y exist. The first-order sensitivity index S_i of model input U_i is defined by:

$$S_i = \frac{\text{var}\left[\mathbb{E}(Y|U_i)\right]}{\text{var}(Y)} \tag{4.3}$$

$S_i \in [0;1]$ measures the main effect contribution of the uncertain model input U_i to the variance of model output Y. Sensitivity indices can be used to identify the model inputs that account for most of the variance of the model output (model inputs U_i with high first-order indices S_i). Sum of S_i is always less than 1 and the difference $1 - \sum_i S_i$ accounts for the contribution of the interactions between model inputs U_i to model output variance $\text{var}(Y)$. Please refer to Saltelli et al. (2008) for more details on VB-GSA theory and on the estimation of sensitivity indices.

➤ **Note to the reader:** Variance-based global sensitivity analysis and variance-based sensitivity indices are described into more details in §1.2.2 on page 29.

To extend VB-GSA to spatial model $\mathcal{F}$, we propose to use different types of sensitivity indices to describe the relative contribution of the uncertain model inputs $\mathbf{U}$ and $Z(\mathbf{x})$ to the variance of the model output: an index on a point support (i.e., with respect to output of interest $Y(\mathbf{x}^*)$) and an index on a larger support (i.e., with respect to output of interest Y_v). First-order sensitivity indices of model inputs with respect to $Y(\mathbf{x}^*)$ are called "*site sensitivity indices*". Under the stationary hypothesis on SRF $Z(\mathbf{x})$, these indices do not depend on site $\mathbf{x}^*$ and thus will simply be denoted by $S_{\mathbf{U}}$ and S_Z:

$$S_{\mathbf{U}} = \frac{\text{var}\left[\mathbb{E}(Y(\mathbf{x}^*) \mid \mathbf{U})\right]}{\text{var}\left[Y(\mathbf{x}^*)\right]} \quad ; \quad S_Z = \frac{\text{var}\left[\mathbb{E}(Y(\mathbf{x}^*) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\})\right]}{\text{var}\left[Y(\mathbf{x}^*)\right]} \tag{4.4}$$

First-order sensitivity indices of model inputs with respect to the block average Y_v are called "*block sensitivity indices*" and are denoted by $S_{\mathbf{U}}(v)$ and $S_Z(v)$:

$$S_{\mathbf{U}}(v) = \frac{\text{var}\left[\mathbb{E}(Y_v \mid \mathbf{U})\right]}{\text{var}\left[Y_v\right]} \quad ; \quad S_Z(v) = \frac{\text{var}\left[\mathbb{E}(Y_v \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\})\right]}{\text{var}\left[Y_v\right]} \tag{4.5}$$

The ratio $S_Z(v)/S_{\mathbf{U}}(v)$ gives the relative contribution of model inputs $Z(\mathbf{x})$ and $\mathbf{U}$ to the variance of the output of interest Y_v. When $S_Z(v)/S_{\mathbf{U}}(v)$ is greater than 1, the variance of Y_v is mainly explained by the variability of the 2D input field $Z(\mathbf{x})$; when $S_Z(v)/S_{\mathbf{U}}(v)$ is less than 1, it is the non spatial input $\mathbf{U}$ that accounts for most of $\text{var}\,(Y_v)$.

4.1.2.3 Illustrative example

The proposed formalism for spatial VB-GSA is illustrated by the following example. A model $Y = \mathcal{F}(\mathbf{U}, Z)$ is used for the economic assessment of flood risk over a given floodplain Ω. $Z(\mathbf{x})$ is the map of maximal water levels reached during a flood event. $Z(\mathbf{x})$ is assumed to be a Gaussian random field with mean $\mu = 50$ and exponential covariance $C(\mathbf{h})$ with $C(0) = 100$, correlation length $a = 5$ and nugget parameter $\eta = 0.1$. $\mathbf{U}$ is a set of three economic parameters U_1, U_2 and U_3 that determine a so-called "*damage function*" that links water levels to monetary costs. U_1, U_2 and U_3 are assumed to be independent random variables following Gaussian distributions $\mathcal{N}(1.5, 0.5)$, $\mathcal{N}(55, 5)$ and $\mathcal{N}(10, 10)$, respectively. Random field $Z(\mathbf{x})$ and random vector $\mathbf{U}$ are supposed to be independent. Model output $Y(\mathbf{x})$ is the map of expected economic damages due to the flood over the area; these damages depend on $\mathbf{U}$ and $Z(\mathbf{x})$ through the mapping $\mathcal{F}_{\text{loc}}$:

$$\forall \mathbf{x} \in \Omega, \quad Y(\mathbf{x}) = \mathcal{F}_{\text{loc}}\left[\mathbf{U}, Z(\mathbf{x})\right] = U_1 \cdot Z(\mathbf{x}) - U_2 \cdot e^{-0.036 \cdot Z(\mathbf{x})} - U_3 \tag{4.6}$$

Stakeholders are interested in two outputs: the flood damage $Y(\mathbf{x}^*)$ on a specific building $\mathbf{x}^* \in \Omega$ and the total damage $|v| \cdot Y_v$ over a district v (here, a disc of radius $r = 50$). Here, the expression of mapping $\mathcal{F}_{\text{loc}}$

Table 4.1: Sensitivity analysis results over $N_{\text{tot}} = 4\,096$ model runs with respect to the outputs of interest $Y(\mathbf{x}^*)$ and $|v| \cdot Y_v$. Mean values with ±s.d. computed by bootstrapping (100 replicas).

Support	**Site $\mathbf{x}^*$**	**Block v**
Output of interest	$Y(\mathbf{x}^*)$	$\lvert v\rvert \cdot Y_v$
Mean of output	66.5 ± 4.2	$539 \cdot 10^3 \pm 3.6 \cdot 10^3$
Variance of output	$1\,393 \pm 188$	$9 \cdot 10^9 \pm 0.2 \cdot 10^9$
	Site indices:	Block indices:
Sensitivity indices	$S_{\mathbf{U}} = 0.09 \pm 0.03$	$S_{\mathbf{U}}(v) = 0.86 \pm 0.02$
	$S_Z = 0.89 \pm 0.02$	$S_Z(v) = 0.12 \pm 0.02$

and the statistical characterisation of model inputs may be simple enough that exact values of sensitivity indices could be derived, but this is usually not the case in real applications in which the model is very complex. A usual alternative is to consider model $\mathcal{F}$ as a "*black box*" and estimate sensivity indices with Monte-Carlo simulation. We chose to use the estimators and the computational procedure described by Lilburne and Tarantola (2009 §3.2), based on a quasi-random sampling design, using $N_{\text{tot}} = 4\,096$ model runs (Table 4.1). It appears that on a given site $\mathbf{x}^*$, the variability of the water level map explains most of the variance of $Y(\mathbf{x}^*)$: $S_Z = 0.89$. On a larger spatial support, the variance of the total flood damage $|v| \cdot Y_v$ is mainly due to the economic parameters U_1, U_2 and U_3: $S_{\mathbf{U}}(v) = 0.86$. Thus, to improve the accuracy of damage estimation for a specific building, the uncertainty should first be reduced on the water level map $Z(\mathbf{x})$; however, to improve the accuracy of total damage estimation over a large district v, the modeller should focus on reducing the uncertainty of economic parameters U_1, U_2 and U_3.

➤ **Another illustrative example:** we also studied a second analytical test case to illustrate change of support effect in VB-GSA. This test case was published in the proceedings of the *Mathematical Geosciences at the Crossroads of Theory and Practice, IAMG 2011* conference with the following title: "*Sensitivity analysis of spatial models using geostatistical simulation*" (Saint-Geours et al. 2011b). It is reproduced in Appendices.

4.1.3 Change of support effect on block sensitivity indices

In this section, we assess how the ranking of uncertain model inputs based on their block sensitivity indices vary under a change of support v of model output.

4.1.3.1 Relation between site sensitivity indices and block sensitivity indices

Site sensitivity indices and block sensitivity indices are related. Let $\mathbb{E}_Z Y(\mathbf{x})$ denote the conditional expectation of $Y(\mathbf{x})$ given $Z(\mathbf{x})$, that is:

$$\forall \mathbf{x} \in \Omega, \quad \mathbb{E}_Z Y(\mathbf{x}) = \mathbb{E}\left[Y(\mathbf{x}) \mid Z(\mathbf{x})\right] \tag{4.7}$$

$\mathbb{E}_Z Y(\mathbf{x})$ is the transform of the input SRF $Z(\mathbf{x})$ via the function $\bar{\mathcal{F}}_{\text{loc}}(z) = \int_{\mathbb{R}^k} \mathcal{F}_{\text{loc}}(\mathbf{u}, z) p_U(\mathbf{u}) d\mathbf{u}$ [Eqn. (4.2) on page 142] where $p_U(\cdot)$ is the multivariate pdf of random vector $\mathbf{U}$. Under our as-

sumptions concerning $Y(\mathbf{x})$, $\mathbb{E}_Z Y(\mathbf{x})$ is a second-order SRF. Let $C^*(\cdot)$ denote its covariance function, $V^* = C^*(0)$ its variance and V_v^* its block variance over support v, that is, the variance of block average $1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x}) d\mathbf{x}$. Block variance V_v^* is equal to the mean value of $C^*(\mathbf{h})$ when the two extremities of lag vector $\mathbf{h}$ describe support v, which we denote by $\overline{C^*}(v,v)$ (Journel and Huijbregts 1978 p.78) Note: see Appendix H.5 on page 232 for a proof . Using these notations, it follows from Eqn. (4.4) to Eqn. (4.5) on page 143 that site sensitivity indices and block sensitivity indices are related by (see §4.1.6 on page 151 for a proof):

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{V_v^*}{V^*} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{\overline{C^*}(v,v)}{C^*(0)} \tag{4.8}$$

4.1.3.2 Change of support effect

Consider now that model $\mathcal{F}$ was initially developed to study the spatial average Y_v over the support v, and that after model upscaling the modeller is interested in the spatial average $Y_\mathcal{V}$ over the support $\mathcal{V}$, where $\mathcal{V} \gg v$. We know from Krige's relation (Journel and Huijbregts 1978 p.67) that the block variance V_v^* decreases with increasing size of support: $V_\mathcal{V}^* \leq V_v^*$. It follows from Eqn. (4.8) that:

$$\frac{S_Z(\mathcal{V})}{S_{\mathbf{U}}(\mathcal{V})} \leq \frac{S_Z(v)}{S_{\mathbf{U}}(v)} \tag{4.9}$$

The fraction of the variance of the aggregated model output explained by the input random field $Z(\mathbf{x})$—compared to the fraction explained by $\mathbf{U}$—is thus smaller on support $\mathcal{V}$ than on support v. More specifically, let us suppose that the covariance function $C^*(\cdot)$ of the random field $\mathbb{E}_Z Y(\mathbf{x})$ has a finite effective range and that the support v is large with respect to this range. To a first approximation, the block variance V_v^* is of the form $V_v^* \simeq V^* A^*/|v|$, where A^* is the so-called "*integral range*" of $C^*(\cdot)$ and is defined by $A^* = 1/V^* \int C^*(\mathbf{h}) d\mathbf{h}$ (Chilès and Delfiner 1999 p.73). It follows from Eqn. (4.8) that:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} \simeq \frac{|v|_c}{|v|} \quad \text{with} \quad |v|_c = A^* \cdot \frac{S_Z}{S_{\mathbf{U}}} \tag{4.10}$$

Eqn. (4.10) shows that the ratio $|v|_c/|v|$ determines the relative contribution of the model inputs $Z(\mathbf{x})$ and $\mathbf{U}$ to the output variance $\text{var}(Y_v)$. The larger that this ratio is, the larger the part of the output variance $\text{var}(Y_v)$ is that is explained by the input random field $Z(\mathbf{x})$. For a small ratio (i.e., when the area of the support v is large compared with the critical size $|v|_c$), the variability of $Z(\mathbf{x})$ is mainly "*local*", and the spatial correlation of $Z(\mathbf{x})$ over v is weak. This local variability averages over the support v when the aggregated model output Y_v is computed; hence, input 2D random field $Z(\mathbf{x})$ explains a small fraction of the output variance $\text{var}(Y_v)$. However, for a greater ratio (i.e., when the area of the support v is small compared with the critical size $|v|_c$), the spatial correlation of $Z(\mathbf{x})$ over v is strong. The averaging-out effect is weaker; hence, model input $Z(\mathbf{x})$ explains a larger fraction of the output variance $\text{var}(Y_v)$.

➤ **Extra comment:** the ratio $S_Z(\mathcal{V})/S_{\mathbf{U}}(\mathcal{V})$ of block sensitivity indices on zone $\mathcal{V}$ can also be written as a function of i) the ratio $S_Z(v)/S_{\mathbf{U}}(v)$ on zone v, and ii) the dispersion variance $D^2(v \mid \mathcal{V})$ of v within $\mathcal{V}$ for $\mathbb{E}_Z Y(\mathbf{x})$ random field:

$$\frac{S_Z(\mathcal{V})}{S_{\mathbf{U}}(\mathcal{V})} = \frac{S_Z(v)}{S_{\mathbf{U}}(v)} - \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{D^2(v \mid \mathcal{V})}{V^*} \tag{4.11}$$

See Appendix H.6 on page 233 for details.

4.1.3.3 Link between covariance function and block sensitivity indices

Critical size $|v|_c = A^* \cdot \frac{S_Z}{S_U}$ depends on the covariance function $C^*(\cdot)$ of the random field $\mathbb{E}_Z Y(\mathbf{x})$, which is itself driven by the covariance function $C(\cdot)$ of the input SRF $Z(\mathbf{x})$. Let us now assume that $Z(\mathbf{x})$ is a Gaussian random field (GRF). $\mathbb{E}_Z Y(\mathbf{x})$ is then square-integrable with respect to the standard normal density. It can be decomposed into an Hermitian expansion and its covariance function $C^*(\cdot)$ can be written as (Chilès and Delfiner 1999 p.396-399; see §4.1.7 on page 152 for a proof):

$$C^*(\mathbf{h}) = \sum_{j=0}^{\infty} \gamma_j^2 \cdot [C(\mathbf{h})]^j \tag{4.12}$$

For most of the usual transition covariance functions (e.g., spherical, exponential and Gaussian models), the covariance $C(\mathbf{h})$ is a monotically increasing function of correlation length a. In this case, it follows from Eqn. (4.12) that the integral range $A^* = 1/V^* \int C^*(\mathbf{h}) d\mathbf{h}$ also increases with correlation length a. An increase in correlation length a thus leads to an increase in the critical size $|v|_c$, and the ratio of block sensitivity indices $S_Z(v)$ and $S_U(v)$ satisfies [Eqn. (4.10) on the previous page]:

$$\frac{\partial}{\partial a}\left[\frac{S_Z(v)}{S_U(v)}\right] \geq 0 \tag{4.13}$$

The relative contribution of the uncertain model input $Z(\mathbf{x})$ to the variance of the output of interest Y_v increases when the correlation length of $Z(\mathbf{x})$ increases. Indeed, when correlation length a increases, the averaging-out effect that occurs when the model output is aggregated over spatial support v weakens; thus, the fraction of the output variance $\text{var}(Y_v)$ which is explained by the input random field $Z(\mathbf{x})$ increases.

Nugget parameter's impact on the block sensitivity indices can be interpreted in the same manner. The nugget parameter η controls the relative part of "*pure noise*" in the input random field $Z(\mathbf{x})$ [Eqn. (4.1) on page 142]. The smaller η is, the weaker the averaging-out effect will be when the block average Y_v is computed over the support v, and the larger the part of output variance $\text{var}(Y_v)$ will be that is explained by $Z(\mathbf{x})$. The critical size $|v|_c$ is thus a decreasing function of nugget parameter η, and the ratio of block sensitivity indices $S_Z(v)$ and $S_U(v)$ satisfies [Eqn. (4.1) on page 142, Eqn. (4.8), Eqn. (4.12)]:

$$\frac{\partial}{\partial \eta}\left[\frac{S_Z(v)}{S_U(v)}\right] \leq 0 \tag{4.14}$$

➤ **Extra comment (limits):** Based on Eqn. (4.8) on the previous page, we can also write the ratio of block sensitivity indices $S_Z(v)$ and $S_U(v)$ for two limit situations. First, when the covariance function $C(\cdot)$ of random field $Z(\mathbf{x})$ is constant over Ω and non-null, then the ratio satisfies:

$$\frac{S_Z(v)}{S_U(v)} = \frac{S_Z}{S_U} \tag{4.15}$$

Next, when $Z(\mathbf{x})$ is a random field without spatial auto-correlation, that is, when its covariance function $C(\mathbf{h})$ is null except for $\mathbf{h} = \mathbf{0}$, then the ratio of block sensitivity indices verifies:

$$\frac{S_Z(v)}{S_U(v)} = 0 \tag{4.16}$$

See Appendix H.9 on page 234 and H.10 on page 234 for proofs and details.

4.1.3.4 Illustrative example

To illustrate the change of support effects on sensitivity analysis results, we performed spatial VB-GSA on our numerical example in the following settings: varying disc-shaped support v of increasing size (Figure 4.2 on the following page); varying correlation length from $a = 1$ to $a = 10$ (Figure 4.3 on page 149); varying nugget parameter from $\eta = 0$ to $\eta = 0.9$ (Figure 4.4 on page 149). For each setting, we computed estimates of the output variance $\text{var}(Y_v)$, the block sensitivity indices $S_{\mathbf{U}}(v)$, $S_Z(v)$ and the ratio $S_Z(v)/S_{\mathbf{U}}(v)$ over $N_{\text{tot}} = 4096$ model runs. Mean values with a 95% confidence interval were then computed for each estimate using bootstrapping (100 replicas). In accordance with Eqn. (4.9) to Eqn. (4.14) on pages 145–146, it appears that the block sensitivity index $S_Z(v)$ (i) decreases when the support v increases (Figure 4.2 (b)), (ii) increases with the correlation length a (Figure 4.3 (b)), and (iii) decreases with the nugget parameter η (Figure 4.4 (b)). The opposite trends are observed for sensitivity index $S_{\mathbf{U}}(v)$. The change of support effect is clearly highlighted in Figure 4.2 (b): the contribution of the economic parameters U_1, U_2 and U_3 to the variance of total flood damage $|v| \cdot Y_v$ exceeds the contribution of the water level map $Z(\mathbf{x})$ when the radius r of v is greater than $r_c \simeq 18$; for radius $r < r_c$, the variance of total flood damage over the support v is mainly explained by the variability of the water levels $Z(\mathbf{x})$. Finally, Figure 4.2 on the next page(c) shows that the ratio $S_Z(v)/S_{\mathbf{U}}(v)$ is proportional to $1/|v|$ when the support v is large enough. The theoretical curve $S_Z(v)/S_{\mathbf{U}}(v) = |v|_c/|v|$ [Eqn. (4.10) on page 145] was fitted (least squares - $R^2 = 0.99$) on data points (for $r \geq 20$ only), yielding an estimate of the critical size $|v|_c \simeq 1,068$. All calculations and figures were realized in R (R Development Core Team 2009): random realisations of $Z(\mathbf{x})$ were generated with the `GaussRF()` function from the `RandomFields` package (Schlather 2001), while computation of sensitivity indices was based on a modified version of the `sobol()` function from the `sensitivity` package.

➤ **Another illustrative example:** these results are corroborated by the outcomes of the second analytical test case we investigated (published in the proceedings of the *IAMG 2011* conference). In addition, we also observed in that test case that the block sensitivity index $S_Z(v)$ increased with the variance $C(0)$ of Gaussian Random Field $Z(\mathbf{x})$. Indeed, in the expression of the critical size $|v|_c$ [Eqn. (4.10) on page 145], the integral range A^* does not depend on the variance $C(0)$, but the site sensitivity index S_Z does. See Appendices for details.

4.1.4 Discussion

Our first goal was to provide a formalism that extends the VB-GSA approach to spatial models when the modeller is mainly interested in the linear average or the sum of a point-based model output $Y(\mathbf{x})$ over some spatial unit v. Our approach is strongly motivated by various prior publications. Other authors had already computed site sensitivity indices (Marrel et al. 2011; Pettit and Wilson 2010) and block sensitivity indices (Lilburne and Tarantola 2009), but did so without naming them or exploring their analytical properties or their relationship. Our work is an attempt to do so. Eqn. (4.8) on page 145 provides an exact relation between the site and block sensitivity indices, it may prove useful in the case of a model with a simple enough analytical expression.

Our research also sought to account for the change of support effects in the propagation of uncertainty through spatial models, within a VB-GSA framework. We proved that the fraction of the variance of the model output that is explained by a spatially distributed model input $Z(\mathbf{x})$ decreases under model upscaling; when the support v is large enough, the ratio of the block sensitivity index of spatially dis-

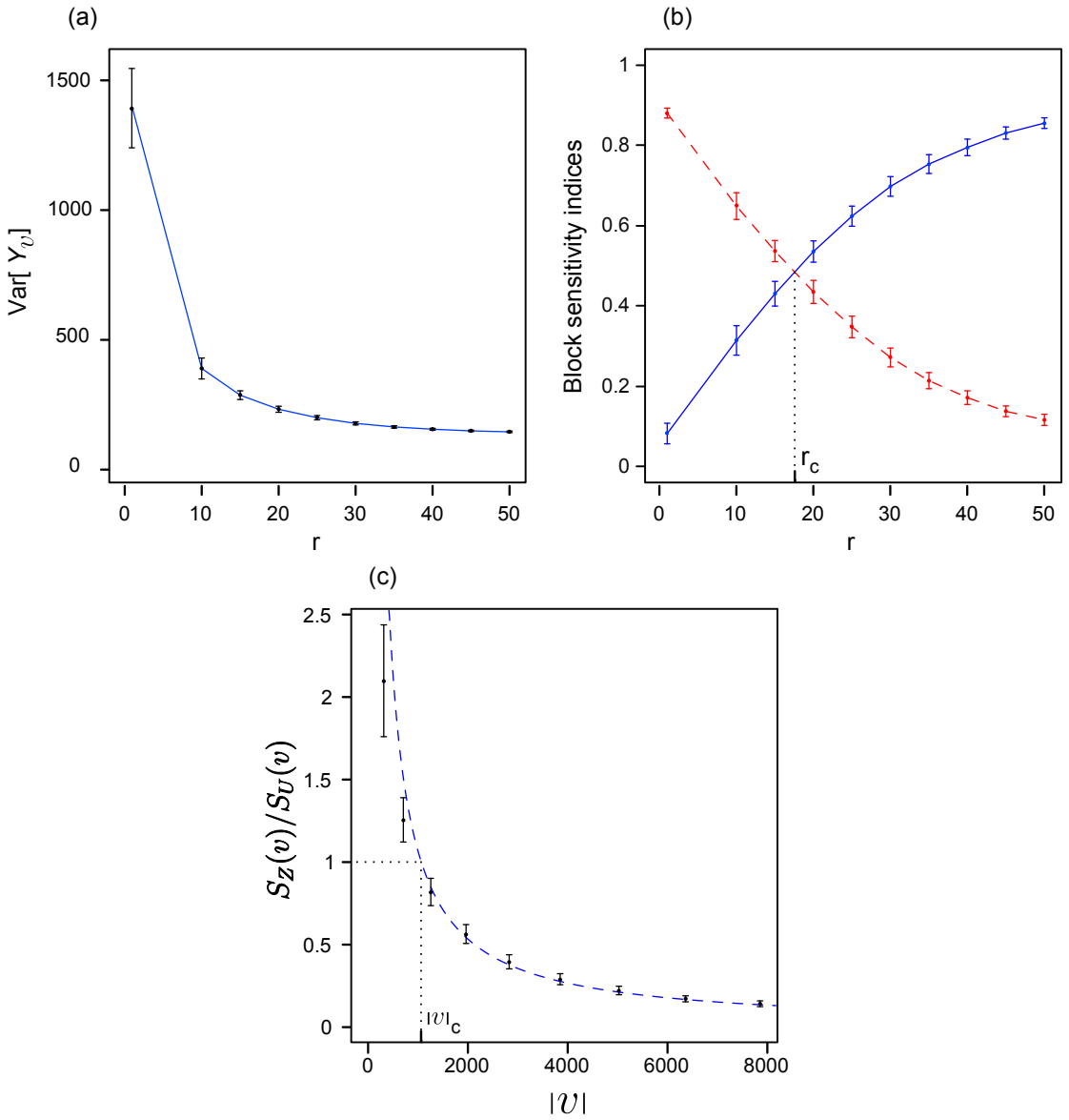

Figure 4.2: VB-GSA results depending on the size of disc-shaped support ν (with radius r and area $|\nu| = \pi r^2$), for $a = 5$, $\eta = 0.1$: **(a)** total variance of Y_v, **(b)** block sensitivity indices $S_{\mathbf{U}}(v)$ (solid line) and $S_Z(v)$ (dashed line), **(c)** ratio $S_Z(v)/S_{\mathbf{U}}(v)$ with fitted curve $S_Z(v)/S_{\mathbf{U}}(v) = |v|_c/|v|$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

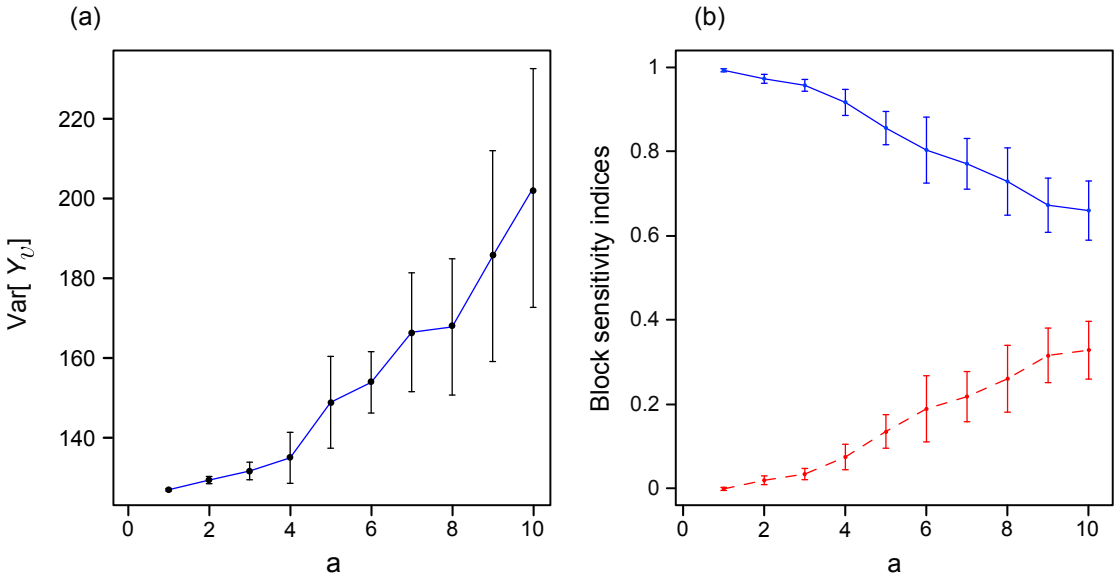

Figure 4.3: VB-GSA results depending on correlation length a, for $\eta = 0.1$ and a disc-shaped support v of radius $r = 50$: **(a)** total variance of Y_v, **(b)** block sensitivity indices $S_{\mathbf{U}}(v)$ (solid line) and $S_Z(v)$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

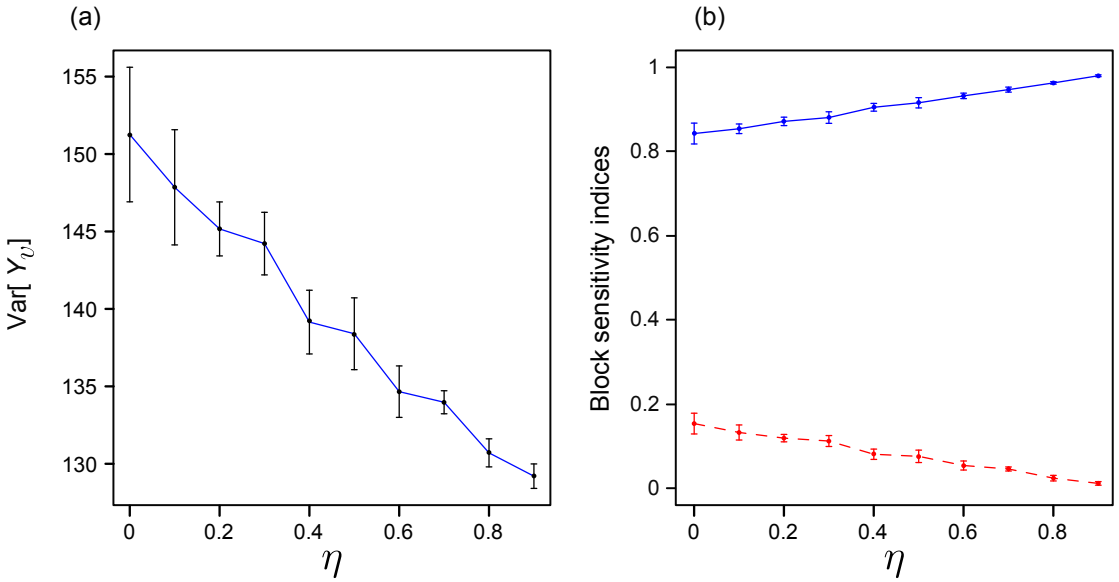

Figure 4.4: VB-GSA results depending on covariance nugget parameter η, for $a = 5$ and a disc-shaped support v of radius $r = 50$: **(a)** total variance of Y_v, **(b)** block sensitivity indices $S_{\mathbf{U}}(v)$ (solid line) and $S_Z(v)$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

tributed input to the block sensitivity index of non-spatial inputs is proportional to $|v|_c/|v|$. The critical size $|v|_c$ depends on the covariance function of the input SRF $Z(\mathbf{x})$; it usually increases with an increase of the correlation length a or a decrease of the nugget parameter η. These findings are a translation into VB-GSA formalism of the averaging-out effect clearly exhibited by Journel and Huijbregts (1978) in the regularization theory. Our contribution is to discuss this issue from the point of view of VB-GSA practitioners. Formalizing the effect of a change of support on sensitivity analysis results may help modellers when they consider model upscaling; it will orientate future data gathering by identifying model inputs that will explain the largest fraction of the variance of the model output over a new spatial support. Our contribution also promotes an increased awareness of the issue of sharing out efficiently, among the various inputs used by a complex computer code, the cost of collecting field data. At some point of the model building process, the modeller will usually aim at reducing the variance of the output below a given threshold, that will depend on the model use. To do so, the modeller may have to improve his knowledge on the "*real*" value of some of the model inputs, usually by collecting extra data. In this case, gathering extra field data on inputs maps that have small sensitivity indices ($S_Z(v) < 0.1$) would be unefficient, as it would be costly but could not reduce the variance of the model output by a large fraction. Saint-Geours et al. (2011a) discuss this issue on a flood risk assessment case study.

It should be noted that our approach is based on conditions that may not be met in some practical cases. First, we considered a model $\mathcal{F}$ with a single spatially distributed input $Z(\mathbf{x})$. In real applications, modellers may have to deal with several spatial inputs $Z_1(\mathbf{x}), \ldots, Z_p(\mathbf{x})$, with different covariance functions $C_i(\cdot)$, correlation lengths a_i and nugget parameters η_i. In this case, it can be shown that Eqn. (4.8) on page 145 still holds separately for each spatial input $Z_i(\mathbf{x})$. However, no conclusion can be drawn "*a priori*" regarding how a change of support affects the relative ranking of two spatial inputs $Z_i(\mathbf{x})$ and $Z_j(\mathbf{x})$; the ratio of their block sensitivity indices $S_{Z_i}(v)/S_{Z_j}(v)$ will depend on the ratio of block variances $V_{v,i}/V_{v,j}$. Second, some environmental models are not point-based and involve spatial interactions (e.g., erosion and groundwater flow models). In this case, it still may be possible to build a point-based surrogate model as a coarse approximation of the original model; if not, then the change of support properties discussed in §4.1.3 may not hold. Third, we assumed the input random field $Z(\mathbf{x})$ to be stationary; if it is not, site sensitivity indices depend on site $\mathbf{x}^\star$ [Eqn. (4.4) on page 143]. It is then possible to compute maps of these indices (Marrel et al. 2011; Pettit and Wilson 2010) to discuss the spatial variability of model inputs sensitivities.

Finally, we focused on the case in which the modeller's interest is in the spatial linear average or the sum of model output $Y(\mathbf{x})$ over the support v. As discussed by Lilburne and Tarantola (2009), other outputs of interest may be considered, such as the maximum value of $Y(\mathbf{x})$ over v (e.g., maximal pollutant concentration over a zone), some quantile of $Y(\mathbf{x})$ over v (Heuvelink et al. 2010b), or the percentage of v for which $Y(\mathbf{x})$ exceeds a certain threshold. To our knowledge, no study has investigated the properties of sensitivity indices computed with respect to such outputs of interest.

4.1.5 Conclusion

This paper provides a formalism to apply variance-based global sensitivity analysis to spatial models when the modeller's interest is in the average or the sum of the model output $Y(\mathbf{x})$ over a given spatial unit v. Site sensitivity indices and block sensitivity indices allow us to discuss how a change of support modifies the relative contribution of uncertain model inputs to the variance of the output of interest. We demonstrate an analytical relationship between these two types of sensitivity indices. Our results show that the block sensitivity index of an input random field $Z(\mathbf{x})$ increases with the ratio $|v|_c/|v|$, where $|v|$ is the area of the spatial support v and the critical size $|v|_c$ depends on the covariance function of $Z(\mathbf{x})$. Our formalization is made with a view toward promoting the use of sensitivity analysis in model-based spatial

decision support systems. Nevertheless, further research is needed to explore the case of non-point-based models and extend our study to outputs of interest other than the average value of model output over support v.

4.1.6 Appendix A: Proof of the relation between site sensitivity indices and block sensitivity indices

➤ **Note to the reader:** This subsection was published in (Saint-Geours et al. 2012) as an appendix.

As mentioned in §4.1.2.1 on page 141, we assume that the first two moments of $Y(\mathbf{x})$ exist. The ratio of block sensitivity indices gives [Eqn. (4.5) on page 143]:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{\mathrm{var}\left(\mathbb{E}\left[Y_v \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}\right]\right)}{\mathrm{var}\left(\mathbb{E}\left[Y_v \mid \mathbf{U}\right]\right)} \tag{4.17}$$

The conditional expectation of block average Y_v given $Z(\mathbf{x})$ gives:

$$\begin{aligned}
\mathbb{E}\left[Y_v \mid Z\right] &= \mathbb{E}\left[\left(1/|v| \int_v Y(\mathbf{x}) d\mathbf{x}\right) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}\right] \quad \text{(definition of } Y_v\text{)} \\
&= 1/|v| \int_v \mathbb{E}\left[Y(\mathbf{x}) \mid Z(\mathbf{x})\right] d\mathbf{x} \quad \text{(for a point-based model)} \\
&= 1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x}) d\mathbf{x} \quad \text{(definition of } \mathbb{E}_Z Y(\mathbf{x})\text{)}
\end{aligned} \tag{4.18}$$

Thus we have $\mathrm{var}\left(\mathbb{E}\left[Y_v \mid Z\right]\right) = \mathrm{var}\left(1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x}) d\mathbf{x}\right) = V_v^*$ (definition of V_v^*). Moreover, the conditional expectation of block average Y_v given input $\mathbf{U}$ gives:

$$\begin{aligned}
\mathbb{E}\left[Y_v \mid \mathbf{U}\right] &= \mathbb{E}\left[\left(1/|v| \int_v Y(\mathbf{x})\, d\mathbf{x}\right) \mid \mathbf{U}\right] \quad \text{(definition of } Y_v\text{)} \\
&= 1/|v| \int_v \mathbb{E}\left[Y(\mathbf{x}) \mid \mathbf{U}\right] d\mathbf{x} \quad \text{(Fubini's theorem)}
\end{aligned} \tag{4.19}$$

$\mathbb{E}\left[Y(\mathbf{x}) \mid \mathbf{U}\right]$ does not depend on site $\mathbf{x}$ under the stationarity of SRF $Z(\mathbf{x})$; thus, we have in particular $\mathbb{E}\left[Y_v \mid \mathbf{U}\right] = \mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right]$, and $\mathrm{var}\left(\mathbb{E}\left[Y_v \mid \mathbf{U}\right]\right) = \mathrm{var}\left(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right]\right)$. Combining these expressions with Eqn. (4.17) yields:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{V_v^*}{\mathrm{var}(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right])} \tag{4.20}$$

The ratio of site sensitivity indices gives [Eqn. (4.4) on page 143]:

$$\cdot\frac{S_Z}{S_{\mathbf{U}}} = \frac{\mathrm{var}(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}\right])}{\mathrm{var}(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right])} \tag{4.21}$$

We notice that for point-based models $\mathrm{var}\left[\mathbb{E}(Y(\mathbf{x}^*) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}\right] = \mathrm{var}\left[\mathbb{E}_Z Y(\mathbf{x}^*)\right] = V^*$ (definition of $\mathbb{E}_Z Y(\mathbf{x})$ [Eqn. (4.7) on page 144)]). Finally, it follows from Eqn. (4.20) and Eqn. (4.21) that:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{V_v^*}{V^*} \tag{4.22}$$

4.1.7 Appendix B: Hermitian expansion of random field $\mathbb{E}_Z Y(\mathbf{x})$

➤ **Note to the reader:** This subsection was published in (Saint-Geours et al. 2012) as an appendix.

The random field $\mathbb{E}_Z Y(\mathbf{x})$ can be written [Eqn. (4.2) on page 142, Eqn. (4.7) on page 144] as a transformation of the Gaussian random field $Z(\mathbf{x})$ through the function $\bar{\mathcal{F}}_{\text{loc}} : z \mapsto \int_{\mathbb{R}^k} \mathcal{F}_{\text{loc}}(u, z) \cdot p_U(u)\, du$:

$$\mathbb{E}_Z Y = \bar{\mathcal{F}}_{\text{loc}}(Z)$$

where $p_U(\cdot)$ is the multivariate pdf of random vector $\mathbf{U}$. Under the hypothesis that the first two moments of $Y(\mathbf{x})$ exist, random field $\mathbb{E}_Z Y(\mathbf{x})$ has finite expected value and finite variance. Thus, $\bar{\mathcal{F}}_{\text{loc}}$ belongs to the Hilbert space $L^2(\mathcal{N})$ of functions $\mathcal{F}_{\text{loc}} : \mathbb{R} \to \mathbb{R}$, which are square-integrable with respect to Gaussian density $n(.)$. Hence, $\bar{\mathcal{F}}_{\text{loc}}$ can be expanded on the sequence of Hermite polynomials $(\chi_j)_{j\in\mathbb{N}}$, which forms an orthonormal basis of $L^2(\mathcal{N})$ (Chilès and Delfiner 1999 p.399):

$$\bar{\mathcal{F}}_{\text{loc}} = \sum_{j=0}^{\infty} \pi_j \cdot \chi_j \quad \text{with} \quad \chi_j(z) = \frac{1}{\sqrt{j!}} \cdot \frac{1}{n(z)} \cdot \frac{\partial^j}{\partial z^j} n(z)$$

where coefficients π_j are given by: $\pi_j = \int_{\mathbb{R}} \chi_j(z) \bar{\mathcal{F}}_{\text{loc}}(z) n(z)\, dz$.

➤ **Note to the reader:** See Appendix §H on page 230, properties H.2 and H.3 for a more detailed proof.

It follows that $\mathbb{E}_Z Y(\mathbf{x})$ can be written as an infinite expansion of polynomials of $Z(\mathbf{x})$:

$$\forall \mathbf{x} \in \Omega, \quad \mathbb{E}_Z Y(\mathbf{x}) = \sum_{j=0}^{\infty} \pi_j \cdot \chi_j [Z(\mathbf{x})]$$

Its covariance then gives (Chilès and Delfiner 1999 p.396, Eqn.(6.23) and p.399, Eqn.(6.25)):

$$\text{cov}\left(\mathbb{E}_Z Y(\mathbf{x}), \mathbb{E}_Z Y(\mathbf{x}+\mathbf{h})\right) = \sum_{j=0}^{\infty} \pi_j^2 \cdot \left[\frac{C(\mathbf{h})}{C(0)}\right]^j = \sum_{j=0}^{\infty} \gamma_j^2 \cdot [C(\mathbf{h})]^j \tag{4.23}$$

where $C(\mathbf{h})$ is the covariance function of GRF $Z(\mathbf{x})$ and $\gamma_j = \pi_j \cdot C(0)^{-j/2}$.

➤ **Note to the reader:** See Appendix H.4 on page 231 for a more detailed proof.

4.2 Influence of spacing

4.2.1 Introduction

In this section, we investigate how the results of VB-GSA of a spatial model may be influenced by another component of the "*scale triplet*": the "*spacing*". As discussed in the introduction of this chapter, we limit our study to the case of point-based and spatially additive models only.

Let consider a model $\mathcal{F}$, in which the quantity of interest is the aggregated value of model output $Y(\mathbf{x})$ over the entire spatial domain Ω. In many applications, this aggregated output Y_Ω is approximated by a weighted sum of the output values $Y(\mathbf{x}_i)$ computed at a finite number of points $\mathbf{x}_i \in \Omega$. This approximation may result in errors when estimating sensitivity indices of the model inputs with respect to the aggregated output of interest. The caracteristic distance between the data points $\mathbf{x}_i$ is related to the notion of "*spacing*" as presented in Chapter 1 (§1.1.3 on page 20). Intuitively, we understand that the approximation error on sensitivity indices will be smaller when the number of data points x_i is large (small spacing), and on the contrary will be larger when there are few data points x_i (large spacing) (Skøien and Blöschl 2006).

The purpose of this section is to give an expression of these approximation errors. We investigate this issue using the same analytical test case as the one described in the previous section §4.1, in which the spatially distributed input $Z(\mathbf{x})$ is modelled as a stationary random field. Besides, we consider the limitative case in which data points $\mathbf{x}_i$ are uniformly positioned on a regular square grid.

4.2.2 Model description

Let consider the same point-based and spatially additive model $\mathcal{F}$ that was already presented in the previous section §4.1.2.1 on page 141:

$$\forall \mathbf{x} \in \Omega\,, \quad Y(\mathbf{x}) = \mathcal{F}_{\text{loc}}\left[\mathbf{U}, Z(\mathbf{x})\right] \tag{4.24}$$

in which $\mathbf{U} = (U_1, \ldots, U_k)$ is a random vector, $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is a second-order stationary random field with covariance function $C(\cdot)$, and $\mathcal{F}_{\text{loc}}$ is a mapping from $\mathbb{R}^k \times \mathbb{R}$ to $\mathbb{R}$.

In addition, we will assume that $\Omega \in \mathbb{R}^2$ is a fixed square-shaped domain and denote by $|\Omega|$ its surface area. We also assume that Ω is covered by a set of points $\mathbf{x}_i$, uniformly positioned on the nodes of a regular square grid as shown in Figure 4.5 on the following page. The distance between two neighbouring points $\mathbf{x}_i$ and $\mathbf{x}_j$ is denoted by s and refered to as the "*spacing*" of the set of points, according to the definitions discussed in §1.1.3 on page 20. We denote by G the total number of points $\mathbf{x}_i$: it is equal to $G = |\Omega|/s^2$.

Finally, we will consider as an output of interest the average value Y_Ω of the output random field $Y(\mathbf{x})$ over the entire spatial domain Ω:

$$Y_\Omega = \frac{1}{|\Omega|} \int_{\mathbf{x}\in\Omega} Y(\mathbf{x})\, d\mathbf{x} \tag{4.25}$$

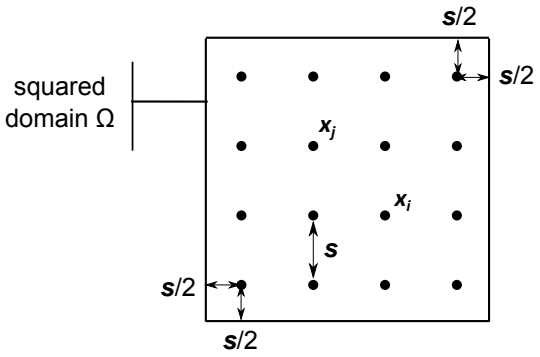

Figure 4.5: Spatial domain Ω, regularly positionned points $\mathbf{x}_i$, spacing s (total number of points: $G = |\Omega|/s^2$)

4.2.3 Approximation on a grid of points

The aggregated output of interest Y_Ω can be approximated by the empirical mean of values $Y(\mathbf{x}_i)$ at each point $\mathbf{x}_i$ of the grid. We denote by $\tilde{Y}_\Omega$ and $\mathbb{E}_Z\tilde{Y}_\Omega$ the random variables defined by:

$$\tilde{Y}_\Omega = \frac{1}{G}\sum_{i=1}^{G} Y(\mathbf{x}_i) \qquad \text{and} \qquad \mathbb{E}_Z\tilde{Y}_\Omega = \frac{1}{G}\sum_{i=1}^{G} \mathbb{E}_Z Y(\mathbf{x}_i) \tag{4.26}$$

Let also denote by $\tilde{V}_\Omega^*$ the variance of $\mathbb{E}_Z\tilde{Y}_\Omega$: $\tilde{V}_\Omega^*$ can be used as an approximation of block variance V_Ω^* of random field $\mathbb{E}_Z Y(\mathbf{x})$ (see §4.1.3.1 on page 144 for a definition of V_Ω^*). To obtain an approximation of block sensitivity indices $S_\mathbf{U}(\Omega)$ and $S_Z(\Omega)$ defined with respect to the output of interest Y_Ω, we can calculate the first-order sensitivity indices of model inputs $\mathbf{U}$ and Z with respect to the approximated output of interest $\tilde{Y}_\Omega$: let denote by $\tilde{S}_\mathbf{U}(\Omega)$ and $\tilde{S}_Z(\Omega)$ these indices. We can show that the ratio of these proxies for block sensitivity indices writes (see Appendix §J on page 242 for a proof):

$$\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_\mathbf{U}(\Omega)} = \frac{S_Z}{S_\mathbf{U}} \cdot \frac{\tilde{V}_\Omega^*}{V^*} \tag{4.27}$$

in which S_Z and $S_\mathbf{U}$ are site sensitivity indices and V^* is the variance of $\mathbb{E}_Z Y(\mathbf{x})$. This relation is similar to Eqn. (4.8) on page 145, except that V_Ω^* and block sensitivity indices $S_Z(\Omega)$ and $S_\mathbf{U}(\Omega)$ have been replaced by their gridded approximation.

4.2.4 Expression of the approximation error and convergence

We want to study the difference between: i) the true values $S_\mathbf{U}(\Omega)$ and $S_Z(\Omega)$ of block sensitivity indices defined with respect to the aggregated output of interest Y_Ω, and ii) the approximated values $\tilde{S}_\mathbf{U}(\Omega)$ and $\tilde{S}_Z(\Omega)$ defined with respect to the proxi output $\tilde{Y}_\Omega$. Using Eqn. (4.27) and the expression of the ratio of sensitivity indices $S_\mathbf{U}(\Omega)$ and $S_\mathbf{U}(\Omega)$ given in the previous section [Eqn. (4.8) on page 145], we obtain an expression of the approximation error:

$$\frac{S_Z(\Omega)}{S_\mathbf{U}(\Omega)} - \frac{\tilde{S}_Z(\Omega)}{\tilde{S}_\mathbf{U}(\Omega)} = \frac{S_Z}{S_\mathbf{U}} \cdot \frac{V_\Omega^* - \tilde{V}_\Omega^*}{V^*} \tag{4.28}$$

Let denote by ϵ the error made when approximating block variance V_Ω^* with $\tilde{V}_\Omega^*$:

$$\epsilon = V_\Omega^* - \tilde{V}_\Omega^* \tag{4.29}$$

The error ϵ depends on the spacing s of the set of point $\mathbf{x}_i$. It can be expressed as a function of the total number of points $G = |\Omega|/s^2$ and of the covariance structure $C^*(\cdot)$ of random field $\mathbb{E}_Z Y(\mathbf{x})$ (see Appendix J.2 on page 243 for a proof):

$$\epsilon = \left[\frac{C^{*,+}(0) - C^*(0)}{G}\right] + \left[\overline{C^*}(\Omega,\Omega) - \frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] \tag{4.30}$$

where $C^{*,+}(\cdot)$ is a modified version of the covariance function $C^*(\cdot)$ of random field $\mathbb{E}_Z Y(\mathbf{x})$, continuous at the origin $h = 0$:

$$C^{*,+}(h) = \begin{cases} C^*(h) & \forall h > 0 \\ \lim\limits_{h\to 0^+} C^*(h) & \text{if } h = 0 \end{cases} \tag{4.31}$$

The first part of the approximation error ϵ is known as the "*zero-effect*" described by Journel and Huijbregts (1978 p.96). For most usual spatial covariance structures of random fields, the inequality $C^*(0) \geq \lim\limits_{h\to 0^+} C^*(h)$ holds (nugget effect). Hence, the first part of the approximation error ϵ is negative and the "*zero-effect*" results in an over-estimation of block variance V_Ω^*. The second part of ϵ is the error made when approximating the double integral $\overline{C^*}(\Omega,\Omega) = 1/|\Omega|^2 \iint\limits_{\mathbf{x},\mathbf{x}'\epsilon v^2} C^*(\mathbf{x}-\mathbf{x}')\,d\mathbf{x}\,d\mathbf{x}'$ with a double Riemann sum over the set of points $\mathbf{x}_i$. Generally speaking, it is not possible to know whether ϵ is positive or negative, that is, if block variance V_Ω^* is under-estimated or over-estimated by proxy $\tilde{V}_\Omega^*$.

Nevertheless, from Eqn. (4.28) on the facing page and the expression of error ϵ [Eqn. (4.30)], we can show that the ratio of proxies of sensitivity indices $\tilde{S}_Z(\Omega)/\tilde{S}_{\mathbf{U}}(\Omega)$ converges toward the ratio of exact sensitivity indices when the number of points $\mathbf{x}_i$ grows, that is, when the spacing s of the set of points tends to zero (see Appendix J.3 on page 244 for a proof):

$$\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_{\mathbf{U}}(\Omega)} \underset{s\to 0}{=} \frac{S_Z(\Omega)}{S_{\mathbf{U}}(\Omega)} + \mathcal{O}(s) \tag{4.32}$$

In the particular case in which there is no spatial correlation in the spatially distributed input $Z(\mathbf{x})$, that is, if the covariance function $C^*(h)$ is null except for $h = 0$, then the approximation error ϵ is equal to $\epsilon = -s^2 \cdot C^*(0)/|\Omega|$.

4.2.5 Conclusion

As mentioned in the introduction, this study is only valid for the point-based and spatially additive model $\mathcal{F}$ described in §4.1.2.1 on page 141, in which the spatially distributed input $Z(\mathbf{x})$ is modelled as a stationary random field. We investigated the case in which the aggregated output of interest Y_Ω is approximated over a set of points $\mathbf{x}_i$ regularly positionned on a square grid. This results in an approximation error on block sensitivity indices $S_{\mathbf{U}}(\Omega)$ and $S_Z(\Omega)$. This error converges to 0 when the number of points grows. The speed of convergence is $\mathcal{O}(s)$ where s is the "*spacing*" of the set of points $\mathbf{x}_i$, that is, the distance between two neighbouring points $\mathbf{x}_i$ and $\mathbf{x}_j$ on the grid.

4.3 Application to the NOE code on the Orb Delta case study

4.3.1 Introduction

In this section, we offer to illustrate on the NOE modelling framework on the Orb Delta case study some of the scale issues related to VB-GSA of spatially additive and point-based models. We will focus on the influence of "*support*", which we theoretically described in §4.1. As mentioned in Chapter 2, the NOE code is a point-based and spatially additive model with two spatially distributed inputs: the assets map and the set of hazard maps (water depths). Our purpose in this section is to answer the following question on the Orb Delta case study: *how do the results of the uncertainty and sensitivity analysis depend on the spatial support v over which the NOE model outputs (damage estimates, AAD and ΔAAD indicators) are aggregated?*

This study is a supplement to the first sensitivity analysis of the NOE modelling framework on the Orb Delta presented in Chapter 3 (§3.3 on page 111). This extra study only differs from the previous one by the set of outputs of interest, which is detailed in the next subsection §4.3.2: sensitivity indices of the NOE model inputs are now computed with respect to the sum of the ΔAAD indicator over different spatial supports v of increasing sizes. Besides, we also build maps of ΔAAD uncertainty as well as maps of sensitivity indices at different cell sizes. The results (§4.3.3) show that i) uncertainty on the ΔAAD indicator is not spatially homogeneous; ii) the maps of sensitivity indices give a new insight on the spatial distribution of the influence of each model input on the variance of model output; and iii) the sensitivity indices of spatially distributed inputs (assets map, hazard maps) decrease with the size of support v. We discuss the main outcomes of this case study and its limits in §4.3.4.

4.3.2 Methods

4.3.2.1 Overview

Our goal is to investigate the sensitivity of the NOE model inputs with respect to the aggregated value of the NOE output indicators over various spatial supports v. For the sake of clarity, we limit our study to the ΔAAD indicator (average annual avoided damages, §2.2.5 on page 44). To perform this sensitivity analysis, we follow the same general procedure as the one used in the first analysis of the NOE code presented in §3.3 on page 111. We consider the same sources of uncertainty and their modelling remains unchanged (§3.3.3.1 on page 117), except for the uncertain costs of the flood management plan, which are ignored here because they are not used in the computation of the ΔAAD indicator. The pseudo Monte Carlo procedure for the estimation of sensitivity indices is also the same, with a total sample size $N_{\text{tot}} = 28\,672$ (§3.3.3.2 on page 124). Sensitivity indices of spatially distributed inputs (the hazard maps and the assets map) are computed following the "*map labelling*" method (§3.1.2.4 on page 84).

The only difference with the first analysis presented in Chapter 3 is the set of outputs of interest we consider: we will use two different settings in our analysis, which will be refered to as "*setting A*" and "*setting B*". Both settings A and B aim at answering the same question, in a different way: *how does the uncertainty on the ΔAAD indicator, and the associated sensitivity indices, depend on the spatial support v over which it is aggregated?*

Setting A is meant only to be *illustrative*: we consider three individual spatial supports randomly selected, of increasing surface area, and show how uncertainty on the ΔAAD indicator and related

sensitivity indices vary from one support to another. This setting A is just a simplified example that aim at making our point more understandable;

Setting B. The goal of Setting B is to carry out a more rigorous analysis. The output of interest we consider is the map of the ΔAAD indicator, first transformed into a number of grids (raster data) of increasing cell sizes (i.e., of increasing spatial support). For each cell size, we compute maps of ΔAAD uncertainty as well as maps of associated sensitivity indices. We then study how these maps vary for increasing cell size (i.e., for increasing spatial support).

In the following subsections §4.3.2.2 and §4.3.2.3 we give a detailed description of settings A and B.

4.3.2.2 Setting A: ΔAAD indicator on individual supports

In a first setting, we consider three individual spatial supports v_1, v_2 and v_3 of increasing surface area (Figure 4.6 on the next page). These three supports were selected at random, and their study is only meant to be illustrative. By way of example, we chose to consider spatial supports that are related to real-world entities: first support v_1 is a single house located on the western bank of the Orb Delta ($|v_1| \leq 1$ ha), second support v_2 is the administrative district of Sauvian ($|v_2| = 13$ sq. km), and third support v_3 is the entire floodplain ($|v_3| = 63$ sq. km). For each spatial support, the output of interest is the sum of the ΔAAD indicator over this support, denoted by ΔAAD_{v_1}, ΔAAD_{v_2} and ΔAAD_{v_3}, respectively. These "*local*" indicators can be used to assess whether a given area v (a house, a district, etc.) will benefit or suffer from the flood control measures on the Orb Delta: if ΔAAD_v is positive (resp. negative), the flood risk management plan under study will result in a decrease (resp. increase) of the average annual damages on the investigated support v. The nominal values of the ΔAAD_v indicator—produced using the nominal values of the NOE model inputs—aggregated over spatial supports v_1, v_2 and v_3 are as follows:

ΔAAD_{v_1}	1.725 k€/year
ΔAAD_{v_2}	221.3 k€/year
ΔAAD_{v_3}	6 523 k€/year

Total-order sensitivity indices are then computed with respect to each of these three outputs of interest ΔAAD_{v_1}, ΔAAD_{v_2} and ΔAAD_{v_3}. They will be denoted by $ST_i(\text{house})$, $ST_i(\text{district})$ and $ST_i(\text{floodplain})$, respectively.[a]

4.3.2.3 Setting B: raster maps of the ΔAAD indicator

Raster maps of the ΔAAD indicator In a second setting, the output of interest we consider is the map of the ΔAAD indicator. This map is first transformed into a number of grids (raster maps) of increasing cell sizes: as previously explained in §2.3.2.4 on page 65, an ΔAAD grid is obtained by computing at each cell c_i the sum of the Average Annual Avoided Damages over all assets (or parts of assets) contained in the cell.[b] We consider four different grids with cells of 200 m by 200 m, 400 m by 400 m, 800 m by 800 m and 1 600 m by 1 600 m, with corresponding cell sizes $|c| = 0.04$, 0.16, 0.64 and 2.56 sq. km, respectively. Figure 4.7 on page 159 shows two ΔAAD raster maps produced using the nominal values of the NOE model inputs, for cell size $|c| = 0.04$ sq. km and $|c| = 2.56$ sq. km, respectively.

[a] We chose to calculate total-order rather than first-order sensitivity indices because their confidence bounds computed by bootstrap proved to be narrower.

[b] If an asset has a large surface area and overlaps many cells of the grid, then the value of the ΔAAD indicator over this asset is shared out among the cells in proportion to the overlaped areas.

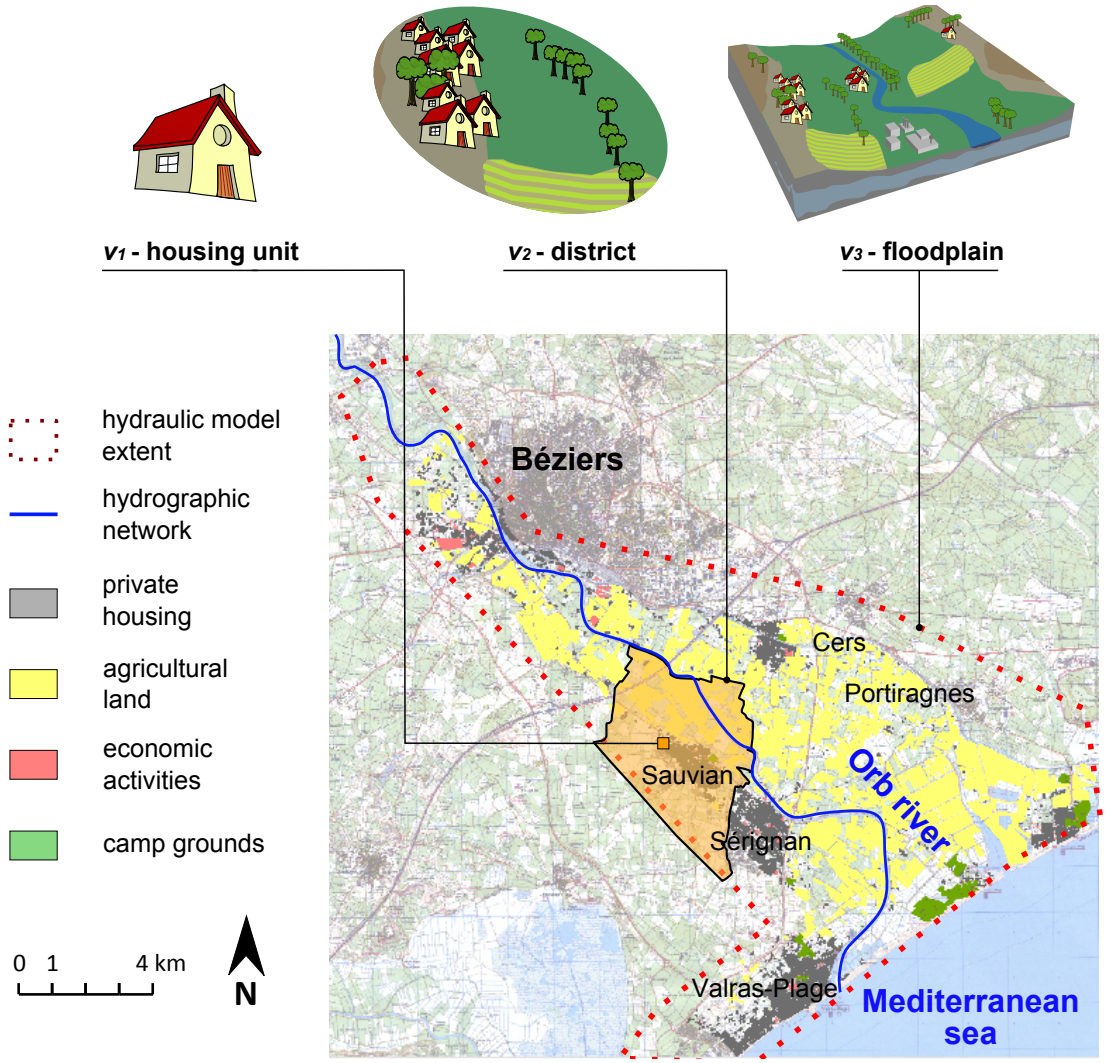

Figure 4.6: Setting A with three different spatial supports for the ΔAAD indicator: house, district, and floodplain

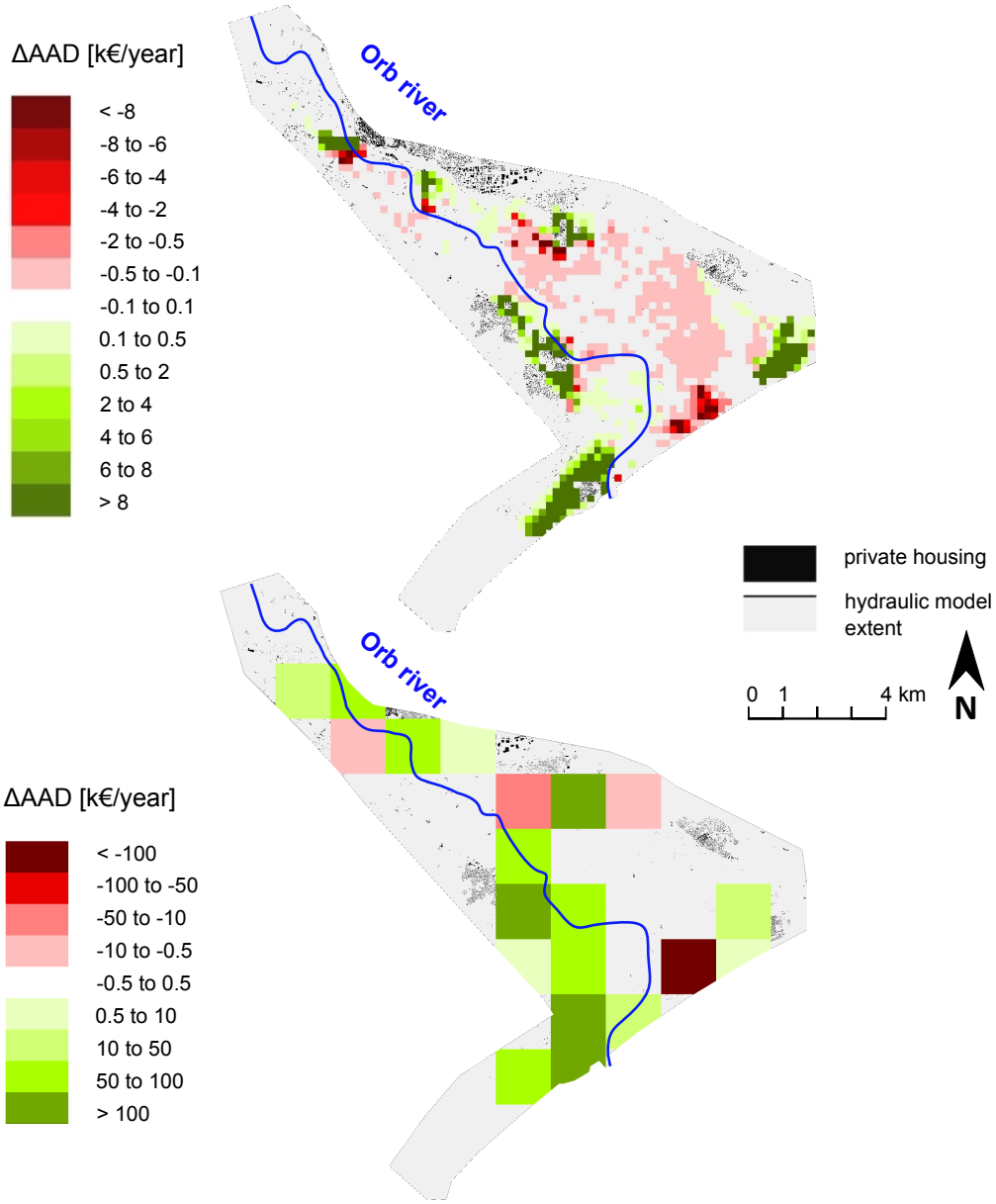

Figure 4.7: Setting B: ΔAAD grids for nominal values of the model inputs, cell size $|c|$ = 0.04 sq. km (top) and $|c|$ = 2.56 sq. km (bottom)

Maps of sensitivity indices In this second setting, maps of sensitivity indices are then produced by computing total-order sensitivity indices with respect to the value of the ΔAAD indicator at each cell of the various ΔAAD raster maps considered. More precisely, for each cell size $|c|$ = 0.04, 0.16, 0.64, and 2.56 sq. km, and for the i^{th} model input, we use the following procedure:

1. at each cell c_j of the ΔAAD grid of cell size $|c|$, we compute the total-order sensitivity index $ST_{i,|c|}(c_j)$ of the i^{th} model input with respect to the sum of the ΔAAD indicator over this cell;
2. the set of sensitivity indices $\left\{ST_{i,|c|}(c_j)\right\}$ over all the cells c_j builds a grid (raster map) that we will denote by $\mathscr{S}_{i,|c|}$ and call "*map of sensitivity indices of the i^{th} model input of cell size $|c|$*";

Following this procedure, we obtain $4 \times 4 = 16$ maps of sensitivity indices $\mathscr{S}_{i,|c|}$, one for each of the four model inputs and each cell size $|c|$ = 0.04, 0.16, 0.64, and 2.56 sq. km.

Average values of the maps of sensitivity indices In order to compare the maps of sensitivity indices obtained for various cell sizes $|c|$, we need to summarize these maps by a single scalar measure. Hence, for each cell size $|c|$ = 0.04, 0.16, 0.64, and 2.56 sq. km., and for the i^{th} model input, we calculate the average value of the map of sensitivity indices $\mathscr{S}_{i,|c|}$: we denote this average value by $\overline{ST}_{i,|c|}$. If $G_{|c|}$ denotes the number of cells in the raster map of cell size $|c|$, $\overline{ST}_{i,|c|}$ is defined by:

$$\overline{ST}_{i,|c|} = \frac{1}{G_{|c|}} \sum_{j=1}^{G_{|c|}} ST_{i,|c|}(c_j) \tag{4.33}$$

The average index $\overline{ST}_{i,|c|}$ is a scalar measure that summarises the average contribution of the i^{th} model input to the variance of the ΔAAD indicator aggregated over small cells c_j of area $|c|$. It is by no means a measure of importance of the i^{th} model input with respect to the ΔAAD indicator aggregated over the entire floodplain. We will further explain this point in the discussion.

4.3.3 Results

4.3.3.1 Setting A: ΔAAD indicator on individual supports

Uncertainty analysis Table 4.2 on the facing page summarises the outcome of the uncertainty analysis in setting A: it gives descriptive statistics of the ΔAAD_v indicator for each of the three spatial supports v_1 to v_3 (single house, district, floodplain), over N_{tot} = 28 672 model runs. The mean value and the standard deviation of the ΔAAD_v indicator naturally increase with the surface area of the support v over which it is aggregated (house, district or total floodplain), ranging from 2.060 ±1.21 k€/year for support v_1 to 5 459 ±1 110 k€/year for support v_3. However, if we consider a dimensionless measure of variability such as the coefficient of variation, we observe a different behaviour: the coefficient of variation of the ΔAAD_v indicator decreases with the surface area of the support v.[c] This finding corroborates the idea that some spatial "*averaging-out effects*" result in a reduction of the relative uncertainty when the ΔAAD indicator is aggregated over a large surface area.

[c] We would get an identical result if we looked at the coefficients of variations computed with respect to the aggregated value ΔAAD_v normalized by the surface area $|v|$.

Table 4.2: Descriptive statistics over $N_{\text{tot}} = 28\,672$ simulations; setting A: mean, s.d. and coefficient of variation of the aggregated value of ΔAAD on spatial supports v_1 to v_3; setting B: average values (over non-zero cells) of mean, s.d. and coefficient of variation of ΔAAD at each cell

Setting A

support	area [sq. km]	mean [k€/year]	s.d. [k€/year]	c.var. [%]
v_1 (house)	0.03	2.060	1.210	58.74
v_2 (district)	13	183.7	47.62	25.92
v_3 (floodplain)	63	5 459	1 110	20.33

Setting B

support	cell area [sq. km]	average mean [k€/year]	average s.d. [k€/year]	average c.var. [%]
200 m cells (1 463*)	0.04	3.731	1.380	385
400 m cells (416*)	0.16	13.12	4.105	247
800 m cells (128*)	0.64	42.65	11.72	96
1 600 m cells (43*)	2.56	127.0	32.26	51

∗ number of non-zero cells

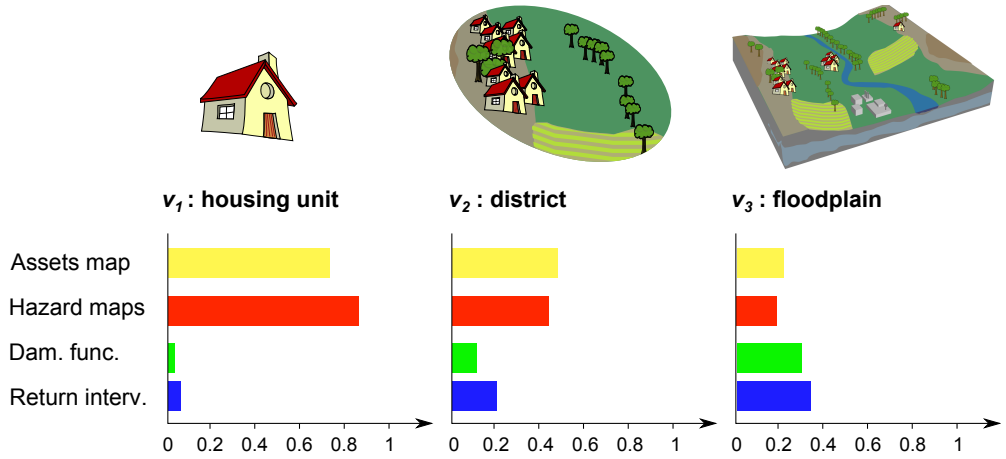

Figure 4.8: Total-order sensitivity indices with respect to the sum of the ΔAAD indicator over three spatial supports: house (left), district (center), and floodplain (right)

Sensitivity analysis Figure 4.8 displays the total-order sensitivity indices computed for each uncertain model input with respect to the aggregated value of the ΔAAD indicator over spatial supports v_1 (house), v_2 (district) and v_3 (floodplain). It clearly suggests that the ranking of uncertainty sources depends on the surface area of the spatial support v. The variance of ΔAAD_{v_1} (smallest support) appears to be mainly explained by the uncertainty on the two spatially distributed inputs, that is, the assets map and the hazard maps (sensitivity indices: 0.8 and 0.65, respectively). On the contrary, the non spatially distributed inputs (depth-damage curves and flood return intervals) prove to be the most important sources of uncertainty when computing the ΔAAD_{v_3} indicator over the total floodplain (sensitivity indices: 0.29 and 0.33, respectively). These results are in line with our theoretical developments of §4.1 and offer clear evidence of the "*change of support*" effect on variance-based sensitivity indices: the sensitivity indices of spatially distributed inputs decrease with the size of the support v while the sensitivity indices of non-spatially distributed inputs symmetrically increase with the size of v.

4.3.3.2 Setting B: raster maps of the ΔAAD indicator

Uncertainty analysis Figure 4.9 on the facing page shows a spatially explicit representation of the uncertainty on the ΔAAD raster map over the $N_{\text{tot}} = 28\,672$ model runs—cell size $|c| = 0.04$ sq. km is taken as an example. A first map displaying the maximum values of ΔAAD for each cell over all model runs (Figure 4.9 a) is compared to the map of minimum values (Figure 4.9 b). It appears that for a large number of cells, the minimum and maximum values of ΔAAD have opposite signs, which we interpret to mean that, due to the input uncertainties in the NOE code, it is impossible to assess with certainty whether these areas will benefit or suffer from the implementation of the flood-control measures on the Orb Delta.

By comparing these maps with that of land use on the study site (Figure 2.19 on page 64), it can be noted that the cells with uncertain sign are mostly covered with agricultural land and show relatively small values of positive or negative ΔAAD. On the contrary, for cells that include urban areas, campgrounds and other economic activities, the ΔAAD indicator proves to keep a constant sign over all model runs, with larger positive or negative values.[a] Hence, in spite of the numerous uncertainties that were considered in the analysis, we can conclude that the flood risk management plan will almost certainly result in a reduction of the average annual damages on urban areas, and almost certainly result in an increase of annual damages on campgrounds. In addition, cells that include urban areas or campgrounds show large standard deviations and low coefficients of variation of the ΔAAD indicator (Figure 4.9 c and d), while cells only covered with agricultural land have small standard deviations but larger coefficients of variation.[a]

Finally, Table 4.2 on the preceding page gives for each cell size $|c| = 0.04$, 0.16, 0.64 and 2.56 sq. km the average value (over the cells c_j) of the mean, standard deviation and coefficient of variation of the ΔAAD indicator.[d] It indicates that the mean and standard deviation of the ΔAAD indicator increase with the surface area $|c|$ of the cells, while its coefficient of variation decreases. These results are consistent with the ones obtained in Setting A (§4.3.3.1 on page 160).

Sensitivity analysis Figure 4.11 on page 165 displays the maps of sensitivity indices $\mathscr{S}_{i,|c|}$ for each model input and for both the smallest cell size $|c| = 0.04$ sq. km and the largest cell size $|c| = 2.56$ sq. km. Spatial distribution of sensitivity indices proves to be heterogeneous. By comparing the first maps of sensitivity indices ($|c| = 0.04$ sq. km) with the map of land use on the study site (Figure 2.19 on page 64), we can identify two different types of areas: urban areas and agricultural land. On the cells that include urban areas, the assets map and the hazard maps display smaller sensitivity indices than on the cells covered with agricultural land. Symmetrically, damage functions and flood return intervals have larger sensitivity indices in urban areas than on agricultural land. This finding might be explained by comparing the characteristics of depth-damage curves for private housing assets and agricultural assets. In particular, depth-damage curves for agricultural land are simple step functions with a number of threshold water levels: when water levels are uncertain, they may induce a "*jump*" from damage amounts below or above these important thresholds. These jumps might explain that the water depth maps have a larger contribution to the variance on the ΔAAD indicator for agricultural land than on urban areas.

Besides, the spatial heterogeneity of the maps of sensitivity indices indicate that we fall out of the hypotheses used in our theoretical developments presented in §4.1. In particular, nor the mean nor the

[a] These qualitative analyses could be improved, for example by computing summary statistics of ΔAAD average value, s.d. and proportion of sign changes over all cells depending on their land use type.

[d] The cells c_j for which the mean value of the ΔAAD indicator over N_{tot} model runs was equal to 0 were not considered to compute these average values.

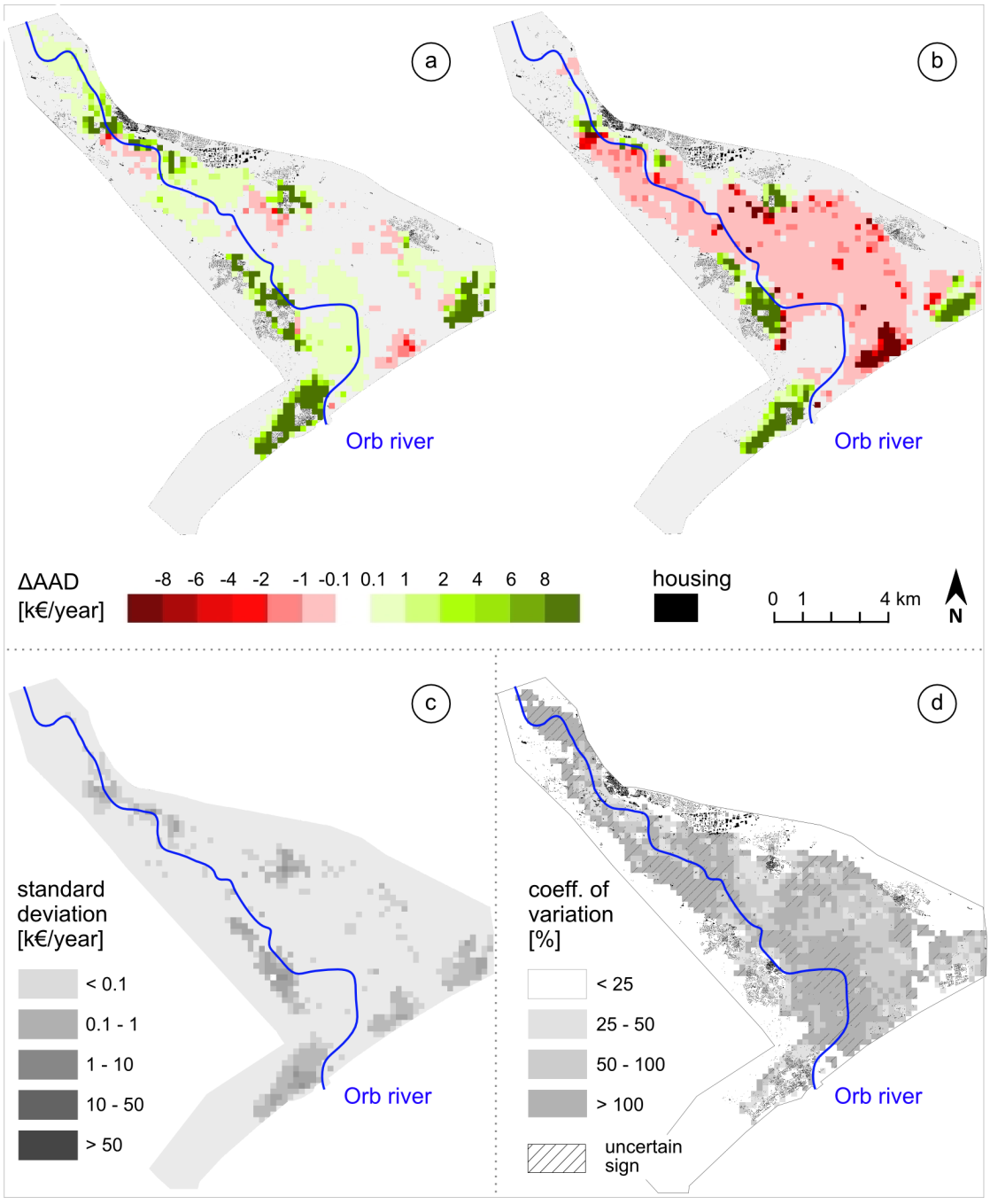

Figure 4.9: Uncertainty on the ΔAAD raster map of cell size $|c| = 0.04$ sq. km over $N_{\text{tot}} = 28\,672$ model runs: maximum values (a), minimum values (b), standard deviations (c), and coefficients of variation (d) at each cell c_j. Dashed cells indicate that the sign of ΔAAD over the cell changes for more than 20% of model runs.

variance of the spatially distributed inputs $Z(\mathbf{x})$ are spatially homogeneous here[e]. This non stationarity results in sensitivity indices being different from one location to one another, whereas they were necessarily constant over the spatial domain in our theoretical developments.

In addition, we can investigate the change of support effect on the NOE modelling framework by comparing the maps of sensitivity indices $\mathscr{S}_{i,|c|}$ for cell sizes $|c|$ = 0.04 sq. km and $|c|$ = 2.56 sq. km (Figure 4.11 on the facing page top and bottom, respectively). The sensitivity indices of the spatially distributed inputs (the assets map and the hazard maps) seem to decrease from $|c|$ = 0.04 to $|c|$ = 2.56 sq. km, while the maps of sensitivity indices of the damage functions and of the flood return intervals display larger values for cell size $|c|$ = 2.56 sq. km than for $|c|$ = 0.04 sq. km. These results are in line with those observed with Setting A (§4.3.3.1 on page 160), that is, the sensitivity indices of spatially distributed inputs decrease with the size of the support v while the sensitivity indices of non-spatially distributed inputs symmetrically increase with the size of v.

To better highlight this change of support effect, Figure 4.10 displays the average values $\overline{ST}_{i,|c|}$ [Eqn. (4.33) on page 160] of the maps of sensitivity indices for each model input and each cell size $|c|$. The sensitivity indices of spatially distributed inputs (assets map and hazard maps) prove to decrease with an increase of the area $|c|$ over which the model output ΔAAD is aggregated, while the sensitivity indices of non spatially distributed inputs (damage functions and flood return intervals) increase symmetrically. The critical cell size $|c|_c$, for which spatially and non-spatially distributed inputs contribute equally to the variance of the model output, falls somewhere between 5 and 50 sq. km.

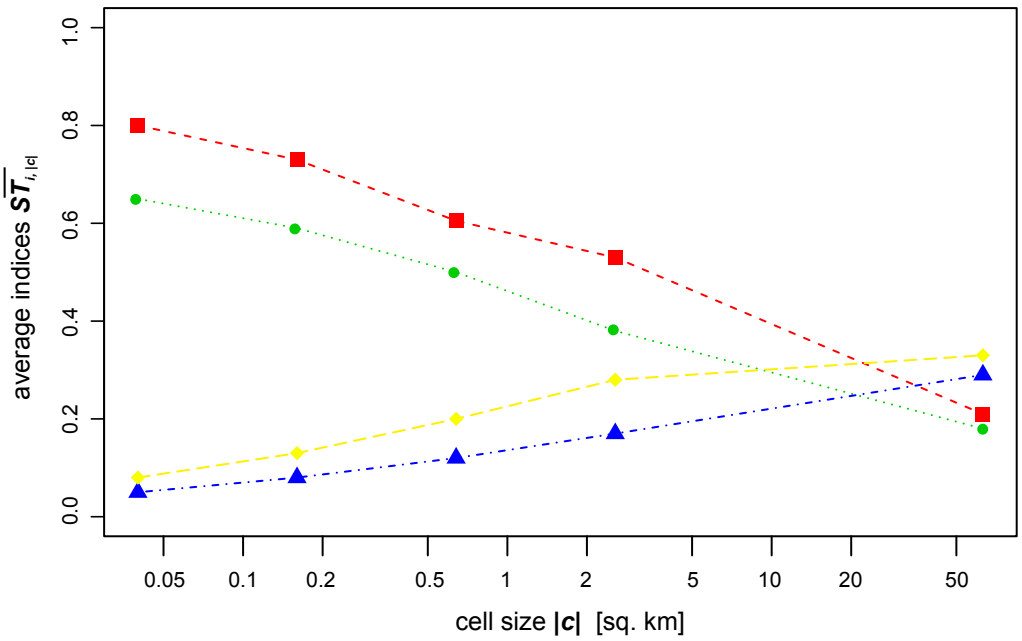

Figure 4.10: Average values $\overline{ST}_{i,|c|}$ of the maps of sensitivity indices with increasing cell size $|c|$ (logarithmic scale) for the asset map (□), the hazard maps (●), the depth-damage curves (△) and the flood return intervals (◇)

[e]Indeed, the mean value of water depths is not constant over the floodplain, and the random realisations of the DEM were generated with conditional simulations, resulting in low variances of simulations close to the ground control points, and larger variances far from the control points.

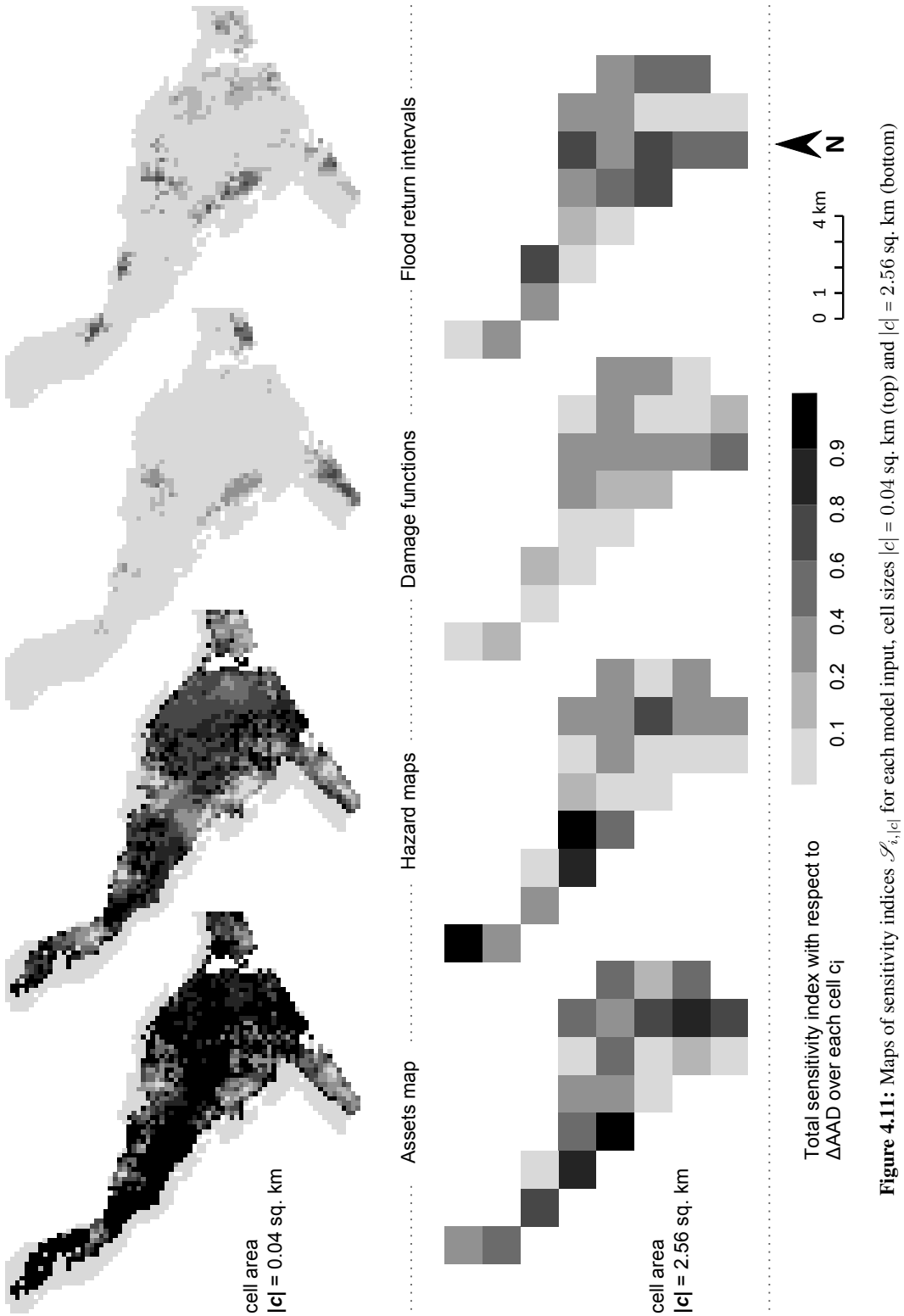

Figure 4.11: Maps of sensitivity indices $\mathscr{S}_{i,|c|}$ for each model input, cell sizes $|c| = 0.04$ sq. km (top) and $|c| = 2.56$ sq. km (bottom)

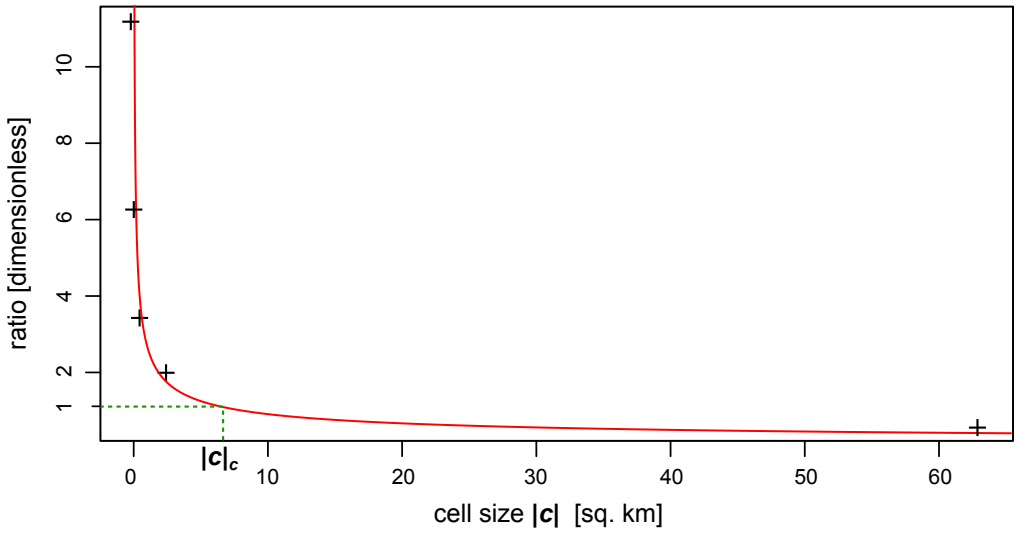

Figure 4.12: Ratio π with fitted curve $\pi = |c|_c/|c|$

To better estimate this critical cell size, we computed for each cell size $|c|$ the ratio π of the average value of sensitivity indices for the group of spatially distributed inputs (assets map and hazard maps) divided by the average value of sensitivity indices for the group of non spatially distributed inputs (damage functions and flood return intervals). Figure 4.12 shows that this ratio is proportional to $1/|c|$: the theoretical curve $\pi = |c|_c/|c|$ derived from Eqn. (4.10) on page 145 was fitted (least squares - $R^2 = 0.79$) on data points, yielding an estimate of the critical cell size $|c|_c \simeq 6.72$ sq. km.

4.3.4 Discussion

Change of support and ranking of uncertainty sources In accordance with the theoretical developments of §4.1, we found that the ranking of uncertainty sources in the NOE modelling framework depend on the spatial support over which the ΔAAD indicator is aggregated. Both settings A and B bring to the same conclusions. For large supports (e.g., the entire floodplain or a district), the main source of uncertainty is the return intervals of flood scenarios. This point is in line with the results obtained in the first sensitivity analysis of the NOE modelling framework and is already discussed in Chapter 3. However, for much smaller supports (e.g., a house, a cell), it appears that the variance of the ΔAAD indicator is mainly due to the uncertainty on the hazard maps (water depths) and the assets map. The critical area $|c|_c$, for which spatially distributed inputs (water depth maps, assets map) and scalar inputs (flood return intervals, damage functions) contribute equally to the variance of the ΔAAD indicator, falls somewhere between 5 and 50 sq. km. Hence, to produce accurate maps of flood damages with an horizontal resolution finer than these critical values, one must try first and foremost to reduce the uncertainty on water depth maps and assets map, which are the key sources of uncertainty on small spatial supports. On the contrary, if the NOE modelling framework is used to produce estimates of total flood damages over a large floodplain, then the uncertain return intervals of flood scenarios contribute more to the variance of the model output than any other source of uncertainty.

Spatial variability of uncertainty of the ΔAAD indicator This analysis also brings a better understanding of the NOE modelling framework by investigating the uncertainty and sensitivity of the maps of

the ΔAAD indicator. First, we observed that the uncertainty on the ΔAAD maps is not spatially homogeneous. In particular, the sign of the ΔAAD indicator is almost certainly constant in some parts of the study area (urban areas: positive ΔAAD; seaside campgrounds: negative ΔAAD), while in other areas (those mostly covered with cultivated land), the sign of the ΔAAD indicator is highly uncertain. This spatially explicit description of uncertainty brings new information for the model end-user, compared to the first analysis (§3.3 on page 111) in which only the aggregated model outputs over the entire floodplain were scrutinized. It may lead the model end-user to better express his concerns about the uncertainty on the ΔAAD maps, and define new quantities of interest to quantify this uncertainty and decide whether it is bearable or not. For example, in this case study, the decision-maker could be especially concerned with the absolute standard deviation of the ΔAAD indicator: he would then pay more attention to urban areas. To decide whether the uncertainty on the ΔAAD map can be tolerated or not in this specific perspective, he could define as a new quantity of interest the maximum ΔAAD standard deviation obtained on an individual cell, or the 90% quantile of ΔAAD standard deviations over all cells. On the contrary, he could be worried not so much about the ΔAAD standard deviation, but rather about the ΔAAD changing sign: in that case he would focus on cultivated land. A possible quantity of interest would then be the proportion of cells on the map for which the ΔAAD indicator changes sign on more than 20% of model runs. To sum up, the maps of ΔAAD uncertainty are certainly valuable tools to better account for the variability of the NOE model outputs, and to discuss what level of uncertainty and what type of uncertainty can be tolerated or not by the model end-user.

Spatial variability of sensitivity indices To identify which sources of uncertainty contribute the most to the variability of the ΔAAD maps, we also produced maps of sensitivity indices computed at each cell of a regular grid. These maps clearly suggest that the contribution of the NOE model inputs to the variance of the ΔAAD indicator is not spatially homogeneous. For example, the sensitivity indices of the water depth maps and the assets map are smaller in urban areas than in areas covered with cultivated land. Such different ranking of uncertainty sources from one location to another may be explained by a number of factors, including the main land use type at that location, the shape of the associated depth-damage curves, the average water depth at that location, etc. Even if we did not explore this point further, the maps of sensitivity indices clearly appear to be promising tools to better explore the behavior of the NOE modelling framework. In particular, an interesting question is how to summarize the information contained in a map of sensitivity indices into a single scalar measure. In this exploratory study, we simply computed the non-weighted average $\overline{ST}_i = 1/G \sum_j ST_i(c_j)$ of sensitivity indices defined with respect to the ΔAAD indicator on each cell c_j of the map. However, we could design other measures, in order to answer the various questions of the model end-user. For example, if the model end-user is mostly concerned with reducing the absolute standard deviation of the ΔAAD indicator, then he may compute the average of cell-based sensitivity indices $ST_i(c_j)$ weighted by the ΔAAD variance on each cell c_j. On the contrary, if he is more worried with the ΔAAD indicator changing signs, he will calculate the average of cell-based sensitivity indices $ST_i(c_j)$ weighted by the proportion of ΔAAD changing signs over all model runs on the cell c_j. These various measures would probably give different conclusions on the key model inputs that drive the uncertainty on the map of the ΔAAD indicator at a given cell size $|c|$. It may be an interesting research item to further explore their properties.

Other averaging-out effects in the NOE modelling framework The theoretical framework we built to explain change of support in VB-GSA can be extended, by analogy, to a number of other averaging-out effects in the NOE modelling framework. For example, we observed in §3.3.5.2 on page 129 an averaging-out effect related to the number of flood-exposed assets of different types on the Orb Delta: the sensitivity indices of the hazard maps was smaller for private housing (large number of assets on the study

site: 16 436) than for other economic activities (small number of assets: 691). A possible explanation is that the dispersion of water depth errors is averaged over the large number of housing polygons scattered across the study area. Hence, we can somehow write that the sensitivity index of the hazard maps with respect to the aggregated value of the ΔAAD indicator over the assets of type "*private housing*" verifies:

$$S_{\text{hazard maps}} \propto \frac{1}{\#T} \tag{4.34}$$

in which $\#T$ is the number of assets of this type. This relation could be applied to other types of assets (agricultural land, campgrounds, etc.). For a given type of assets, if the number of assets grows, then the "*averaging-out effect*" on the water depth errors will get stronger, and the relative contribution of the hazard maps to the variance of the aggregated ΔAAD indicator will be smaller. A similar "*number averaging-out effect*" may partly explain why the uncertainty on depth-damage curves appears to be more influential on the private housing sector, which is described with only one depth-damage curve, than for the other economic activities, which are described by 60 damage curves that are assumed to vary independently.

This analogy may also be useful to understand the contribution of the uncertain return intervals of flood scenarios to the variance of the ΔAAD indicator. As explained in §2.2.7 on page 49, computing the ΔAAD indicator requires to estimate the return intervals (or annual exceedance probabilities) of a number m of flood scenarios. These estimates of flood return intervals are affected by errors that are statistically correlated. By analogy with Eqn. (4.10) on page 145, we may expect that the sensitivity index related to these uncertain flood return intervals (or exceedance probabilities) with respect to the ΔAAD indicator follows a law similar to:

$$S_{\text{return intervals}} \propto \frac{m_c}{m} \tag{4.35}$$

in which m is the number of flood scenarios considered to compute the ΔAAD indicator, and m_c is a measure of the intensity of correlation between the errors on these flood return intervals. In other words, the sensitivity index of uncertain flood return intervals will: i) decrease if the number m increases; and ii) increase if the intensity of correlation between flood return interval estimates increase. Of course, further research is needed to confirm or invalidate these explanations that are only based on analogies. Besides, the implications of these various averaging-out effects will be discussed in the Chapter conclusion on page 172 and in the general conclusion on page 177.

4.4 Chapter conclusion

4.4.1 On scale issues in VB-GSA of spatial models

Sensitivity indices depend on the spatial support, spacing and extent of model output In §4.1, we investigated the case of a point-based and spatially additive model $\mathcal{F}$, in which the spatially distributed input is modelled as a Stationary Random Field $Z(\mathbf{x})$. Using the vocabulary of geostatistics, we suggested to use the term "*site sensitivity indices*" when sensitivity indices of model inputs are computed with respect to the point value $Y(\mathbf{x})$ at some location $\mathbf{x} \in \Omega$, and the term "*block sensitivity indices*" when the variance-based sensitivity indices of model inputs are computed with respect to the aggregated value Y_v of model output over a spatial support v. Then, building on the regularization theory in geostatistics, we proved that the ratio $S_Z(v)/S_U(v)$ of block sensitivity indices of spatially distributed input $Z(\mathbf{x})$ and scalar inputs U_i is proportional to the ratio $|v|_c/|v|$, in which $|v|$ is the surface area of the spatial support over which model output is aggregated, and $|v|_c$ some critical area [Eqn. (4.10) on page 145]. This equation summarizes the effect of spatial support on variance-based sensitivity indices in a point-based and spatially additive model: when the surface area $|v|$ increases, the relative contribution of spatially distributed inputs to the variance of the aggregated model output Y_v decreases. In our contribution, we also specified how the critical size $|v|_c$ depends on the covariance structure $C(\cdot)$ of the spatially distributed input $Z(\mathbf{x})$.

Next, in §4.2, we discussed the influence of spatial "*spacing*" on the VB-GSA of the same point-based and spatially additive model $\mathcal{F}$. We proved that if the aggregated model output Y_v is approximated on a regular grid of points, then the ratio $S_Z(v)/S_\mathbf{U}(v)$ of block sensitivity indices of spatial and scalar inputs will also be approximated. We showed that the approximation error is a $\mathcal{O}(s)$, in which s is the "*spacing*" of the grid of points, that is, the distance between two neighbouring points of the grid.

Finally, we did not consider in our work the last component of the scale triplet (§1.1.3 on page 20), that is, the "*extent*"—the entire area $\Omega \subset \mathbb{R}^2$ covered by the model $\mathcal{F}$. However, for the case of a point-based and spatially additive model in which $Z(\mathbf{x})$ is modelled as a stationary random field, the "*change of support*" properties discussed in §4.1 give an appropriate framework to discuss the impact of "*extent*" as well. If the output of interest for sensitivity analysis is the aggregated value Y_Ω of model output over the entire model extent Ω, then the ratio $S_Z(\Omega)/S_U(\Omega)$ is inversely proportional to the surface area $|\Omega|$ of model extent.

These contributions may prove useful to give a complementary insight on scale issues in spatially distributed modelling. When a spatial model is developed to represent some physical, biological or anthropogenic processes, the choice of a modelling scale (i.e., support, spacing and extent) depends on a number of constraints, some of which are controlled by the modeller, and others not. Among these constraints are the understanding of the processes under study, the intended use of the model results, but also the characteristics of the field data that can be collected, the computationnal power available, etc. What we have shown in this chapter is that the choice of a modelling scale will also partly determine which are the key sources of uncertainty in the model. In particular, if the support of model output is large compared to the characteristic length of correlation in a spatially distributed input, then we can expect that the contribution of this input to the variance of the aggregated model output will be small. Hence, such an input does not deserve too much attention, and extra data gathering or field data collecting should be dedicated to other model inputs.

Limits of our analysis As mentioned in the conclusion of section §4.1, our analysis of scale issues in VB-GSA of spatial models is based on a number of limitative hypotheses that may not be met in some

practical cases. These limits are already discussed in §4.1.4 on page 147, hence we just give here some extra comments on the following key points:

(1) Stationarity of Z(x). In the analytical test case under study in §4.1 and §4.2, we assumed the input random field $Z(\mathbf{x})$ to be stationary; if it is not, site sensitivity indices defined in Eqn. (4.4) on page 143 depend on site $\mathbf{x} \in \Omega$. It is then possible to compute maps of these indices (Marrel et al. 2011; Pettit and Wilson 2010) to discuss the spatial variability of model inputs sensitivities. We used this approach to carry out VB-GSA on the NOE code on the Orb Delta case study in §4.3. We produced map of sensivity indices of different cell sizes (Figure 4.11 on page 165). We think that these maps are valuable tools to investigate the spatial variability of model input sensitivities, identify areas where some inputs are more influential, and finally better understand the behaviour of the model under study.

(2) Random Field model. We also assumed in §4.1 and §4.2 that the spatially distributed input in model $\mathcal{F}$ can be modelled by a Random Field $Z(\mathbf{x})$. The properties that were obtained on the influence of "*support*" and "*spacing*" on variance-based sensitivity indices are only valid for this specific case. However, there are many other theoretical frameworks that could describe the variability of a spatially distributed input (e.g., spatial point processes). Even if we could not prove it in our research, we are of the firm opinion that the properties obtained for Random Field inputs would in most cases hold for other types of spatially distributed inputs. In particular, the influence of the support of model output on variance-based sensitivity indices would most probably be the same: the relative contribution of spatially distributed inputs to the variance of the aggregated output over a support v is a decreasing function of $|v|$.

(3) Case of a non spatially additive model. We can wonder whether the results presented in §4.1 and §4.2 would hold or not for non spatially additive numerical models. A model is non spatially additive when the output of interest is not the spatial average Y_v (or the spatial sum) of spatially distributed ouput $Y(\mathbf{x})$ over a given spatial unit $v \subset \Omega$ (§1.1.5 on page 23). For example, the output of interest could be the maximum value of $Y(\mathbf{x})$ over v (e.g., the maximal pollutant concentration over a study area), some quantile of $Y(\mathbf{x})$ over v, or the percentage of v for which $Y(\mathbf{x})$ exceeds a certain threshold. In theses cases, there is no "*averaging-out effect*" associated with linearity, and there is no reason why the contribution of an uncertain input $Z(\mathbf{x})$ to the variance of the output of interest would increase or decrease under a change of spatial support v. Properties of sensitivity indices discussed in §4.1 and §4.2 no longer hold.

(4) Case of a non-point based model. A model is non-point based when there are some spatial interactions involved in the description of the physical processes under study (§1.1.4 on page 22). For example, models that simulate river flow routing are usually not point-based, as the water flow at a location $\mathbf{x} \in \Omega$ depends on the flow at other locations $\mathbf{x}' \in \Omega$ upstream. Another example is given by Brémond (2011), who developed a model for flood damage assessment in which the damage on a farm located at a given point $\mathbf{x} \in \Omega$ depends on the flood intensity parameters (water depth, velocity, etc.) at this location $\mathbf{x}$ but also on a number of induced damages on crops, warehouses or infrastructures, related to flood intensity parameters at other locations $\mathbf{x}'$. In this case, the change of support properties discussed in §4.1 do not hold either, because the value of the spatial average Y_v does no depend only on the values of the spatially distributed input $Z(\mathbf{x})$ over zone v, but also on the values of $Z(\mathbf{x})$ over total domain Ω. Nevertheless, in many applications, a non-point based model can be replaced by a point-based model as a first-order approximation. Hence, for these "*quasi point-based*" models, the behaviour described for point-based models in §4.1 would also be observed: the block sensitivity indices of spatially distributed inputs over a spatial support v will decrease when v gets bigger.

From change of spatial support to other averaging-out effects We proved in §4.1 that the ratio of block sensitivity indices of spatially distributed model input $Z(\mathbf{x})$ and scalar inputs U_i with respect to the aggregated model ouput Y_v: i) decreases when the size $|v|$ of the spatial support v increases, and ii) increases when the spatial auto-correlation in random field $Z(\mathbf{x})$ increases (larger range parameter a). By analogy, we could describe other "*averaging-out effects*" that occur in VB-GSA of numerical models. For example, we can consider a time-dependent model $\mathcal{F}$, in which one of the model inputs is a time serie $Z(t)$ that exhibits auto-correlation, and model output is a time serie $Y(t)$. If the model is point-based[f] and time additive[g], then the ratio of sensitivity indices of time dependent input $Z(t)$ and scalar inputs U_i will: i) decrease with the length $|t_2 - t_1|$ of the time interval $[t_1, t_2]$ over which the model output $Y(t)$ is aggregated; and ii) increase when the auto-correlation in the time dependent input $Z(t)$ grows. If Δt_c is a characteristic duration that measures the intensity of auto-correlation in the time dependent input $Z(t)$, then a possible portray of "*time averaging out effect*" in VB-GSA might be:

$$\frac{S_Z([t_1, t_2])}{S_U([t_1, t_2])} \propto \frac{\Delta t_c}{|t_2 - t_1|} \tag{4.36}$$

A similar explanation may be used to discuss the case of other non-scalar inputs, such as tabular inputs. We also used the same argument to try and explain some averaging-out effects observed in the NOE modelling framework on the Orb Delta case study (§4.3.4 on page 167). Of course, further research is needed to confirm or invalidate these explanations that are only based on analogies.

4.4.2 On the NOE modelling framework

We carried out in §4.3 a second sensitivity analysis of the NOE modelling framework on the Orb Delta case study, in which we considered as outputs of interest the aggregated value of the ΔAAD indicator over different spatial support v. We briefly summarize here the main outcomes of this analysis, which are discussed into more details in §4.3.4 on page 166.

Ranking sources of uncertainty Following the theoretical developments of §4.1, we found that the ranking of uncertainty sources in the NOE modelling framework on the Orb Delta case study depend on the spatial support over which the ΔAAD indicator is aggregated. Hence, the strategy to reduce the variability of flood damage estimates on the Orb Delta will heavily depend on the choice of a spatial support for model output. If the NOE modelling framework is used to produce maps of flood damages with a resolution finer than 5×5 km, one must try to reduce first the uncertainty on water depth maps and assets map, which are the key sources of uncertainty on small spatial supports. On the contrary, if the NOE modelling framework is used to produce estimates of total flood damages over a large floodplain, then the uncertain return intervals of flood scenarios are the main sources of uncertainty.

Generalizing the results obtained on the Orb Delta case study The "*change of support*" properties that we highlighted in §4.3 for the Orb Delta case study would certainly hold for other floodplains, because they only require the model under study to be point-based and spatially additive. These properties also offer a clear explanation to the empirical observations we made when comparing VB-GSA results on two case studies: the Orb Delta (63 sq. km) and the Vilaine floodplain (10 sq. km), as detailed in §3.3.7 on page 131. It appeared that the contribution of the assets map and water depth maps to the variance

[f] i.e., model output at a given time t only depends on scalar inputs U_i and on the value of the input time serie $Z(t)$ at that same time t

[g] i.e., the output of interest is the average value $Y_{[t_1,t_2]}$ of model output over some time interval $[t_1, t_2]$

of the ΔAAD indicator aggregated over the entire extent of the study site was more important in the Vilaine floodplain (smaller extent) than on the Orb Delta (larger extent). This observation is in line with Eqn. (4.10) on page 145 that describe how block sensitivity indices depend on the spatial support of model output. At the time of writing, it is planned to carry out a more complete comparative analysis of VB-GSA results on another two floodplains along the Rhône River, briefly described in §2.3.3: Fourques-Beaucaire case study (125 sq. km), and ZEC case study (650 sq. km). The outcomes of this comparative analysis will hopefully be used to illustrate and validate change of support effect in VB-GSA of the NOE modelling framework.

Other averaging-out effects in the NOE modelling chain As mentioned in §4.3.4 on page 167, the theoretical framework we built to explain change of support in VB-GSA can be extended, by analogy, to a number of other averaging-out effects in the NOE modelling framework. In particular, we showed that there are such averaging-out effects related to the surface area of flood-exposed assets, to the number of assets of each type (private housing, agricultural land, etc.), to the number of damage functions used, and to the number of flood scenarios considered for the computation of the ΔAAD indicator. These various parameters control the ranking of the uncertainty sources in the NOE modelling framework.

Choice of uncertainty models and averaging-out effects As mentioned in the conclusion of Chapter 3, a key limit to our analysis of the NOE modelling chain is the choice of an uncertainty model for each model input. For some model inputs, this choice is sometimes supported by few or even no data (e.g., damage functions for the Orb Delta case study). What appears now is that it is of the greatest importance to characterise the auto-correlation that uncertain non-scalar model inputs may exhibit. Indeed, we have shown in this chapter that the sensitivity index of a non-scalar model input will depend on its auto-correlated structure: the more (positive) auto-correlation, the largest the contribution of this input to the variance of the model output. However, it is often very difficult to characterise properly the correlation structure in such model inputs. For example, on the Orb Delta case study, we assumed that the uncertainty on each depth-damage curve was independent from the uncertainty on other curves; introducing some sort of correlation in this description would mechanically increase the sensitivity index of depth-damage curves with respect to the total flood damages over the study site. Along the same line, introducing stronger spatial auto-correlation in the description of the DTM would result in an increase of its sensitivity index. Finally, in the same manner, we can discuss the choice that was made to model the variability of the annual exceedance probabilities of flood scenarios. In the Orb Delta study, the errors associated with the return interval of each flood scenario are perfectly related—they all stem from the same **Q-f** curve. Hence, the contribution of the uncertain exceedance probabilities to the variance of the ΔAAD indicator is larger than if the errors on exceedance probabilities were assumed independent for each flood scenario.

Unfortunately, this discussion will be of no help for a modeller who wants to carry out a sensitivity analysis of his model, but has few or no data to support the choice of an uncertainty description for some of the model inputs. However, our contribution may help him anticipate the impact of the arbitrary choices that he will have to do. In particular, the sensitivity indices of a complex model input will increase if it is modelled with a strong auto-correlated structure, and decrease if not. This point will be further discussed in the general conclusion on page 177.

Conclusion and prospects

THE goal of this thesis was to investigate the issue of sensitivity analysis for spatial models, through the detailed study of a model for cost-benefit analysis of flood risk management plans. In this conclusion, we discuss both the methodological and applied contributions of this study. For both aspects, we give a brief summary of our work followed by a number of directions for future research. Finally, we conclude this document by some general comments on the use of sensitivity analysis in environmental modelling.

Variance-based sensitivity analysis for spatial models

A brief summary

The methodological goal of this thesis was to investigate the use of variance-based global sensitivity analysis (VB-GSA) for spatial models. As explained in the introduction, we focused our research on two questions, namely: i) how to handle spatially distributed inputs in VB-GSA? and ii) how to account for scale issues within the VB-GSA framework? These questions are respectively discussed at the end of Chapter 3 and Chapter 4: we just reproduce here our main conclusions.

Handling spatially distributed inputs in VB-GSA. We tried to adress the first research question from a very pragmatic perspective. We displayed an up-to-date review of the numerical tricks that can be used to compute variance-based sensitivity indices for a model input which exhibits spatial auto-correlation. We discussed the pros and cons of these methods and built a decision-tree to choose among them (Figure 3.7 on page 96). We then focused on the "*map labelling*" method. This technique is limited to non CPU intensive numerical models, but it has the desirable feature to allow a complex description of the spatial structure of uncertainty in the model inputs, whose random realisations can be generated with any stochastic process or ad-hoc algorithm. We carried out two numerical studies to investigate how the number of random realisations, and the algorithm used to generate them, may influence the estimation of sensitivity indices. These empirical studies were of an exploratory nature: they raised a number of questions but did not result in any conclusion that could be turned into general rules. Finally, we illustrated the use of the "*map labelling*" method on the NOE code for the Orb Delta case study, in which two spatially distributed model inputs were considered: the assets map and the hazard maps.

Scale issues in VB-GSA. We adressed the second research question with a more theoretical perspective, using the concepts, notations and results from the geostatistics theory. Our contribution lies in a better understanding of the link between the scale of a model output and the variance-based sensitivity indices. Theoretical results are limited to the specific class of point-based and spatially additive models, in which

the spatially distributed inputs are modelled as Stationary Random Fields. We used the notion of the "*scale triplet*" to decompose the research question into three items, namely the influence of support, spacing and extent of model output. We suggested to define "*site sensitivity indices*" and "*block sensitivity indices*" to account for the role of the spatial support of model output, and established some properties of these indices. In particular, we proved that the relative contribution of an uncertain spatially distributed input to the variance of the model output: i) decreases with the size of the spatial support considered for model output aggregation [Eqn. (4.10) on page 145]; and ii) increases with its correlation length, and more generally depends on its auto-correlated structure. To be more precise, we gave a formula that expresses the ratio of block sensitivity indices of a spatially distributed input $Z(\mathbf{x})$ and scalar inputs U_i as a function of i) the surface area considered for model output aggregation, and ii) the covariance function of $Z(\mathbf{x})$. The role of the other two components of scale (spacing and extent) were less scrutinized.

Prospects and directions for future work

The methods and results described in this thesis may be applied straightforwardly to a wide range of non CPU intensive models, in different fields of earth and environmental sciences. The main difficulty the modeller will encounter to apply these methods is to characterise the uncertainty in the model inputs: this issue was not the focus of our work, even if we suggested a few strategies to model spatial uncertainty (§3.2.2 on page 98). Besides, we would like to stress three points that could extend the reach of our study.

Other types of complex model inputs. The methods we described to handle spatially distributed inputs in VB-GSA could easily be used to deal with other types of high-dimensional model inputs, such as functional inputs, time series, tabular data, etc. For example, in the sensitivity analysis of the NOE code on the Orb Delta case study, uncertain depth-damage relations were represented as tabular data: they were handled with the "*map labelling*" method. Of course, all the issues discussed for spatially distributed inputs would have to be carefully investigated for these other types of inputs. In particular, attention should be paid to carefully model the auto-correlation in such non scalar model inputs.

Other averaging-out effects. By analogy with the "*change of support*" effect that we detailed in §4.1 on page 140, it is possible to explain other "*averaging-out effects*" that occur in VB-GSA of numerical models. This point is extensively explained in the conclusion of Chapter 4 on page 171. We will also discuss it for the specific case of the NOE modelling framework later in this conlusion.

Scaling issues in spatial modelling. Our contributions on scale issues in VB-GSA may prove useful to give a complementary insight on the problem of scaling in spatially distributed modelling. When a model is developed to represent some physical, biological or anthropogenic processes, the choice of a modelling spatial scale (i.e., support, spacing and extent) depends on a number of constraints, some of which are controlled by the modeller, and others not. Among these constraints are the understanding of the processes under study, the intended use of the model results, but also the characteristics of the field data that can be collected, the computational power available, etc. What we have shown in Chapter 4 is that the choice of a modelling scale will also partly determine which are the main sources of uncertainty in the model.

Finally, to further expand the findings of this thesis, it would be necessary to explore new research topics. Here are four research items that could be investigated.

(1) Assessing the influence of uncertain map attributes. As mentioned in §3.1.4 on page 92, none of the techniques encountered in the literature can answer the following question: if a spatially dis-

tributed input proves to be influential, which of its "*attributes*"[h] contribute the most to the variance of the model output? A possible way to answer this question would be to first carry out VB-GSA with the "*map labelling*" technique, then to perform a complementary graphical analysis based on scatterplots or Contribution to the Sample Mean plots to assess the relative influence of the different map attributes.

(2) Optimal sampling of spatially distributed inputs. Our exploratory study of section §3.2 on page 98 clearly suggets that further research is needed to find optimal sampling techniques of spatially distributed inputs in VB-GSA. An interesting path of research is opened by the work of Scheidt and Caers (2009) who suggest to draw optimal samples of random realisations of $Z(\mathbf{x})$ based on the definition of a distance between realisations.

(3) VB-GSA for CPU intensive spatial models. We noted that none of the available techniques described in Chapter 3 is appropriate to handle a time-consuming model with several spatially distributed model inputs $Z_i(\mathbf{x})$ whose dimension cannot be reduced. Hence, further research is needed to solve this situation, in which intensive simulation is computationally intractable.

(4) Relaxing hypotheses on scaling results. The results we obtained on scale issues in VB-GSA are conditioned to a number of limitative hypotheses, namely: i) the spatially model input is modelled as a Stationary Random Field; ii) the model under study is spatially additive; and iii) the model under study is point-based. An interesting research path would be to try and relax some of these conditions. This point is extensively discussed in the conclusion of Chapter 4 on page 169.

Apart from these four points, we suggest to add a last direction for future work. We have not fully adressed in this manuscript the issue of computing variance-based sensitivity indices when the model output Y is spatially distributed. We have focused on two specific situations, in which the output of interest for the modeller is either the output value $Y(\mathbf{x}^*)$ at some specific location $\mathbf{x}^*$, or the aggregated value Y_v over a given support v. A third situation is the following: the modeller may be interested in the spatially distributed model output $\{Y(\mathbf{x}) : \mathbf{x} \in \Omega\}$ as a whole, and may want to define sensitivity indices with respect to the overall variability of the output map. Here are two ideas to adress this issue.

(5) Generalized sensitivity indices. A first idea is to build up on the work of Campbell et al. (2006), who investigated the issue of sensitivity analysis with a functional or multivariate output. They suggest to: i) use any dimension reduction technique—such as Principal Component Analysis— to extract a small number of scalar components $Y^{(k)}$ from the multivariate output Y, then ii) estimate sensitivity indices $S_i^{(k)}$ with respect to each of these scalar components. Lamboni et al. (2011) apply this approach for a time-dependent output $Y(t)$, and further define a new "*generalized sensitivity index*" $S_i^{(gal)} = \sum_k p_k S_i^{(k)}$ as a weighted average of indices $S_i^{(k)}$, in which weights p_k represent the energy content of each independent scalar component $Y^{(k)}$—see also Auder et al. (2012). With a reasonable effort, it should be possible to adapt this method to a spatially distributed output $Y(\mathbf{x})$: the main difficulty would be to find the most appropriate dimension reduction technique for spatially distributed data.

(6) A new measure of importance? Another research path would to be build on the work of Liu and Homma (2010) to define a new importance measure dedicated to spatially distributed outputs $Y(\mathbf{x})$. Let assume that we have defined a measure of dissimilarity δ that could represent the overall "*dissimilarity*" of a set of maps—Scheidt and Caers (2009) give an example of such a measure. A possible way to assess the contribution of the i^{th} model input to the variability of the model output would be to compute the difference between: i) the dissimilarity δ of a set of output maps obtained

[h] e.g., any scalar descriptor such as the mean value of the input map, the value at a particular location, etc.

by running simulations of the model under study with a full sampling is the space of the uncertain model inputs, and ii) the reduced dissimilarity $\delta^{(i)}$ of a set of output maps obtained from a conditional sampling of model inputs with a fixed value of the i^{th} model input. Of course, such an importance measure would probably lack most of the nice properties of variance-based sensitivity indices (sum up to 1, etc.).

Lessons on cost-benefit analyses of flood risk management plans

The applied goal of our thesis was to investigate the propagation of uncertainty in cost-benefit analyses of flood risk management plans, and to identify the most influential sources of uncertainty with VB-GSA. We display a brief summary of this part of our research, and give directions for future work.

A brief summary

A preliminary but essential step of our work was to clarify the perimeter and characteristics of the model we had to study. Hence, a first applied contribution of this thesis was to designed a modelling framework named NOE to describe cost-benefit analyses of flood risk management plans. Particular attention was paid in this framework to two points that are usually not emphasized in the literature: i) the proper definition of the average annual damages (AAD indicator); and ii) the description of the spatial overlay procedure between the hazard maps and the assets map.

Next, we implemented the NOE modelling framework on five study sites, and carried out sensitivity analysis for three of them: the Orb Delta, the Vilaine floodplain and the ZEC project along the Rhône river. Only the results obtained on the Orb Delta case study were extensively presented and discussed in this manuscript (§3.3 on page 111 and §4.3 on page 156). However, in this conclusion, we offer to draw general conclusions from the various studies that were carried out. A key result is that **it is impossible to establish a fixed and general ranking of the sources of uncertainty in CBA studies** applied to flood risk management plans. On the contrary, we proved that the contribution of each source of uncertainty to the variance of the NOE model outputs depends on a number of factors that may change from one case study to another. The second applied contribution of this thesis was to investigate three of these factors, which are listed below.

Economic sector. We found that the ranking of uncertainty sources with respect to their contribution to the variance of the flood damage estimates depends on the economic sector considered (private housing, agriculture, industry, etc.).

Spatial extent of the study area. All other things being equal, the relative contributions of the assets map and water depth maps to the variance of the flood damage estimates are a decreasing function of the extent of the study area. Symmetrically, the relative contribution of non spatially distributed inputs (damage functions, flood return intervals) will increase with the extent of the study area. These empirical observations are in line with the theoretical results we obtained on scale issues in VB-GSA of point-based and spatially additive models.

Other non-spatial averaging-out effects. As mentioned in §4.3.4 on page 166, the theoretical framework we built to explain change of support in VB-GSA can be extended, by analogy, to a number of other averaging-out effects that occur in the NOE modelling framework. In particular, we showed that there are such averaging-out effects related to the surface area of flood-exposed assets, to the number of assets

of each type (private housing, agricultural land, etc.), to the number of damage functions used, and to the number of flood scenarios considered for the computation of the ΔAAD indicator. These various parameters control the ranking of the uncertainty sources in the NOE modelling framework.

Prospects and directions for future work

One initial motivation for the applied part of this thesis was to better address uncertainty in CBA studies applied to flood risk management plans. It is obviously a challenging task to make firm and detailed recommendations on the basis of our results, which show that the ranking of uncertainty sources in flood damage assessments differ from one case study to another. However, the following comments might prove useful to modellers.

What resolution for flood damage maps? Experts that produce flood risk assessment studies often have to choose a spatial resolution for the production of flood damage maps. This thesis does not give a firm answer to this difficult question, but yields a better understanding of the following point: the choice of a given resolution (i.e., a spatial support for the aggregation of the flood damage indicator) will determine which sources of uncertainty are the most influential and which are not. If flood damage maps are produced at a fine resolution, one must try to reduce first the uncertainty on water depth maps and assets map, which will probably be the key sources of uncertainty on small spatial supports. On the contrary, if average annual flood damages are aggregated over a large floodplain, then the uncertain return intervals of flood scenarios are likely to be the main sources of uncertainty.

Flood risk analyses—how detailed do we need to be? This is the title of a research paper by Apel et al. (2009), who discussed the question of the required model complexity in flood risk analyses. As they explain it, "*the methods used in research and practical applications range from very basic approaches with numerous simplifying assumptions up to very sophisticated, data and calculation time demanding applications both on the hazard and on the vulnerability part of the risk.*" One contribution of this thesis is to shade some new light on this tough issue. The level of "*detail*" or "*complexity*" of a model for flood risk assessment can somehow be described as a level of aggregation (not necessarily spatial) as defined in §1.1.3 on page 20[i]. Disaggregated models often give the—false—impression that they produce more "*precise*" results than aggregated models. Yet, only a careful uncertainty and sensitivity analysis can assess the variability of the model output and estimate the contribution of each model input to this variability. In the light of our work, we want to underline the following point: **the more disaggregated a model is, the more difficult it will be to carry out a sensitivity analysis of it**. Indeed, we have shown that it is of the greatest importance in sensitivity analysis to properly characterise the auto-correlation that disaggregated model inputs may exhibit, because the strength of this auto-correlation will partly determine their sensitivity indices.[j] Unfortunately, it is most of the time difficult or even impossible to collect enough data to support the choice of an auto-correlated uncertainty model for such disaggregated inputs.[k] What is dangerous for the modeller is the tantalising solution to just ignore auto-correlation and consider the elements of a disaggregated input as either statistically independent or completely colinear.

[i] For example, in the flood damage assessment that was carried out on the Orb Delta study site, only one depth-damage curve was used to model damage for private housing. On the contrary, the CEPRI guidelines for cost-benefit analysis of flood risk management plans suggest to use up to seven different depth-damage relations depending on the caracteristics of the houses (CEPRI 2011). We can say that the CEPRI model is more "*detailed*"—or "*complex*"—than the Orb one. We can also say that it operates at a "*less aggregated*" level. There are many other examples of flood damage assessments that are more or less aggregated than these two.

[j] This issue has been further discussed in the conclusion of Chapter 4 (§4.4 on page 169)

[k] For example, in the CEPRI guideline—like in any other case where depth-damage relations are designed at a low level of aggregation—the modeller would have to quantify the correlation between the seven uncertain depth-damage curves for private housing.

One contribution of this thesis is to anticipate the impact of such arbitrary choices: in particular, the sensitivity indices of a disaggregated model input will likely be overestimated if its components are modelled as completely colinear, and underestimated if its components are modelled as purely independent. To sum up, we are of the opinion that disaggregated models make uncertainty and sensitivity analyses more difficult to carry out, and are thus more difficult to evaluate with regards to uncertainty management.

Directions for future work. Further research is needed to extend our study on uncertainties in cost-benefit analyses applied to flood risk management plans. Here is a list of important research topics that could be investigated.

(1) Hydraulic modelling. One major limit of our study is that we considered the hazard maps as model inputs: we did not include the hydraulic model that produces them as a part of the NOE modelling framework. To overcome this limit, further work could build on the extensive literature that already deals with uncertainties in hydraulic modelling (de Rocquigny et al. 2010; Arnaud et al. 2006). A notable challenging issue is that hydraulic models are not point-based models—the water depth, water velocity or flood duration at one location depend on the water flow upstream.

(2) Ignored sources of uncertainty. Some uncertainty sources were identified but not taken into account in the sensitivity analysis of the NOE modelling chain. In further work, at least two of them–which are uncertain modelling assumptions rather than uncertain model inputs—could be included with a reasonable effort: i) the choice of a technical procedure for spatial overlay analysis between hazard maps and assets maps; and ii) the initial choice of flood scenarios considered (number of scenarios and their characteristics).

(3) Mischaracterised sources of uncertainty. Another difficulty in sensitivity analysis is to properly describe and quantify uncertainty on the various model inputs. Regarding the NOE modelling framework, there is room for improvement in the modelling of the input uncertainties. We can for example mention the case of uncertainties on the annual exceedance probabilities of flood scenarios: one aspect that we did not adress is the estimation of flood exceedance probabilities and associated uncertainty when many river tributaries and lateral inflows contribute to the main river under study.

(4) Grouping of model inputs. The numerous inputs of the NOE code were gathered into five groups to make the sensitivity analysis computationally more tractable. The outcomes of the analysis heavily depend on the initial choices that were made to compose these groups, and further work should try to divide these groups into subgroups to better identify the most influential inputs. This point has been discussed in the conclusion of Chapter 3 on page 135.

Besides, the NOE modelling framework does not account for all possible subtleties in CBA-AD studies applied to flood risk management plans. To extend the reach of our discussion on the uncertainties in these CBA-AD studies, it would be necessary to investigate the following extra research topics.

(5) Continuous scenarios. As mentioned in §2.2.6 on page 48, flood scenarios that are used in the computation of the annual average damages could be randomly generated over a very large length of time in order to build a plausible chronicle of flooding events over time. An open research item is how to account for uncertainty in such chronicles and related computation of the AAD indicator.

(6) Non point-based damage assessment. As discussed earlier (limits of the analysis in the conclusion of Chapter 4 on page 169), models for flood damage assessment are not necessarily point-based. For example, the damage on a farm located at a given point may depend on the flood intensity

parameters at this location but also on a number of induced damages on crops, warehouses or infrastructures, related to flood intensity parameters at other locations (Brémond 2011). Another example is that of flood damage assessment for roads or energy supply networks, in which the damage on one part of the network heavily depends on the flood impacts on other parts of the network (Gleyze 2005). An open research item is how to account for uncertainty in such models, and how scale issues that were discussed in this thesis would hold or not in that non point-based case.

(7) Evolution of land use over time. In the NOE modelling framework, the state of the whole system under study (land use, hydrologic and hydraulic characteristics, etc.) is assumed to be fixed through the length of time over which the flood risk management plan is evaluated (typically 30 to 50 years). Relaxing this hypothesis would require a large amount of work and would open a number of new research questions. In particular, if we acknowledge that the climate and subsequent hydrologic characteristics of a floodplain vary over time, then the probabilistic framework in which we defined the average annual flood damages is not appropriate anymore.

Practice of sensitivity analysis in environmental modelling

We would like to conclude this thesis by a few practical comments on the outcomes of sensitivity analysis in environmental modelling. During this research work, we partly acted as modellers who had to build and implement a numerical model, carry out a sensitivity analysis, and communicate the results of this sensitivity analysis to the model end-users. These comments are modest testimonies from our experience.

Outcomes of sensitivity analysis: a modeller's point of view

The main reason given in the literature to justify the use of sensitivity analysis is to reduce the variability of the model output by identifying the key sources of uncertainty[1]. However, in our experience, this goal is often difficult to reach because reducing the variability of the key model inputs may be impossible. Nevertheless, sensitivity analysis brings some other invaluable outcomes that we detail below.

First, from a practical perspective, the most challenging step of an uncertainty/sensitivity analysis is to identify and describe the various sources of uncertainty involved in a model—uncertain inputs, modelling assumptions, etc. In our view, this first step is also the most instructive for the modeller. Indeed, by carefully discussing the nature of uncertainty in his model, the modeller will be led to foresee problems that he ignored so far and may come up with new ideas. For example, in our case, investigating the nature of uncertainty in cost-benefit analyses of flood risk management plans was a strong incentive to better define two parts of the NOE modelling framework: i) the multi-dimensional definition of the Average Annual Damages (§2.2.5 on page 44); and ii) the spatial overlay procedure to assess flood exposure from the hazard maps and the assets map.

Next, sensitivity analysis is also a wonderful excuse for striking up conversations with the various experts involved in a complex modelling project. A common situation in environmental modelling is the following: a large modelling chain is built to support decision-making on a given issue (e.g., flood damage assessment); this chain is composed of a number of submodels that summarize various fields of knowledge (e.g., hydrology, hydraulic, economy, etc.). These various submodels or input data are often produced by different experts, and sensitivity analysis of the global modelling chain requires to collect

[1] This rationale is called "*variance cutting*" in Saltelli et al. (2008)

relevant information on uncertainty for each of them. This is a daunting task, but it offers the opportunity to better understand the behaviour of each submodel and to promote a shared view of uncertainty treatment with all the different partners involved in a modelling project.

Finally, sensitivity analysis has also proven its worth as an aid in better understanding the limits of a model. In particular, the exploration of the space of uncertain model inputs may help identify some particular range of input values in which the model has an unexpected behaviour. Besides, sensitivity analysis —or more precisely uncertainty analysis, which can be seen as the first step of SA—also produces empirical confidence intervals on the outputs of the model. These are essential outcomes that should help the modeller to decide what use can be done of the model and to what extent he can draw firm conclusions and recommendations on the basis of the model outputs.

Communicating sensitivity analysis results to model end-users

During our research, we organized a number of meetings to communicate the results of sensitivity analysis of the NOE modelling framework to the water managers involded on the various study sites (not only the Orb Delta but also the Rhône River). We list below three major difficulties we encountered: they are likely to be faced by anyone trying to communicate to the model end-users the results of a sensitivity analysis of a spatial model.

A first recommendation is to clearly present and discuss the results of the uncertainty analysis *before* switching to sensitivity analysis conclusions. Our experience shows that it is useless to discuss sensitivity analysis results if uncertainty information has not been correctly interpreted by the decision makers. Indeed, only the model end-user is legitimate to decide whether the variability of a model output is small enough or needs to be reduced. Besides, it is also his role—with the help of the modeller—to define what measure of variability (variance of the model output, probability to exceed a threshold value, etc.) should be chosen to perform sensitivity analysis.

A second recommendation is to insist on the limitations in the interpretation of the sensitivity analysis results. Indeed, we observed a trend toward an over-interpretation of sensitivity indices by the model end-users. A small value of S_i is often interpreted by decision makers as "*model input U_i is not important and we can completely forget about it, in any future application of this model*", when a cautious interpretation would rather be: "*in this specific model, with the specific uncertainty ranges that were specified for the different input variables, the uncertainty on input U_i does not contribute much to the overall variability of model output*". In particular, the results of a sensitivity analysis do not say what would happen if the uncertainty ranges on model inputs were severely modified. Hence, as a general rule, it is necessary to carry out a new sensitivity analysis, with appropriate uncertainty ranges, if a model is to be applied to a new data set. Communicating the results of sensitivity analysis to model end-users thus requires to insist on these limitations.

Finally, a last challenging issue is how to communicate to model end-users the outcomes of an uncertainty and sensitivity analysis, especially when the model outputs are maps. An unmissable starting point on this vast subject is the RIVM/MNP Guidance on Uncertainty Assessment and Communication (Van der Sluijs et al. 2003), which includes a volume dedicated to the visualisation of spatial uncertainty (Visser et al. 2006). Many other authors have also brought valuable contributions on this issue: Thomson et al. (2005) displayed an extensive review on the question; Kunz et al. (2011) focused on uncertainty visualisation methods in natural hazards assessments and suggested various visual variables to map uncertainties; finally, Viard et al. (2011) showed that spatial uncertainty visualisation does influence decision making, and investigated the difference between adjacent versus coincident display of spatial uncertainties. These

papers essentially lead us to the conclusion that a clear communication of uncertainties associated with a spatial decision support system is both a necessity and a difficult challenge.

References

Achleitner, S., Huttenlau, M., Gems, B., Stötter, J., and Aufleger, M. (2010). Sensing the uncertainty of cumulative flood losses due to varying flood peak distribution functions in a Monte-Carlo based loss modelling framework. In *IAHR 2010 European Congress*. 39, 48, 49

Aerts, J. C. J. H., Heuvelink, G. B. M., and Goodchild, M. F. (2003). Accounting for spatial uncertainty in optimization with spatial decision support systems. *Transactions in GIS*, 7(2):211–230. 99

Allamano, P., Laio, F., and Claps, P. (2011). Effects of disregarding seasonality on the distribution of hydrological extremes. *Hydrol. Earth. Syst. Sci.*, 15:3207–3215. 51

Allard, D. (2010). Introduction to non-gaussian random fields: a journey beyond gaussianity. 232

Almansa, C. and Martínez-Paz, J. (2011). What weight should be assigned to future environmental impacts? A probabilistic cost benefit analysis using recent advances on discounting. *Sci. Total Environ.*, 409(7):1305–1314. 123

Apel, H., Aronica, G. T., Kreibich, H., and Thieken, A. H. (2009). Flood risk analyses: how detailed do we need to be? *Nat. Hazards*, 49(1):79–98. 177

Apel, H., Merz, B., and Thieken, A. (2008). Quantification of uncertainties in flood risk assessments. *Int. J. River. Basin. Manage.*, 6(2):149–162. 3, 11, 112, 129

Apel, H., Thieken, A. H., Merz, B., and Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Nat. Hazards Earth Syst. Sci.*, 4(2):295–308. 3, 11, 112, 129

Archer, G., Saltelli, A., and Sobol, I. (1997). Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *J. Stat. Comput. Simul.*, 58:99–120. 30, 89, 219

Armatte, M. and Dalmedico, A. D. (2004). Modèles et modélisations, 1950-2000 : Nouvelles pratiques, nouveaux enjeux. *Revue d'histoire des sciences*, 57(2):243–303. 1, 9

Arnaud, A., Goutal, N., and Moncenix, L. (2006). Prise en compte des incertitudes dans les modèles hydrauliques. Rapport H-T56-2006-03618-FR, EDF R&D. 178

Arnell, N. W. (1989). Expected annual damages and uncertainties in flood frequency estimation. *J. Water Resour. Plann. Manage.*, 115(1):94–107. 47, 112, 116

Ascough, J., Maier, H., Ravalico, J., and Strudley, M. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecol. Model.*, 219:383–399. 3, 11, 112

Auder, B., De Crecy, A., Iooss, B., and Marquès, M. (2012). Screening and metamodeling of computer experiments with functional outputs. Application to thermal-hydraulic computations. *Reliability Engineering and System Safety*, 107:122–131. 175

Baecher, G., Foufoula-Georgiou, E., Keeney, R., and Lave (2000). *Risk analysis and uncertainty in flood damage reduction studies*. National Academy Press, Washington, D.C. 39

Bales, J. and Wagner, C. (2009). Sources of uncertainty in flood inundation maps. *J. Flood Risk Manage.*, 2:139–147. 3, 11, 112, 119

BCEOM (2000). Étude de gestion du risque inondation dans le bassin versant de l'Orb. Rapport technique, Syndicat Mixte de la Vallée de l'Orb. 59, 61, 113, 115

Bernardara, P., de Rocquigny, E., Goutal, N., Arnaud, A., and Passoni, G. (2010). Uncertainty analysis in flood hazard assessment: hydrological and hydraulic calibration. *Can. J. Civ. Eng.*, 37(7):968–979. 112

Blöschl, G. and Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrol. Processes*, 9:251–290. 5, 13, 20, 21, 22, 138

Bolado-Lavin, R., Castaings, W., and Tarantola, S. (2009). Contribution to the sample mean plot for graphical and numerical sensitivity analysis. *Reliab. Eng. Syst. Saf.*, 94(6):1041 – 1049. 28, 94, 220

Bonin, O. (2006). Sensitivity analysis and uncertainty analysis for vector geographical applications. In *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisbon, Portugal, July 5-7. 86, 99, 100, 121

Bordin, P. (2002). *SIG: concepts, outils et données*. Hermès Science Publications. 19

Borgonovo, E., Castaings, W., and Tarantola, S. (2011). Moment independent importance measures: new results and analytical test cases. *Risk Anal.*, 31(3):404–428. 28

Borgonovo, E., Castaings, W., and Tarantola, S. (2012). Model emulation and moment-independent sensitivity analysis: an application to environmental modeling. *Environ. Modell. Softw.*, 34:105–115. 28

Bouleau, N., Chatzis, K., and Coutard, O. (2004). Les modèles, obstacles et enjeux de la critique. Rapport CDU RST RCPVS04-002, Ministère de l'Équipement, DRAST, Centre de prospective et de veille scientifique et technologique. 1, 9

Bournot, A. (2008). *Evaluation de la pertinence des mesures de gestion du risque d'inondation. Manuel des pratiques existantes*. Centre Européen de Prévention du Risque d'Inondation. 39, 42, 57

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth, Belmont CA. 28

Briant, M. (2001). HEC-FDA sensitivity and uncertainty analysis. Research Document 46, USACE Hydrologic Engineering Center. 112

Brémond, P. (2011). *Caractérisation et évaluation économique de la vulnérabilité des exploitations agricoles aux inondations*. PhD thesis, Université Montpellier I. 73, 170, 179

Brown, J. D. and Heuvelink, G. B. M. (2007). The data uncertainty engine (DUE): a software tool for assessing and simulating uncertain environmental variables. *Comput. Geosci.*, 33(2):172–190. 140

Brown, J. D., Heuvelink, G. B. M., and Refsgaard, J. C. (2005). An integrated methodology for recording uncertainties about environmental data. *Water Sci. Technol.*, 52(6):153–160. 100

Busby, D., Romary, T., Feraille, M., and Touzani, S. (2010). An integrated approach for uncertainty and sensitivity analysis in reservoir forecasting. *J. Comput. Geosci.* Submitted. 83

Campbell, K., Mckay, M. D., and Williams, B. J. (2006). Sensitivity analysis when model outputs are functions. *Reliab. Eng. Syst. Saf.*, 91(10-11):1468–1472. 88, 175

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environ. Modell. Softw.*, 22(10):1509–1518. 28

Campolongo, F., Saltelli, A., and Cariboni, J. (2011). From screening to quantitative sensitivity analysis. a unified approach. *Comput. Phys. Commun.*, 182(4):978–988. 32

Cariboni, J., Gatelli, D., Liska, R., and Saltelli, A. (2007). The role of sensitivity analysis in ecological modelling. *Ecol. Model.*, 203(1-2):167–182. 2, 10, 25, 112, 140

Carlos García-Díaz, J. and Gozalvez-Zafrilla, J. (2012). Uncertainty and sensitive analysis of environmental model for risk assessments: An industrial case study. *Reliab. Eng. Syst. Saf.*, 107:16–22. 28

CEPRI (2011). L'ACB (analyse coût/bénéfice) : une aide à la décision au service de la gestion des inondations. Guide à l'usage des maîtres d'ouvrage et de leurs partenaires, Centre Européen de Prévention des Risques d'Inondation. 39, 42, 177

Challinor, A. J., Wheeler, T. R., Craufurd, P. Q., Slingo, J. M., and Grimes, D. I. F. (2004). Design and optimisation of a large-area process-based model for annual crops. *Agric. For. Meteorol.*, 124(1-2):99–120. 18

Chan, K., Saltelli, A., and Tarantola, S. (2000). Winding stairs: a sampling tool to compute sensitivity indices. *Stat. Comput.*, 10:187 – 196. 28, 32

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*, chapter 7 - Conditional Simulations, pages 449–507. Wiley, New York. 22, 99, 103, 105, 141, 145, 146, 152, 232

Countryman, J. and Tustison, B. (2008). Flood frequency confidence bounds: art, science or guess. In *Proceedings of the World Environmental and Water Resources Congress 2008: Ahupua'a*, volume 316. 119

CREM (2009). Guidance on the development, evaluation, and application of environmental models. Technical report, US Environmental Protection Agency, Council for Regulatory Environmental Modeling. 2, 10, 25, 112, 140

Cressie, N. (1993). *Statistics for spatial data, revised edn*. Wiley, New York. 142

Crosetto, M. and Tarantola, S. (2001). Uncertainty and sensitivity analysis: tools for GIS-based model implementation. *Int. J. Geogr. Inf. Sci.*, 15:415–437. 83, 84

Cukier, R., Levine, H., and Schuler, K. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *J. Comput. Phys.*, 26:1–42. 28, 32

Da-Veiga, S. (2007). *Analyse d'incertitudes et de sensibilité : application aux modèles de cinétique chimique*. PhD thesis, Université Toulouse III. 29

de Kort, I. A. T. and Booij, M. J. (2007). Decision making under uncertainty in a decision support system for the Red River. *Environ. Modell. Softw.*, 22(2):128–136. 112

de Moel, H. and Aerts, J. C. J. H. (2011). Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. *Nat. Hazards Earth Syst. Sci.*, 58:407–425. 3, 11, 112, 119, 121, 122, 129

de Moel, H., Asselman, N., and Aerts, J. C. J. H. (2012). Uncertainty and sensitivity analysis of coastal flood damage estimates in the west of the Netherlands. *Nat. Hazards Earth Syst. Sci.*, 12(4):1045–1058. 3, 4, 11, 112

de Rocquigny, E. (2010). An applied framework for uncertainty treatment and key challenges in hydrological and hydraulic modeling. *Can. J. Civ. Eng.*, 37(7):941–954. 112

de Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice*. Wiley, New York. 24, 25, 26, 33, 40

de Rocquigny, E., Soulaimani, A., and Morse, B. (2010). Uncertainty in hydrological and hydraulic modeling: Editorial Introduction. *Can. J. Civ. Eng.*, 37(7, Sp. Iss. SI):III–V. 178

Delenne, C., Cappelaere, B., and Guinot, V. (2012). Uncertainty analysis of river flooding and dam failure risks using local sensitivity computations. *Reliab. Eng. Syst. Saf.*, 107:171–183. Article in press. 28

DNRM (2002). Guidance on the assessment of tangible flood damages. Report QNRM02081, The State of Queensland, Department of Natural Resources and Mines. 39, 56

EEA, WHO, and JRC (2008). Impacts of Europe's changing climate. 2008 indicator-based assessment. Report EEA No 4/2008, European Environment Agency, Copenhagen, Denmark. 111

Egorova, R., Van Noortwijk, J., and Holterman, S. (2008). Uncertainty in flood damage estimation. *Int. J. River. Basin. Manage.*, 6:139–148. 122

Eleuterio, J., Payraudeau, S., and Rozan, A. (2008). Sensibilité de l'évaluation des dommages associés aux inondations en fonction de la caractérisation de la vulnérabilité des bâtiments. *Ingénieries Eau-Agriculture-Territoires*, 55-56:29–44. 55

Erdlenbruch, K., Germano, V., Gilbert, E., Grelot, F., and Lescoulier, C. (2007). Étude socio-économique des inondations sur le bassin versant de l'Orb. Rapport GRI 60306K, Conseil Général de l'Hérault. 60, 61

Erdlenbruch, K., Gilbert, E., Grelot, F., and Lescouliers, C. (2008). Une analyse coût-bénéfice spatialisée de la protection contre les inondations : application de la méthode des dommages évités à la basse vallée de l'Orb. *Ingénieries Eau-Agriculture-Territoires*, 53:3–20. 39, 50, 53, 60, 112, 113, 114

European Commission (2008). Guide to cost-benefit analysis of investment projects. Technical report, Evaluation Unit, DG Regional Policy. 38, 42, 111

European Commission (2009a). Impact assessment guidelines. Guidelines SEC(2009) 92. 2, 10, 111, 112, 140

European Commission (2009b). Impact assessment guidelines: appendices part III. Guidelines SEC(2009) 92. 25

European Commission (2010). Risk assessment and mapping guidelines for disaster management. Commission staff working paper SEC(2010) 1626 final. 41

Fisher, P., Comber, A., and Wadsworth, R. (2005). *Nature de l'incertitude pour les données spatiales*. Éditions Lavoisier, Paris. 2, 10, 99

Fisher, P. F. (1991). Modelling soil map-unit inclusions by Monte Carlo simulation. *Int. J. Geogr. Inf. Sci.*, 5(2):193–208. 99, 121

Francos, A., Elorza, F. J., Bouraoui, F., Bidoglio, G., and Galbiati, L. (2003). Sensitivity analysis of distributed environmental simulation models: understanding the model behaviour in hydrological studies at the catchment scale. *Reliab. Eng. Syst. Saf.*, 79(2):205 – 218. 83

Gatelli, D., Kucherenko, S., Ratto, M., and Tarantola, S. (2009). Calculating first-order sensitivity measures: A benchmark of some recent methodologies. *Reliab. Eng. Syst. Saf.*, 94(7):1212–1219. 33

GERI (2012). Analyse coût-bénéfices des projets de protection-prévention des inondations. Rapport technique, Comité départemental de l'eau du Gard, Groupe d'Échange Risque Inondation. 39

Gilbert, E. and Ledoux, B. (2012). Analyse coût/bénéfice du schéma d'optimisation des zones d'expansion des crues du Rhône, en aval de Viviers. Rapport technique, DREAL Rhône-Alpes. 55, 70, 216

Girres, J.-F. and Julien, P. (2010). Estimation of digitizing and polygonal approximation errors in the computation of length in vector databases. In *Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Leicester, UK. 100, 121

Gleyze, J.-F. (2005). *La vulnérabilité structurelle des réseaux de transport dans un contexte de risques*. Thèse de doctorat, spécialité géographie, Université de Paris VII, Paris. 179

Gouldby, B. P., Sayers, P. B., Panzeri, M. C., and Lanyon, J. E. (2010). Development and application of efficient methods for the forward propagation of epistemic uncertainty and sensitivity analysis within complex broad-scale flood risk system models. *Can. J. Civ. Eng.*, 37(7):955–967. 112

Grelot, F. (2009). Projet EPI: Perception du risque et évaluation économique de l'exposition aux inondations - étude de deux territoires aux contextes hydrologiques différents. Rapport final du programme de recherche Risque-Décision-Territoire, Cemagref, UMR G-EAU. 70

Grelot, F., Bailly, J.-S., Blanc, C., Erdlenbruch, K., Meriaux, P., Saint-Geours, N., and Tourment, R. (2009). Sensibilité d'une analyse coût-bénéfice: enseignements pour l'évaluation des projets d'atténuation des inondations. *Ingénieries Eau-Agriculture-Territoires*, Spécial inondations:95–108. 215

Grelot, F., Enjalbert, P., Agenais, A.-L., Bailly, J.-S., Brémond, P., Langer, T., and Saint-Geours, N. (2012). Programme d'actions de prévention des inondations sur les bassins versants de l'Orb et du Libron : compléments à l'étude ACB de 2007. Rapport technique, IRSTEA, Syndicat Mixte Béziers - La Mer. 57, 60, 61, 64, 113, 116

Gumiere, S. J. (2009). *Contribution à la modélisation déterministe spatialisée de l'érosion hydrique des sols à l'échelle des petits bassins versants cultivés*. PhD thesis, Centre International D'Études Supérieures en Sciences Agronomiques Montpellier SupAgro. 87

Gumiere, S. J., Raclot, D., Cheviron, B., Davy, G., Louchart, X., J.-C., F., Moussa, R., and Le Bissonnais, Y. (2011). MHYDAS-Erosion: a distributed single-storm water erosion model for agricultural catchments. *Hydrol. Processes*, 25(11):1717–1728. 87

Haasnoot, M. and Van Dewolfshaar, K. (2009). Combining a conceptual framework and a spatial analysis tool, HABITAT, to support the implementation of river basin management plans. *Int. J. River. Basin. Manage.*, 7(4):295–311. 18

Helton, J. and Davis, F. (2003). Latin Hypercube Sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Saf.*, 81(1):23 – 69. 107

Helton, J. C. and Davis, F. J. (2006). Sampling-based methods for uncertainty and sensitivity analysis. *Multimed. Environ. Modell.*, 32(2):135–154. 25, 103, 112

Hengl, T. (2006). Finding the right pixel size. *Comput. Geosci.*, 32:1283–1298. 121

Hession, S., Shortridge, A. M., and Torbick, N. M. (2006). Categorical models for spatial data uncertainty. In *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.* 99

Heuvelink, G. B. M. (1998). Uncertainty analysis in environmental modelling under a change of spatial scale. *Nutr. Cycling Agroecosyst.*, 50:255–26. 21, 141

Heuvelink, G. B. M., Brus, D. J., and Reinds, G. (2010a). Accounting for spatial sampling effects in regional uncertainty propagation analysis. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Ressources and Environmental Sciences*, pages 85–88, Leicester, UK, July 20-23. 23, 141

Heuvelink, G. B. M., Burgers, S. L. G. E., Tiktak, A., and Den Berg, F. V. (2010b). Uncertainty and stochastic sensitivity analysis of the GeoPEARL pesticide leaching model. *Geoderma*, 155(3-4):186–192. 81, 141, 150

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, 19:293–325. 29

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.*, 52(1):1–17. 31, 89, 219

Hubert, G. and Ledoux, B. (1999). *Le coût du risque. Évaluation des impacts socio-économiques des inondations*. Presses de l'ENPC. 38

Hunsaker, C., Goodchild, M. F., Friedl, M., and Case, T. (2001). *Spatial uncertainty in ecology*. Springer. 99

INSEE (2008). French classification of activities - NAF Rev. 2 2008. Available at www.insee.fr/en/methods/default.asp?page =nomenclatures/naf2008/naf2008.htm. Accessed 12 May 2012. 62, 115

Iooss, B. (2008). Treatment of spatially dependent input variables in sensitivity analysis of model output methods. Rapport technique, CEA. 79

Iooss, B. (2011). Revue sur l'analyse de sensibilité globale de modèles numériques. *JSFdS*, 152(1):1–23. 25, 27

Iooss, B. and Ribatet, M. (2009). Global sensitivity analysis of computer models with functional inputs. *Reliab. Eng. Syst. Saf.*, 94:1194 – 1204. 79, 86, 111

ISIS (2012). Isis Flow user guide. Available at www.halcrow.com.isis. Accessed 12 May 2012. 61, 115

ISL (2011). Étude du renforcement de la digue du Rhône rive droite entre Beaucaire et Fourques. Rapport définitif, SYMADREM. 50, 70, 216

Jansen, M., Rossing, W., and Daamen, R. (1994). Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In Grasman, J. and van Straten, G., editors, *Predictability and Nonlinear Modelling in Natural Sciences and Economics*, pages 334–343. Kluwer, Dordrecht. 28, 32

Jensen, S. (1998). Mapping human exposure to traffic air pollution using gis. *J. Hazard. Mater.*, 61(1-3):385–392. 18

Jonkman, S., Brinkhuis-Jak, M., and Kok, M. (2004). Cost benefit analysis and flood damage mitigation in the Netherlands. *Heron*, 49(1):95–111. 39, 45

Journel, A. and Huijbregts, C. J. (1978). *Mining geostatistics*. The Blackburn Press, Caldwell. 141, 145, 150, 155, 230, 231, 232, 233, 240, 243

Khaliq, M., Ouarda, T., Ondo, J.-C., Gachon, P., and Bobée, B. (2006). Frequency analysis of a sequence of dependent and or non-stationary hydro-meteorological observations: a review. *J. Hydrol.*, 329 (3-4):534–552. 119

Koivumäki, L., Alho, P., Lotsari, E., Käyhkö, J., Saari, A., and Hyyppä, H. (2010). Uncertainties in flood risk mapping: a case study on estimating building damages for a river flood in Finland. *J. Flood Risk Manage.*, 3(2):166–183. 3, 11, 112, 119, 121, 122, 129

Kucherenko, S., Albrecht, D., and Saltelli, A. (2011). Comparison of Latin Hypercube and Quasi Monte Carlo sampling techniques. 98

Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., and Shah, N. (2009). Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliab. Eng. Syst. Saf.*, 94:1135–1148. 28

Kucherenko, S., Tarantola, S., and Annoni, P. (2012). Estimation of global sensitivity indices for models with dependent variables. *Comput. Phys. Commun.*, 183(4):937–946. 29, 79

Kunz, M., Grêt-Regamey, A., and Hurni, L. (2011). Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. *Nat. Hazards*, 59(3):1735–1751. 180

Kutschera, G. (2009). What are the uncertainties when calculating flood damage potentials? *Wasser Wirtschaft*, 99(10):21–27. 112, 122

Kyriakidis, P. C. (2005). Sequential spatial simulation using Latin Hypercube Sampling. In Leuangthong, O. and Deutsch, C., editors, *Geostatistics Banff 2004*, volume 14 of *Quantitative Geology and Geostatistics*, pages 65–74. Springer. 110

Lamboni, M. (2009). *Analyse de sensibilité pour les modèles dynamiques utilisés en agronomie et environnement*. PhD thesis, Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech). 29

Lamboni, M., Iooss, B., Popelin, A.-L., and Gamboa, F. (2012). Derivative-based global sensitivity measures: general links with Sobol' indices and numerical tests. 28

Lamboni, M., Monod, H., and Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliab. Eng. Syst. Saf.*, 96(4):450–459. 175

Landcare Research (2011). AquiferSim user documentation. Technical report, IRAP Integrated Research for Aquifer Protection, Lincoln Ventures, Landcare Research. 18

Langer, T. (2011). Mise en place d'outils pour l'analyse coût bénéfice d'aménagements de prévention du risque d'inondation : croisement aléa, enjeux et incertitudes. Master's thesis, Université de Toulouse, Master 2 SIGMA. 56, 212, 213, 215

Lantuéjoul, C. (2002). *Geostatistical simulation: Models and Algorithms*. Springer. 99

Lavergne, C. (2006). Contributions à l'analyse de sensibilité. Lesson at Université Montpellier II. 29

Ledoux, B. (2006). *La gestion du risque inondation*. Ed. Lavoisier. 38, 41

Ledoux, B., Grelot, F., and Reliant, C. (2003). *Synthèse des évaluations socio-économiques des instruments de prévention des inondations*. Ministère de l'Écologie et du Développement Durable, Direction des Études Économiques et de l'Évaluation Environnementale. 38

Ledoux Consultants (2010). Méthode standard pour l'analyse coût-bénéfice des projets de prévention des inondations. Guide d'accompagnement, Mission Rhône, DREAL Rhône-Alpes, Service Prévention des Risques. 39

Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E., and Schoendorf, J. (2010). Global sensitivity analysis for systems with independent and/or correlated inputs. *J. Am. Chem. Soc.* Published on Web 04/26/2010. 29, 79

Lilburne, L., Hewitt, A., and Ferriss, S. (2006). Progress with the design of a soil uncertainty database, and associated tools for simulating spatial realisations of soil properties. In *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. 100

Lilburne, L. and Tarantola, S. (2009). Sensitivity analysis of spatial models. *Int. J. Geogr. Inf. Sci.*, 23(2):151–168. 33, 79, 84, 85, 103, 106, 124, 140, 141, 144, 147, 150

Liu, Q. and Homma, T. (2010). A new importance measure for sensitivity analysis. *J. Nucl. Sci. Technol.*, 47(1):53–61. 28, 175

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartogr. Geogr. Inform.*, 32(3):139–160. 100

Maidment, D. (1993). *Handbook of Hydrology*. McGraw-Hill, New York. 119

Mara, T. and Tarantola, S. (2012). Variance-based sensitivity indices for models with dependent inputs. *Reliab. Eng. Syst. Saf.*, 107:115–121. 29, 79

Marrel, A., Iooss, B., Da Veiga, S., and Ribatet, M. (2012). Global sensitivity analysis of stochastic computer models with joint metamodels. *Stat. Comput.*, 22(3):833–847. 86

Marrel, A., Iooss, B., Jullien, M., Laurent, B., and Volkova, E. (2011). Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, 22(3):383–397. 88, 141, 147, 150, 170

MEDD (2004). Les inondations. Dossier d'information, 20 p., Ministère de l'Écologie et du Développement Durable, Direction de la Prévention de la Pollutions et des Risques, SDPRM. 38

MEDDTL (2011). Programmes d'action de prévention des inondations (PAPI) : de la stratégie aux programmes d'action. Cahier des charges, Ministère de l'Écologie, du Développement durable, des Transports et du Logement. 38, 39, 48, 50, 54, 64, 116

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A. (2010). Review article on assessment of economic flood damage. *Nat. Hazards Earth Syst. Sci.*, 10(8):1697–1724. 39, 55, 56, 57, 112, 116, 122

Merz, B., Kreibich, H., Thieken, A. H., and Schmidtke, R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. *Nat. Hazards Earth Syst. Sci.*, 4(1):153–163. 3, 11, 112, 122

Merz, B. and Thieken, A. (2009). Flood risk curves and uncertainty bounds. *Nat. Hazards*, 51 (3):437–458. 112, 122

Messner, F., Penning-Rowsell, E., Green, C., Meyer, V., Tunstall, S., and van der Veen, A. (2007). Evaluating flood damages: guidance and recommendations on principles and methods. Report T09-06-01, FLOODsite Consortium. 38, 39, 45, 50, 54, 111, 116

Mishra, S., Deeds, N. E., and Ramarao, B. S. (2003). Application of classification trees in the sensitivity analysis of probabilistic model results. *Reliab. Eng. Syst. Saf.*, 79(2):123 – 129. 28

Monod, H. (2012). Méthodes pour l'exploration numérique de modèles agro-écologiques à l'échelle du parcellaire agricole. In *Colloque MAS 2012 - Concentration et applications*, Clermont-Ferrand, 29-31 August. 88

Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174. 28

Morss, R. E., Wilhelmi, O. V., Downton, M. W., and Gruntfest, E. (2005). Flood risk, uncertainty, and scientific information for decision making. *B. Am. Meteorol. Soc.*, pages 1593–1601. 128

Mostert, E. and Junier, S. J. (2009). The European flood risk directive: challenges for research. *Hydrol. Earth Syst. Sci. Discuss.*, 6:4961–4988. 112

Neppel, L., Renard, B., Lang, M., Ayral, P., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., and Vinet, F. (2010). Flood frequency analysis using historical data: accounting for random and systematic errors. *Hydrol. Sci. J.*, 55:192–208. 119

Neubert, M., Herold, H., and Meinel, G. (2008). *Object-based image analysis*, chapter Assessing image segmentation quality – concepts, methods and application, pages 769–784. Springer Berlin Heidelberg. 100

Novak, V., Guinot, V., Jeffrey, A., and Reeve, D. (2010). *Hydraulic modelling - an introduction: principles, methods and applications*. Spon Press. 53

NRC (2000). *Risk analysis and uncertainty in flood damage reduction studies*. The National Academies Press. 45

O'Hagan, A. (2012). Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux. *Environ. Modell. Softw.*, 36(SI):35–48. Article in Press. 26

Oreskes, N., Shraderfrechette, K., and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147):641–646. 2, 9

Ostendorf, B. (2011). Overview: spatial information and indicators for sustainable management of natural resources. *Ecol. Indic.*, 11:97–102. 1, 9

Pappenberger, F. and Beven, K. J. (2006). Ignorance is bliss: or seven reasons not to use uncertainty analysis. *Water Resour. Res.*, 42(5). 112

Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Adv. Water Resour.*, 31:1–14. 3, 11, 112

Pappenberger, F., Harvey, H., Beven, K., Hall, J., and Meadowcroft, I. (2006). Decision tree for choosing an uncertainty analysis methodology: a wiki experiment. *Hydrol. Processes*, 20(17):3793–3798. 25, 112

Parry, M. L., Canziani, O., Palutikof, J., van der Linden, P., and Hanson, C. (2007). Climate change 2007: Impacts, adaptation and vulnerability. Contribution of working group II to the fourth assessment report of the Intergovernmental Panel on Climate Change. Summary for policymakers, IPCC. 38

Pebesma, E. J. and Heuvelink, G. B. M. (1999). Latin Hypercube Sampling of Gaussian random fields. *Technometrics*, 41(4). Reprinted. 103, 105, 107, 109

Penning-Rowsell, E., Johnson, C., Tunstall, S., Tapsell, S., Morris, J., Chatterton, J., and Green, C. (2005). The benefits of flood and coastal risk management: a handbook of assessment techniques. Technical report, Middlesex University Press. 39, 215

Pettit, C. and Wilson, D. (2010). Full-field sensitivity analysis through dimension reduction and probabilistic surrogate models. *Probab. Eng. Mech.*, 25(4):380–392. 141, 147, 150, 170

Pronzato, L. and Mueller, W. G. (2012). Design of computer experiments: space filling and beyond. *Stat. Comput.*, 22(3 (SI)):681–701. 32

R Development Core Team (2009). R: a language and environment for statistical computing. 147

Radoux, J., Bogaert, P., Fasbender, D., and Defourny, P. (2011). Thematic accuracy assessment of geographic object-based image classification. *Int. J. Geogr. Inf. Sci.*, 25(6 (SI)):895–911. 100

Ratto, M., Castelletti, A., and Pagano, A. (2012). Emulation techniques for the reduction and sensitivity analysis of complex environmental models. *Environ. Modell. Softw.*, 34:1–4. 26

Refsgaard, J. C., Van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process: a framework and guidance. *Environ. Modell. Softw.*, 22:1543–1556. 26, 45, 112, 113, 140

Remy, N., Boucher, A., and Wu, J. (2009). *Applied Geostatistics with SGeMS: A User's Guide*. Cambridge University Press. 99

Ruffo, P., Bazzana, L., Consonni, A., Corradi, A., Saltelli, A., and Tarantola, S. (2006). Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques. *Reliab. Eng. Syst. Saf.*, 91(10-11):1155–1162. 86, 141

Saint-Geours, N., Bailly, J.-S., Grelot, F., and Lavergne, C. (2010). Latin Hypercube Sampling of Gaussian random field for Sobol' global sensitivity analysis of models with spatial inputs and scalar output. In *Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, pages 81–84, Leicester, UK, 20-23 July. 76, 98, 101, 110, 141

Saint-Geours, N., Lavergne, C., Bailly, J.-S., and Grelot, F. (2011a). Analyse de sensibilité de Sobol d'un modèle spatialisé pour l'évaluation économique du risque d'inondation. *JSFdS*, 152(1):24–46. 121, 150

Saint-Geours, N., Lavergne, C., Bailly, J.-S., and Grelot, F. (2011b). Sensitivity analysis of spatial models using geostatistical simulation. In *Mathematical Geosciences at the Crossroads of Theory and Practice, Proceedings of IAMG 2011 conference*, pages 178–189, Salzburg, Austria, September 5-9. 144

Saint-Geours, N., Lavergne, C., Bailly, J.-S., and Grelot, F. (2012). Change of support in variance-based spatial sensitivity analysis. *Math. Geosci.*, 44(8):945–958. 129, 138, 140, 151, 152

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output: design and estimator for the total sensitivity index. *Comput. Phys. Commun.*, 181(2):259 – 270. 33, 201, 202

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis - The Primer*. Wiley. 24, 25, 27, 29, 31, 33, 40, 103, 104, 112, 117, 125, 143, 179

Saltelli, A., Tarantola, S., and Chan, K. (1999). A quantitative, model independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56. 28, 32

Scheidt, C. and Caers, J. (2009). Representing spatial uncertainty using distances and kernels. *Math. Geosci.*, 41(4):397–419. 110, 175

Schlather, M. (2001). Simulation of stationary and isotropic random fields. *R-News*, 1(2):18–20. 99, 147

Skøien, J. and Blöschl, G. (2006). Sampling scale effects in random fields and implications for environmental monitoring. *Environ. Monit. Assess.*, 114:521–552. 153

SMVOL (2011). Programme d'actions de prévention des inondations sur les bassins de l'Orb et du Libron (34) pour les années 2011 à 2015. Note de présentation, Syndicat Mixte des Vallées de l'Orb et du Libron. 55, 59, 113

Sobol', I. (1993). Sensitivity analysis for non-linear mathematical model. *Modelling and Computational Experiments*, 1:407–414. 30, 32, 140

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comp. Math. Math. Phys.*, 7:86–112. 28, 124

Stephens, E., Bates, P., Freer, J., and Mason, D. (2012). The impact of uncertainty in satellite data on the assessment of flood inundation models. *J. Hydrol.*, 414-415:162–173. 112

Tarantola, S., Gatelli, D., and Mara, T. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliab. Eng. Syst. Saf.*, 91:717–727. 28, 32

Tarantola, S., Giglioli, N., Jesinghaus, J., and Saltelli, A. (2002). Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making ? *Stoch. Env. Res. Risk A.*, 16:63–76. 2, 10, 25, 112, 140

Te Linde, A. H., Bubeck, P., Dekkers, J. E. C., de Moel, H., and Aerts, J. C. J. H. (2011). Future flood risk estimates along the river Rhine. *Nat. Hazards Earth Syst. Sci.*, 11(2):459–473. 3, 11, 112, 121

Thomson, J., Hetzler, E., and MacEachren, A. (2005). A typology for visualizing uncertainty. In Erbacher, R. F., Roberts, J. C., and Grohn, M. T., editors, *Conference on Visualization and Data Analysis 2005*, volume 5669 of *Proceedings of the Society of photo-optical instrumentation engineers (SPIE)*, pages 146–157, San Jose, CA, Jan 17-18. 180

Tiktak, A., De Nie, D., Van Der Linden, A., and Kruijne, R. (2002). Modelling the leaching and drainage of pesticides in the Netherlands: the GeoPEARL model. *Agronomie*, 22:373–387. 23

Tong, C. (2010). Self-validated variance-based methods for sensitivity analysis of model outputs. *Reliab. Eng. Syst. Saf.*, 95(3):301–309. cited By (since 1996) 0. 32

Torterotot, J.-P. (1993). *Le coût des dommages dus aux inondations : Estimation et analyse des incertitudes*. Thèse de doctorat, spécialité sciences et techniques de l'environnement, École Nationale des Ponts et Chaussées, Paris. 2 volumes. 122

Van Dantzig, D. (1956). Economic decision problems for flood prevention. *Econometrica*, 24:276–287. 45

Van Der Knijff, J. M., Younis, J., and De Roo, A. P. J. (2010). LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Inf. Sci.*, 24(2):189–212. 18

Van der Sluijs, J. P., Risbey, J. S., Kloprogge, P., Ravetz, J. R., Funtowicz, S. O., Quintana, S. C., Guimaraes Pereira, A., De Marchi, B., Petersen, A. C., Janssen, P. H. M., Hoppe, R., and Huijs, S. W. F. (2003). RIVM/MNP Guidance for uncertainty assessment and communication. Detailed guidance, Utrecht University. 180

Vasseure, B., Jeansoulin, R., Devilliers, R., and Frank, A. U. (2005). *Évaluation de la qualité externe de l'information géographique : une approche ontologique*, chapter 15, pages 293–307. Éditions Lavoisier, Paris. 2, 10

Viard, T., Caumon, G., and Levy, B. (2011). Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions? *Comput. Geosci.*, 37(4):511–520. 180

Vidard, A. (2011). Méthodes déterministes d'analyse de sensibilité. In *Journée de l'Atelier Modélisation du Cluster Environnement : Méthodes d'analyse de sensibilité*, Grenoble, France. 28

Visser, H., Petersen, A. C., Beusen, A. H. W., Heuberger, P. S. C., and Janssen, P. H. M. (2006). Guidance for uncertainty assessment and communication: checklist for uncertainty in spatial information and visualising spatial uncertainty. Report 550032001/2006, Netherlands Environmental Assessment Agency. 180

Volkova, E., Iooss, B., and Van Dorpe, F. (2008). Global sensitivity analysis for a numerical model of radionuclide migration from the RRC Kurchatov Institute radwaste disposal site. *Stoch. Env. Res. Risk A.*, 22(1):17–31. 83

Walker, W., Harremoees, P., Rotmans, J., Van der Sluijs, J., van Asselt, M., Janssen, P., and Krayer von Krauss, M. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integr. Assess.*, 4(1):5–17. 2, 10, 26, 112

Wechsler, S. (2007). Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrol. Earth Syst. Sci.*, 11(4):1481–1500. 119

Weichel, T., Pappenberger, F., and Schulz, K. (2007). Sensitivity and uncertainty in flood inundation modelling: concept of an analysis framework. *Adv. Geosci.*, 11:31–36. 112

Wu, J., Jones, K. B., Li, H., and Loucks, O. L., editors (2006). *Scaling and uncertainty analysis in ecology: methods and applications*. Springer. 20

Xu, C. and Gertner, G. (2011). Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). *Comput. Stat. Data Anal.*, 55(1):184 – 198. 28

List of publications

Refereed journal publications

Saint-Geours, N., Lavergne, C., Bailly, J.-S. and Grelot, F. (2012), 'Change of support in variance-based spatial sensitivity analysis', *Mathematical Geosciences*, 44(8) 945-958.
→ **paper reproduced in §4.1 on page 140**

Saint-Geours, N., Lavergne, C., Bailly, J.-S. and Grelot, F. (2011), 'Analyse de sensibilité de Sobol d'un modèle spatialisé pour l'évaluation économique du risque d'inondation', *Journal de la Société Française de Statistique*, 152(1), 24-46.

Grelot, F., Bailly, J.-S., Blanc, C., Erdlenbruch, K., Meriaux, P., Saint-Geours, N. and Tourment, N. (2009), 'Sensibilité d'une analyse coût-bénéfice : enseignements pour l'évaluation des projets d'atténuation des inondations', *Ingénieries Eau-Agriculture-Territoires*, Spécial inondations, 95-108.

Manuscript submitted or in press

Saint-Geours, N., Lavergne, C., Bailly, J.-S. and Grelot, F., 'Ranking sources of uncertainty in flood damage modelling: a case study on the cost-benefit analysis of a flood mitigation project in the Orb Delta, France', submitted to *Journal of Flood Risk Management*.
→ **paper reproduced in §3.3 on page 111**

Abdallah, H., Bailly, J.-S., Baghdadi, N., Saint-Geours, N. and Fabre, F. 'Potential of space-borne LiDAR sensors for global bathymetry in coastal and inland waters', *IEEE Transactions on Magnetics*, Accepted July 2012.

International conferences (with proceedings paper)

Saint-Geours, N., Lavergne, C., Bailly, J.-S. and Grelot, F. (2011) Analyse de sensibilité globale d'un modèle d'évaluation économique du risque d'inondation, *in* '43ème Journées de Statistique de la SFdS', Gammarth, Tunisia, May 23-27.
→ **extended abstract reproduced in Appendices**

Saint-Geours, N., Lavergne, C., Bailly, J.-S. and Grelot, F. (2011) Sensitivity analysis of spatial models using geostatistical simulation, *in* 'Mathematical Geosciences at the Crossroads of Theory and Practice,

Proceedings of IAMG 2011 conference', Salzburg, Austria, September 5-9, pp. 178-189.
→ **extended abstract reproduced in Appendices**

Saint-Geours, N., Bailly, J.-S., Grelot, F. and Lavergne, C. (2010), Latin Hypercube Sampling of Gaussian random field for Sobol global sensitivity analysis of models with spatial inputs and scalar output, *in* 'Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences', Leicester, UK, pp. 81-84.
→ **short paper reproduced in §3.2 on page 98**

Saint-Geours, N. and Lilburne, L. (2010), Comparison of three spatial sensitivity analysis techniques, *in* 'Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences', Leicester, UK, pp. 421-424.
→ **short paper reproduced in Appendices**

International conferences (no proceedings paper)

Saint-Geours, N., Lavergne, C., Bailly, J.-S. and Grelot, F. (2012) Cost-benefit analysis of a flood mitigation policy: an uncertainty and sensitivity analysis, *in* 'European Geosciences Union, General Assembly 2012', Vienna, Austria, April 19-24.

Abdallah, H., Bailly, J.-S., Baghdadi, N. and Saint-Geours, N. (2010) Sensitivity analysis of simulated LIDAR waveforms according to sensor and water parameters variability, *in* '2010 AGU Fall Meeting', San Francisco, USA, December 13-17.

Grelot, F., Bailly, J.-S. and Saint-Geours, N. (2009) Validity of cost-benefit analysis for flood prevention projects: insights from sensitivity analysis, *in* 'European Geosciences Union, General Assembly 2009', Vienna, Austria, April 19-24. Geophysical Research Abstracts, 11.

Other presentations, workshops (no proceedings paper)

(2012) *in* 'Journés MAS de la Société de Mathématiques Appliquées et Industrielles', Clermont-Ferrand, August 29-31, oral presentation.

(2011) *in* 'DOCTISS Post-graduate school day', Montpellier, May 19, poster abstract.

(2011) *in* 'GDR MASCOT NUM Workshop', Villard-de-Lans, March 23, oral presentation.

(2011) *in* 'Seminar of Statistics of Laboratoire Jean Kuntzmann', Grenoble, Feb 3, oral presentation.

(2011) *in* 'Seminar on IPEF post-graduate training', Marne-la-Vallée, Jan 27, oral presentation.

(2010) *in* 'Journées d'étude inter-Zones Ateliers : risque d'inondation, vulnérabilité et résilience: regards croisés Loire Rhône', Blois, Dec 3, oral presentation.

(2010) *in* 'UMR TETIS Post-grad day', Montpellier, Dec 2, oral presentation.

(2010) *in* 'Seminar on statistical and computer tools for post-grad students in Agronomy and Environment of Languedoc-Roussillon', April 28, oral presentation.

(2010) *in* 'GDR MASCOT-NUM Workshop', Avignon, March 18, poster abstract.

Technical reports

Grelot, F., Saint-Geours, N., and Bailly, J.-S. (2012) ‘Guide méthodologique pour la prise en compte des incertitudes dans les ACB appliquées aux projets de prévention du risque d’inondation’, Document en cours d’élaboration pour le compte du Plan Rhône.

Appendices

A	Appendix to §1.2: details and proofs	200
B	Appendix to §1.2: VB-GSA of a model with sub-components	203
C	Appendix to §2.2.5: definition of the Average Annual Damages	206
D	Appendix to §2.2.8.2: overlay analysis for flood exposure assessment	212
E	Appendix to §2.3.3: extra case studies	215
F	Appendix to §3.1: details on the numerical study	219
G	Appendix to §3.3: sensitivity indices with respect to the NPV indicator	226
H	Appendix to §4.1: proofs	230
I	Appendix to §4.1: case of a linear model	237
J	Appendix to §4.2: proofs	242
K	Short paper published in the proceedings of IAMG 2011 conference	247
L	Short paper published in the proceedings of the Accuracy 2010 conference	260
M	Extended abstract - 43th 'Journées de Statistique de la SFdS'	264

A Appendix to §1.2: details and proofs

We give in this appendix some details on VB-GSA basics which are briefly presented in §1.2 on page 24.

A.1 Expression of elementary functions f_α

We consider the elementary functions f_α defined in the functional decomposition Eqn. (1.4) on page 29. We assume that the condition given in Eqn. (1.5) on page 29 is met. We want to give here a proof of the general expression of elementary functions f_α given in Eqn. (1.8) on page 30:

$$\forall \alpha \subseteq \{1,\cdots,K\}, \quad f_\alpha = \sum_{\beta\subseteq\alpha} (-1)^{|\alpha|-|\beta|} \cdot \mathbb{E}(Y \mid \mathbf{U}_\alpha) \tag{A.1}$$

in which $|\alpha|$ is the cardinal of subset $\alpha \subseteq \{1,\cdots,K\}$.

Proof. We prove the general expression of elementary functions f_α by induction.

Consider the following statements $\mathcal{P}_n$ for $n \in \mathbb{N}$:

$$\mathcal{P}_n : \quad \forall \alpha \subseteq \{1,\cdots,K\}, \quad |\alpha| \leq n \quad \Rightarrow f_\alpha = \sum_{\beta\subseteq\alpha} (-1)^{|\alpha|-|\beta|} \cdot \mathbb{E}(Y \mid \mathbf{U}_\alpha) \tag{A.2}$$

Base case By integrating the functional decomposition given in Eqn. (1.4) on page 29 with respect to all model inputs $U_1,\ldots,U_K$, under condition Eqn. (1.5), we get:

$$\int_{[0;1]^K} f(\mathbf{U}) d\mathbf{U} = f_\varnothing \tag{A.3}$$

thus:

$$f_\varnothing = \mathbb{E}(Y) \tag{A.4}$$

Hence, the particular statement $\mathcal{P}_0$ is true.

Inductive hypothesis Consider $n \in \mathbb{N}$. Let us assume we know $\mathcal{P}_n$ is true.

Inductive step Let consider α a subset of $\{1,\cdots,K\}$ such that $|\alpha| = n+1$. Let consider $\bar{\alpha} = \{1,\ldots,K\} \setminus \alpha$ the complementary of α. By integrating the functional decomposition given in Eqn. (1.4) on page 29 with respect to each model input U_i such that $i \in \bar{\alpha}$, under condition Eqn. (1.5) on page 29, we get:

$$\begin{aligned}\int_{[0;1]^{K-n-1}} f(\mathbf{U}_{\bar{\alpha}}) d\mathbf{U}_{\bar{\alpha}} &= \sum_{\beta\subseteq\alpha} f_\beta(\mathbf{U}_\beta) \\ &= f_\alpha(\mathbf{U}_\alpha) + \sum_{\beta\subset\alpha} f_\beta(\mathbf{U}_\beta)\end{aligned}$$

Thus:

$$\begin{aligned}f_\alpha(\mathbf{U}_\alpha) &= \int_{[0;1]^{K-n-1}} f(\mathbf{U}_{\bar{\alpha}}) d\mathbf{U}_{\bar{\alpha}} - \sum_{\beta\subset\alpha} f_\beta(\mathbf{U}_\beta) \\ &= \mathbb{E}(f|\mathbf{U}_\alpha) - \sum_{\beta\subset\alpha} f_\beta(\mathbf{U}_\beta) \quad \text{by definition of } \mathbb{E}(f|\mathbf{U}_\alpha)\end{aligned}$$

The elementary function f_α can thus be written as:

$$\begin{aligned}
f_\alpha &= \mathbb{E}(f|\mathbf{U}_\alpha) - \sum_{\beta\subset\alpha} f_\beta \\
&= \mathbb{E}(f|\mathbf{U}_\alpha) - \sum_{\beta\subset\alpha}\sum_{\delta\subseteq\beta}(-1)^{|\beta|-|\delta|}\mathbb{E}(f|\mathbf{U}_\delta) \quad \text{(according to } \mathcal{P}_n\text{)} \\
&= \mathbb{E}(f|\mathbf{U}_\alpha) - \sum_{\delta\subset\alpha}\sum_{q=0}^{|\alpha|-|\delta|-1} C^q_{|\alpha|-|\delta|}(-1)^q\mathbb{E}(f|\mathbf{U}_\delta) \quad \text{(with } q\text{: cardinal of } \beta\setminus\delta\text{)} \\
&= \mathbb{E}(f|\mathbf{U}_\alpha) - \sum_{\delta\subset\alpha}\left(\left[\sum_{q=0}^{|\alpha|-|\delta|} C^q_{|\alpha|-|\delta|}(-1)^q\right] - (-1)^{|\alpha|-|\delta|}\right)\mathbb{E}(f|\mathbf{U}_\delta) \\
&= \mathbb{E}(f|\mathbf{U}_\alpha) + \sum_{\delta\subset\alpha}(-1)^{|\alpha|-|\delta|}\mathbb{E}(f|\mathbf{U}_\delta) \quad \text{(binomial formula)} \\
&= \sum_{\delta\subseteq\alpha}(-1)^{|\alpha|-|\delta|}\mathbb{E}(f|\mathbf{U}_\delta)
\end{aligned}$$

Thus statement $\mathcal{P}_{n+1}$ holds.

Conclusion Statement $\mathcal{P}_n$ is true for all $n \in \mathbb{N}$, which proves the general expression of elementary functions f_α given in Eqn. (1.8) on page 30. □

A.2 Pseudo-Monte Carlo estimation procedure of variance-based sensitivity indices

We describe here the sampling procedure that was used throughout this thesis to estimate variance-based sensitivity indices from a set of simulations of a numerical model $\mathcal{F}$. This estimation procedure is based on the recommendations of Saltelli et al. (2010). It is composed of the following steps:

1. generate two input samples M_1 and M_2 of size $K\times N$ in the space of uncertain model inputs, where K is the number of model inputs (or groups of model inputs) and N will be refered to as the "*base sample size*". The two samples are $LP-\tau$ sequences where each input U_j is sampled from its pdf p_j. The i^{th} line of sample M_1 or M_2 is a set $(U_1^{(i)},\ldots,U_K^{(i)})$ where $U_j^{(i)}$ is a random value drawn from pdf of model input U_j. The choice of a base sample size N depends on the accuracy needed for sensitivity indices estimates; it has to be of the form 2^κ to ensure that the desired properties of the $LP-\tau$ sequences hold;

2. new samples are then created by combining original samples M_1 and M_2. For $j=1$ to K, a new sample $M_{1,2}^{(j)}$ is created: it is equal to sample M_1, except for the j^{th} column which is taken from sample M_2 (Figure $\mathcal{A}$.1 on the next page);

3. create a total sample M_{tot} by binding base samples M_1, M_2 and all mixed samples $M_{1,2}^{(j)}$. The size N_{tot} of the total sample will be refered to as the "*total sample size*". N_{tot} depends on the base sample size and on the number of model inputs K (or groups of model inputs): $N_{\text{tot}} = (K+2)\cdot N$;

4. calculate an output vector by evaluating the numerical code $\mathcal{F}$ at each line of the total sample M_{tot}. Each model run gives a value for the output of interest Y. We denote by $Y(M_1)$, $Y(M_2)$ and $Y(M_{1,2}^{(j)})$ the vectors of length N giving the value of Y for each line of samples M_1, M_2 and $M_{1,2}^{(j)}$, respectively;

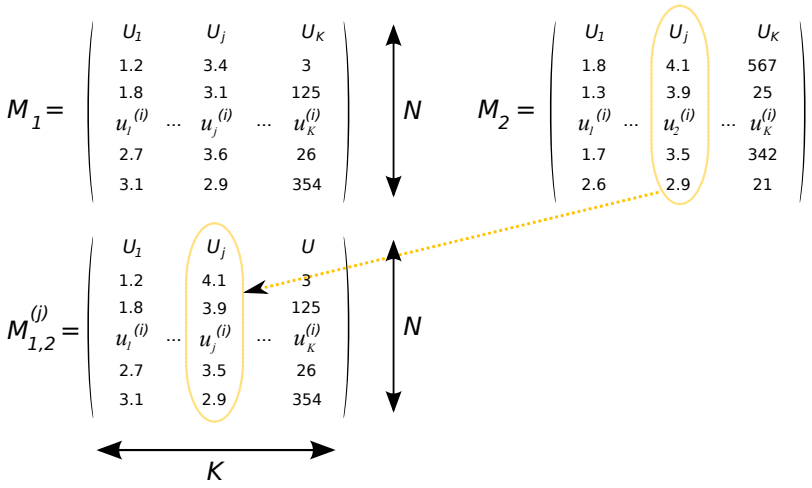

Figure $\mathcal{A}$.1: Creating sample $M_{1,2}^{(j)}$

5. estimate first-order and total-order sensitivity indices for each model input U_i (or each group of model inputs) from the output vectors with Eqn. ($\mathcal{A}$.5) and Eqn. ($\mathcal{A}$.6). These estimators are those suggested by Saltelli et al. (2010).

$$S_j = \frac{\frac{1}{N}\sum_{i=1}^{N} Y(M_2)_i \cdot Y(M_{1,2}^{(j)})_i - \frac{1}{N}\sum_{i=1}^{N} Y(M_2)_i \cdot Y(M_1)_i}{\frac{1}{N}\sum_{i=1}^{N} Y(M_1)_i \cdot Y(M_1)_i - \left(\frac{1}{N}\sum_{i=1}^{N} Y(M_1)_i\right) \cdot \left(\frac{1}{N}\sum_{i=1}^{N} Y(M_2)_i\right)} \qquad (\mathcal{A}.5)$$

$$ST_j = \frac{\frac{1}{2N}\sum_{i=1}^{N} \left[Y(M_1)_i - Y(M_{1,2}^{(j)})_i\right]^2}{\frac{1}{N}\sum_{i=1}^{N} Y(M_1)_i \cdot Y(M_1)_i - \left(\frac{1}{N}\sum_{i=1}^{N} Y(M_1)_i\right) \cdot \left(\frac{1}{N}\sum_{i=1}^{N} Y(M_2)_i\right)} \qquad (\mathcal{A}.6)$$

B Appendix to §1.2: VB-GSA of a model with sub-components

Problématique Le modèle NOE est un modèle construit de manière modulaire : différents sous-modèles (hydraulique, estimation des dommages, calcul de l'indicateur NPV, etc.) sont chaînés pour aboutir au calcul des indicateurs économiques finaux (Chapter 2). De manière générale, on peut représenter la situation de la manière suivante :

- on veut étudier un modèle quelconque $Y = f(U_1, \ldots, U_K)$ avec K variables d'entrées ;
- la variable U_1 est elle-même calculée grâce à un sous-modèle $U_1 = f'(U'_1, \ldots, U'_{K'})$ avec K' sous-variables ;
- on suppose toutes les variables et sous-variables indépendantes entre elles.

On peut alors mener l'analyse de sensibilité du modèle principal f de deux manières : soit en échantillonnant l'ensemble des $K - 1 + K'$ variables d'entrée (exceptée U_1), on obtiendra alors les indices de sensibilité de premier ordre et totaux pour chacune de ces variables ; soit en menant d'abord l'analyse du sous-modèle f', puis en menant l'analyse du modèle global f en utilisant les résultats de l'analyse précédente pour décrire la variabilité du paramètre U_1.

Cette deuxième démarche a été un temps appliquée au modèle NOE pour calculer les indices de sensibilité relatifs à l'indicateur NPV, pour un gain en temps de calcul de l'ordre de 30 %. On donne ici le détail du gain en temps de calcul apporté par cette approche.

Lien entre indices de sensibilité d'un modèle et d'un sous-modèle Dans le contexte défini précédemment, on peut définir plusieurs indices de sensibilité : ceux des sous-variables U'_i par rapport à la variable U_1 ; ceux des variables U_i par rapport à la sortie Y ; ceux des sous-variables U'_i par rapport à la sortie Y. Sous certaines conditions, on peut relier ces différents indices de sensibilité.

On définit la variable aléatoire $\phi = \mathbb{E}(Y \mid U_1)$. L'indice de sensibilité de premier ordre de U'_j par rapport à Y est alors :

$$S^Y_{U'_j} = \frac{\mathrm{var}\left[\mathbb{E}(Y \mid U'_j)\right]}{\mathrm{var}(Y)} = \frac{\mathrm{var}\left[\mathbb{E}(\phi \mid U'_j)\right]}{\mathrm{var}(Y)} \tag{$\mathcal{A}$.7}$$

L'indice de premier ordre de U_1 par rapport à Y est par ailleurs égal à :

$$S^Y_{U_1} = \frac{\mathrm{var}(\phi)}{\mathrm{var}(Y)} \tag{$\mathcal{A}$.8}$$

Et l'indice de premier ordre de U'_j par rapport à ϕ est :

$$S^\phi_{U'_j} = \frac{\mathrm{var}\left[\mathbb{E}(\phi \mid U'_j)\right]}{\mathrm{var}(\phi)} \tag{$\mathcal{A}$.9}$$

On peut donc écrire la relation suivante entre ces trois indices de sensibilité :

$$S^Y_{U'_j} = S^Y_{U_1} \cdot S^\phi_{U'_j} \tag{$\mathcal{A}$.10}$$

Cas particulier Plaçons nous maintenant dans le cas particulier où la variable aléatoire ϕ est une transformation affine de U_1 : $\phi = a \cdot U_1 + b$. Alors l'indice de premier ordre de la sous-variable U'_j par rapport à ϕ est égal à l'indice de sensibilité de U'_j par rapport à U_1 : $S^{\phi}_{U'_j} = S^{U_1}_{U'_j}$. On a alors la relation suivante de transitivité entre les indices de sensibilité du premier ordre du modèle principal f et ceux du sous-modèle f' :

$$S^{Y}_{U'_j} = S^{Y}_{U_1} \cdot S^{U_1}_{U'_j} \tag{$\mathcal{A}$.11}$$

Gain en temps de calcul Notons C le coût en temps de calcul d'une simulation du modèle principal f, et C' le coût d'une simulation du sous-modèle f'. On veut mener l'analyse de sensibilité du modèle principal $Y = f(U_1, \ldots, U_K)$ et estimer les indices de sensibilité de chaque variable principale U_i par une procédure de pseudo-Monte Carlo. Une première option consiste à échantillonner de manière « *brute* » toutes les sous-variables $U'_1, \ldots, U'_{K'}$ et toutes les variables $U_2, \ldots, U_K$ (Figure $\mathcal{A}$.2 on the facing page, gauche). En notant N la taille de base des échantillons de Monte Carlo, le coût total de l'analyse est :

$$C_{brute} = N \cdot (K - 1 + K' + 2) \cdot (C + C') \tag{$\mathcal{A}$.12}$$

Une seconde option est de procéder en deux étapes (Figure $\mathcal{A}$.2 on the next page, droite). On commence par estimer les indices de sensibilité des sous-variables $U'_1, \ldots, U'_{K'}$ dans le sous-modèle f' pour un coût égal à $N \cdot (K' + 2) \cdot C'$. Puis on estime les indices de sensibilité des variables principales $U_1, \ldots, U_K$ du modèle principal f, pour un coût égal à $N \cdot (K + 2) \cdot C$. Le coût total de l'analyse pour cette seconde option est donc :

$$C_{module} = N \cdot [(K' + 2) \cdot C' + (K + 2) \cdot C)] \tag{$\mathcal{A}$.13}$$

L'analyse de sensibilité par module entraîne ainsi un gain de temps égal à :

$$C_{brute} - C_{module} = N \cdot [(K - 1) \cdot C' + (K' - 1) \cdot C] \tag{$\mathcal{A}$.14}$$

Cependant, cette seconde option entraîne aussi une perte d'information : on ne connaît plus les indices de sensibilité des sous-variables U'_j par rapport à la sortie Y, mais seulement les indices des U'_j par rapport à la variable principale U_1 et les indices de la variable principale U_1 par rapport à la sortie Y. Dans le cas particulier où la variable principale U_1 n'intervient qu'à l'ordre 1 dans l'équation définissant Y, alors on connaît aussi les indices de premier ordre des sous-variables U'_j par rapport à la sortie Y grâce à Eqn. ($\mathcal{A}$.11).

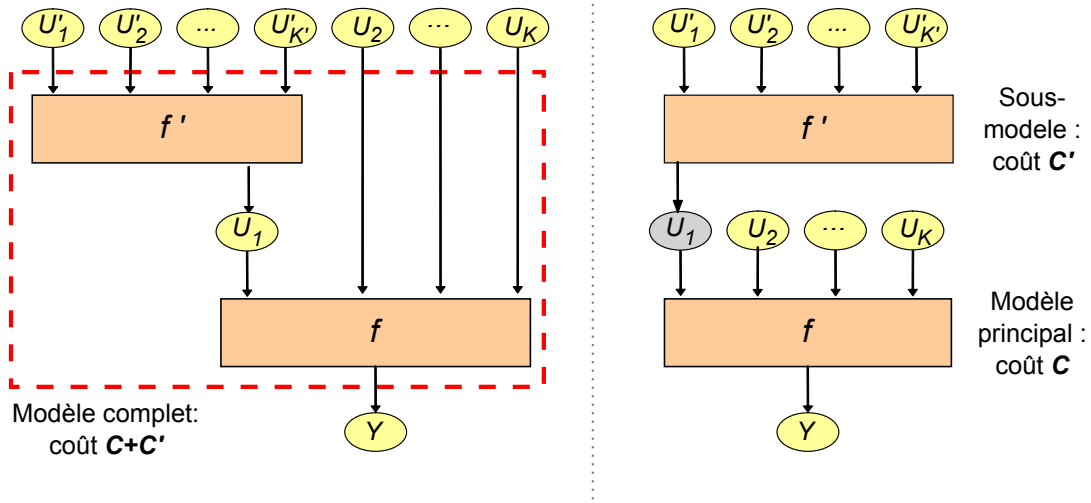

FIGURE $\mathcal{A}$.2 – Estimation des indices de sensibilité : approche brute (gauche), approche par modules (droite)

C Appendix to §2.2.5 : definition of the Average Annual Damages

As mentioned in §2.2.5 on page 44, it is a challenging issue to define rigorously the notion of "*annual avoided damages*". In the literature, the average annual damages (AAD indicator) are usually defined as the integral of flood damages D with respect to the probability of the associated peak discharge q:

$$AAD = \int_{q=0}^{\infty} D(q)p(q)dq \tag{A.15}$$

One of the contributions of this thesis is an attempt to extend the definition of the AAD indicator, in order to account for more than one characteristic of flood scenarios in the computation of the average annual damages. For example, apart from the peak discharge q, other random characteristics of flood scenarios could be taken into account to define the average annual damages: the season of flood occurence, the state of levees along the river stream (failure/no failure), or the location of the possible levee failures. Our attempt proved useful to better identify and discuss the assumptions and limitations that are hidden when the average annual damages are used as a risk indicator.

In this appendix, we display two different formal frameworks to define the AAD indicator. In the first framework, we model the floodings events as the elements of a finite probability space. The AAD indicator is then a finite weighted sum of the damages associated with each flooding event. In the second framework, we model the flooding events as a random vector with values in $\mathbb{R}^\kappa$ with $\kappa \geq 1$. The AAD indicator is then defined as a κ-dimensional integral over $\mathbb{R}^\kappa$. Both frameworks are illustrated by a step-by-step example.

Note: In the NOE modelling framework presented in Chapter 2, we used the second framework (flooding events modelled as a random vector with values in $\mathbb{R}^\kappa$).

C.1 Flooding events as the elements of a finite probability space

C.1.1 Flooding events

Definition (Set of flooding events)**.** We represent the set of "*flooding events*" by a finite set Υ of $\mathcal{M}$ elements. Each element $e \in \Upsilon$ is a unique flooding event. We denote by $\mathscr{P}(\Upsilon)$ the set of subsets of Υ.

Hypothesis (Probability space of flooding events)**.** *We assume there exists a probability measure function P defined on $\mathscr{P}(\Upsilon)$ such that for a given year, $P(e)$ is the probability that flooding event $e \in \Upsilon$ occurs this year. The triplet $(\Upsilon, \mathscr{P}(\Upsilon), P)$ is a finite probability space.*

Definition (Vector of descriptors for a flooding event)**.** We assume that a "*flooding event*" can be entirely described by a finite number κ of scalar descriptors (e.g., peak discharge, inflow volume, water level beyond dams, failure or not failure of levees, the season of occurence of the event, etc.). We denote by $\mathbf{X}$ the random vector from Υ to $\mathbb{R}^\kappa$ which associates each flooding event $e \in \Upsilon$ with a vector of descriptors:

$$\begin{aligned} \mathbf{X} : \Upsilon &\to \mathbb{R}^\kappa \\ e &\mapsto X(e) = (X_1(e), \ldots, X_\kappa(e)) \end{aligned}$$

We denote by P_X the law of random vector $\mathbf{X}$: it is a discrete probability distribution because the set Υ of flooding events is finite.

C.1.2 Average annual damages

Definition (Damages associated with a flooding event)**.** We assume that the damages associated with flooding events can be represented by a scalar random variable $\mathbf{D}$ from Υ to $\mathbb{R}$. For a given flooding event $e \in \Upsilon$, $D(e)$ is the amount of damage costs induced by this event:

$$\begin{array}{rccc} \mathbf{D}: & \Upsilon & \to & \mathbb{R} \\ & \mathbf{e} & \mapsto & D(e) \end{array} \tag{$\mathcal{A}$.16}$$

We denote by note P_D the law (discrete distribution function) of random variable $\mathbf{D}$.

Definition (Average Annual Damages)**.** The Average Annual Damages indicator (AAD) is defined as the expected value of random variable $\mathbf{D}$:

$$\text{AAD} = \mathbb{E}(\mathbf{D}) = \sum_{e \in \Upsilon} D(e)P(e) \tag{$\mathcal{A}$.17}$$

C.1.3 Approximation of the AAD indicator

Definition (Partition of the set of flooding events)**.** The random vector $X(e) = (X_1(e), \ldots, X_\kappa(e))$ of flooding event descriptors has its values in $\mathbb{R}^\kappa$. Let consider a partition $\mathcal{I} = (I_1, \ldots, I_m)$ of $\mathbb{R}^\kappa$, such that each subset I_i contains at least the vector of decriptors $X(e)$ of one flooding event $e \in \Upsilon$. We denote by Υ_i the subset (non-empty) of flooding events such that the vector of descriptors $X(e)$ belongs to the subset I_i:

$$\forall i \in [\![1; m]\!], \quad \Upsilon_i = \{e \in \Upsilon \mid X(e) \in I_i\}$$

The set of subsets Υ_i is a partition of the set Υ of flooding events.

Example for $\kappa = 1$: when the random vector $\mathbf{X}$ is of dimension $\kappa = 1$, a usual partition of $\mathbb{R}$ we can consider is a series of intervals $I_i =]a_i; a_{i+1}]$ (with $a_1 = -\infty$ and $a_{m+1} = +\infty$). Subset Υ_i contains all the flooding events e such that $a_i < X(e) \leq a_{i+1}$.

Definition (Average damage of a subset of flooding events)**.** Let consider $i \in [\![1; m]\!]$. We denote by $D_i \in \mathbb{R}$ the average damage associated with the subset $\Upsilon_i \subset \Upsilon$, that is, the conditional expectation $\mathbb{E}[\mathbf{D} \mid e \in \Upsilon_i]$:

$$D_i = \mathbb{E}[D \mid e \in \Upsilon_i] = \frac{\sum\limits_{e \in \Upsilon_i} D(e)P(e)}{\sum\limits_{e \in \Upsilon_i} P(e)} \tag{$\mathcal{A}$.18}$$

Property C.1 (Another expression of the AAD indicator)**.** *The AAD indicator can be written as the sum of the average damages of all the subsets Υ_i weighted by the measure of each subset I_i with respect to P_X:*

$$\mathit{AAD} = \sum_{i=1}^{m} D_i \cdot P_X(I_i) \tag{$\mathcal{A}$.19}$$

Proof. According to Eqn. ($\mathcal{A}$.17), the AAD indicator is defined as: $\text{AAD} = \sum_{e \in \Upsilon} D(e)P(e)$. Let split

this discrete sum onto the subsets Υ_i:

$$\begin{aligned}
\text{AAD} &= \sum_{i=1}^{m} \sum_{e \in \Upsilon_i} D(e)P(e) \\
&= \sum_{i=1}^{m} \frac{\sum\limits_{e \in \Upsilon_i} D(e)P(e)}{\sum\limits_{e \in \Upsilon_i} P(e)} \cdot \sum_{e \in \Upsilon_i} P(e) \\
&= \sum_{i=1}^{m} D_i \cdot \sum_{e \in \Upsilon_i} P(e) \quad \text{[Eqn. (}\mathcal{A}\text{.18)]}
\end{aligned}$$

Besides, the sum $\sum_{e \in \Upsilon_i} P(e)$ is equal to:

$$\begin{aligned}
\sum_{e \in \Upsilon_i} P(e) &= \mathbb{P}\left(e \in \Upsilon_i\right) \\
&= \mathbb{P}\left(X(e) \in I_i\right) \quad \text{(by definition of } I_i\text{)} \\
&= P_X(I_i)
\end{aligned}$$

Hence we finally obtain the desired expression. □

Definition (Proxy for the AAD indicator)**.** Let assume that we have:

- a proxy $\hat{D}_i$ of the average damages D_i for each subset Υ_i: this estimator can for example be chosen as the damage $D(e)$ associated with a representative flooding event $e \in \Upsilon_i$;
- a proxy $\hat{P_X}(I_i)$ of the measure $P_X(I_i)$ for each subset I_i: this estimator can be derived from a sample of realisations of random vector $\mathbf{X}$ over years;

We then define an approximation of the Average Annual Damages, denoted by $\hat{\text{AAD}}$:

$$\hat{\text{AAD}} = \sum_{i=1}^{m} \hat{D}_i \cdot \hat{P_X}(I_i) \tag{$\mathcal{A}$.20}$$

C.1.4 A step-by-step example

We describe here a simple example to illustrate the notions and notations given in this section.

Finite probability space of flooding events We consider a set Υ of $\mathcal{M} = 5$ flooding events, named e_1 to e_5. $\mathscr{P}(\Upsilon)$ is the set of subsets of Υ. We define a probability measure function P on $\mathscr{P}(\Upsilon)$ based on the following elementary probabilities:

$$\begin{aligned}
P(e_1) &= 1/2 \\
P(e_2) &= 1/3 \\
P(e_3) &= 1/12 \\
P(e_4) &= 1/16 \\
P(e_5) &= 1/48
\end{aligned} \tag{$\mathcal{A}$.21}$$

For a given year, $P(e_i)$ is the probability that flooding event e_i occurs this year. The triplet $(\Upsilon, \mathscr{P}(\Upsilon), P)$ is a finite probability space.

Descriptors of flooding events We define two descriptors for each flooding event: its peak discharge (given in m^3/s), and the state of a levee along the river (0: the levee does not fail; 1: the levee fails). The random vector $\mathbf{X}$ which associates each flooding event $e_i \in \Upsilon$ with a vector of descriptors writes:

$$\begin{aligned} \mathbf{X}(e_1) &= (1\,000\,m^3/s, 0) \\ \mathbf{X}(e_2) &= (2\,000\,m^3/s, 0) \\ \mathbf{X}(e_3) &= (1\,000\,m^3/s, 1) \\ \mathbf{X}(e_4) &= (2\,000\,m^3/s, 1) \\ \mathbf{X}(e_5) &= (4\,000\,m^3/s, 1) \end{aligned} \tag{$\mathcal{A}$.22}$$

Damages associated with a flooding event For a given flooding event $e \in \Upsilon$, $D(e)$ is the amount of damage costs induced by this event:

$$\begin{aligned} \mathbf{D}(e_1) &= 0\,€ \\ \mathbf{D}(e_2) &= 150\,€ \\ \mathbf{D}(e_3) &= 3\,000\,€ \\ \mathbf{D}(e_4) &= 5\,000\,€ \\ \mathbf{D}(e_5) &= 8\,000\,€ \end{aligned} \tag{$\mathcal{A}$.23}$$

Average annual damages The AAD indicator is equal to the expected value of random variable $\mathbf{D}$:

$$AAD = \sum_{i=1}^{5} D(e_1) \cdot P(e_1) \tag{$\mathcal{A}$.24}$$

We find $AAD = 0 \times 1/2 + 150 \times 1/3 + 3000 \times 1/12 + 5000 \times 1/16 + 8000 \times 1/48 = 779\,€$.

Partition of the set of flooding events We now divide the range of possible peak discharges into a partition with three intervals:

$$I_1 =]-\infty, 1000] \quad I_2 =]1000, 2000] \quad I_3 =]2000, +\infty[$$

From this partition on the values of peak discharges, we can build the associated subsets Υ_i of flooding events such that the vector of descriptors $X(e)$ belongs to the subset I_i:

$$\begin{aligned} \Upsilon_1 &= \{e_1, e_3\} \\ \Upsilon_2 &= \{e_2, e_4\} \\ \Upsilon_3 &= \{e_5\} \end{aligned} \tag{$\mathcal{A}$.25}$$

Average damage of a subset of flooding events The average damages D_1, D_2, D_3 associated with the subsets Υ_1, Υ_2, Υ_3 are given by:

$$\begin{aligned} D_1 &= \frac{P(e_1)\cdot D(e_1) + P(e_3)\cdot D(e_3)}{P(e_1)+P(e_3)} && = 428 € \\ D_2 &= \frac{P(e_2)\cdot D(e_2) + P(e_4)\cdot D(e_4)}{P(e_1)+P(e_4)} && = 916 € \\ D_3 &= D(e_5) && = 8000 € \end{aligned} \tag{$\mathcal{A}$.26}$$

Another expression of the AAD indicator The AAD indicator can be written as the sum of the average damages of all the subsets Υ_i weighted by the measure of each subset I_i with respect to P_X:

$$\text{AAD} = D_1 \cdot P_X(I_1) + D_2 \cdot P_X(I_2) + D_3 \cdot P_X(I_3) \tag{$\mathcal{A}$.27}$$

The measure of each subset I_i is defined as the probability that the peak discharge q of a flooding event belongs to I_i:

$$\begin{aligned} P_X(I_1) &= P(q \leq 1000\, m^3/s) && = 1/2 + 1/12 = 7/12 \\ P_X(I_2) &= P(1000 < q \leq 2000\, m^3/s) && = 1/3 + 1/16 = 19/48 \\ P_X(I_3) &= P(q > 2000\, m^3/s) && = 1/48 \end{aligned} \tag{$\mathcal{A}$.28}$$

Using Eqn. ($\mathcal{A}$.27), we find $AAD = 428 \times \frac{7}{12} + 916 \times \frac{19}{48} + 8000 \times \frac{1}{48} = 779$ €.

C.2 Flooding events as a random vector

C.2.1 Flooding events

Definition (Set of flooding events)**.** We assume that a "*flooding event*" can be entirely described by a finite number κ of scalar parameters (e.g., peak discharge, inflow volume, water level beyond dams, failure or not failure of levees, the season of occurence of the event, etc.). We then identify each flooding event with a single realisation e of a random vector $\mathbf{E} = (E_1, \ldots, E_\kappa)$ with values in $\mathbb{R}^\kappa$.

Hypothesis (Probability density function of flooding events)**.** *We assume that random vector* $\mathbf{E}$ *has a probability density function* p_E*. For a given year,* $p_E(e)$ *represents the probability that flooding event* $e \in \mathbb{R}^\kappa$ *occurs this year.*

Definition (Marginal density)**.** For each component E_i of random vector $\mathbf{E} = (E_1, \ldots, E_\kappa)$, we define a marginal density $p_i : \mathbb{R} \to \mathbb{R}$ by:

$$\forall e_i \in \mathbb{R}, \quad p_i(e_i) = \int_{\mathbb{R}^{\kappa-1}} p_E(e)\, de_1 \ldots de_{i-1} de_{i+1} \ldots de_\kappa$$

C.2.2 Average Annual Damages

Hypothesis (Damages associated with a flooding event)**.** *We assume that, on a given study area, the monetized damages associated with flooding events can be represented by a continuous function* D *from* $\mathbb{R}^\kappa$ *to* $\mathbb{R}$*. For a given flooding event* $e \in \mathbb{R}^\kappa$*,* $D(e)$ *is the amount of damage costs [€] induced by this event:*

$$\begin{aligned} D : \quad \mathbb{R}^\kappa &\to \mathbb{R} \\ e &\mapsto D(e) \end{aligned} \tag{$\mathcal{A}$.29}$$

We denote by $\mathbf{D}$ *the random variable* $\mathbf{D} = D(\mathbf{E})$*. We denote by note* P_D *the law of random variable* $\mathbf{D}$*.*

Property C.2 (Probability density function of damages)**.** *Random variable* $\mathbf{D}$ *admits a probability density function* p_D.

Definition (Average Annual Damages)**.** The Average Annual Damages indicator (AAD) is defined as the expected value of random variable $\mathbf{D}$ (if it exists):

$$\text{AAD} = \mathbb{E}(\mathbf{D}) = \int_{\mathbb{R}} u\, p_D(u)\, du = \int_{\mathbb{R}^\kappa} D(e)\, p_E(e)\, de \tag{$\mathcal{A}$.30}$$

C.2.3 Aproximation of the AAD indicator

Definition (Conditional expectation of damages)**.** Let consider a component E_i of random vector $\mathbf{E} = (E_1, \ldots, E_\kappa)$ that describe flooding events. We denote by $\mathbb{E}[\mathbf{D} \mid E_i]$ the conditional expectation of damages $\mathbf{D}$ given E_i:

$$\forall e_i \in \mathbb{R}, \quad \mathbb{E}[\mathbf{D} \mid E_i = e_i] = \frac{1}{p_i(e_i)} \int_{\mathbb{R}^{\kappa-1}} D(e) p_E(e)\, de_1 \ldots de_{i-1} de_{i+1} \ldots de_\kappa \tag{$\mathcal{A}$.31}$$

Property C.3 (Expression of the AAD indicator as a unidimensionnal integral)**.** *The AAD indicator can be written as a unidimensionnal integral on random variable* E_i*:*

$$AAD = \int_{\mathbb{R}} \mathbb{E}[\mathbf{D} \mid E_i = e_i] \cdot p_i(e_i) de_i$$

Proof. According to Eqn. ($\mathcal{A}$.30), the AAD indicator is defined as:

$$\text{AAD} = \int_{\mathbb{R}^\kappa} D(e)\, p_E(e)\, de \tag{$\mathcal{A}$.32}$$

Which we can write:

$$\begin{aligned} \text{AAD} &= \int_{\mathbb{R}} \left(\frac{1}{p_i(e_i)} \int_{\mathbb{R}^{\kappa-1}} D(e)\, p_E(e)\, de_1 \ldots de_{i-1} de_{i+1} \ldots de_\kappa \right) p_i(e_i)\, de_i \\ &= \int_{\mathbb{R}} \mathbb{E}[\mathbf{D} \mid E_i] \cdot p_i(e_i)\, de_i \quad \text{(according to Eqn. ($\mathcal{A}$.31))} \end{aligned}$$

□

To approximate the AAD indicator, one must approximate the following components:

- conditional amount of damages: $\mathbb{E}[\mathbf{D} \mid E_i = e_i]$ is the conditional expectation of damages given that the flood descriptor E_i is equal to e_i. This is the most difficult component to approximate;
- conditional density $p_i(e_i)$: this value can be estimated from a serie of random realisations of E_i over years, for example by fitting a parametric law of extreme values;
- the integral: the integral can be approximated using the usual trapezoidal rule, from a small number of data points;

D Appendix to §2.2.8.2: overlay analysis for flood exposure assessment

Dans cette annexe, nous évoquons très rapidement la question des méthodes de croisement entre données « *aléa* » et données « *enjeux* » dans le cadre de l'estimation des dommages dus aux inondations. Cette question a été brièvement abordée durant le travail de thèse ; elle ne constitue cependant pas le cœur de notre travail et n'a donc pas été présentée dans le corps de ce mémoire.

Croisement aléa-enjeux pour l'estimation des dommages dus aux crues Dans le cadre d'une démarche d'estimation des dommages dus aux inondations telle que celle présentée dans le Chapitre 2, une des étapes consiste à caractériser l'exposition des enjeux situés sur la zone d'étude vis-à-vis de différents scénarios de crues (§2.2.8.2 on page 54). Ce calcul d'exposition se fait généralement à l'aide d'un logiciel SIG en superposant les données d'aléa et la carte des enjeux. L'objectif de cette étape est d'affecter à chaque enjeu une hauteur de submersion, une vitesse de submersion et une durée de submersion pour les différents scénarios de crue considérés. Nous appellerons cette opération le « *croisement aléa-enjeux* ».

Diversité des méthodes de croisement aléa-enjeux Du fait de la dimension spatialisée des données « *aléa* » et « *enjeux* » et de la diversité des formats numériques dans lesquels elles sont stockées, l'opération de « *croisement aléa-enjeux* » peut se faire selon des procédures variées. Ces diverses méthodes aboutiront toutes à affecter aux enjeux des valeurs d'aléa pour les différents scénarios de crue considérés, mais des variations parfois importantes peuvent apparaître dans les valeurs d'aléa affectées d'une méthode à l'autre. Ainsi, dans le cadre de l'appui technique apporté par Irstea aux partenaires du Plan Rhône (qui a constitué le cadre de ce travail de thèse), il avait été prévu d'étudier comment les indicateurs économiques produit par le modèle NOE pouvaient varier selon la méthode de « *croisement aléa-enjeux* » choisie.

Cette question a été traitée par Thibaud Langer dans son mémoire de Master (Langer 2011). Un panel d'une dizaine de méthodes de « *croisement aléa-enjeux* » différentes a été proposé (Figure $\mathcal{A}$.3 on the next page). Ces méthodes diffèrent en autres par : i) la nature des données « *aléa* » (ponctuel, polygone ou raster), et ii) la nature des données « *enjeux* » (ponctuel ou polygone). Ces différentes approches de croisement aléa-enjeux ont été appliquées dans le cadre d'une estimation des dommages dus aux inondations sur le terrain d'étude de Fourques-Beaucaire sur le Rhône (projet de renforcement de la digue entre Fourques et Beaucaire). Les estimations des dommages moyens annualisés obtenus avec les différentes méthodes de croisement aléa-enjeux ont ensuite été comparées, faisant apparaître des écarts relatifs pouvant atteindre les 10%.

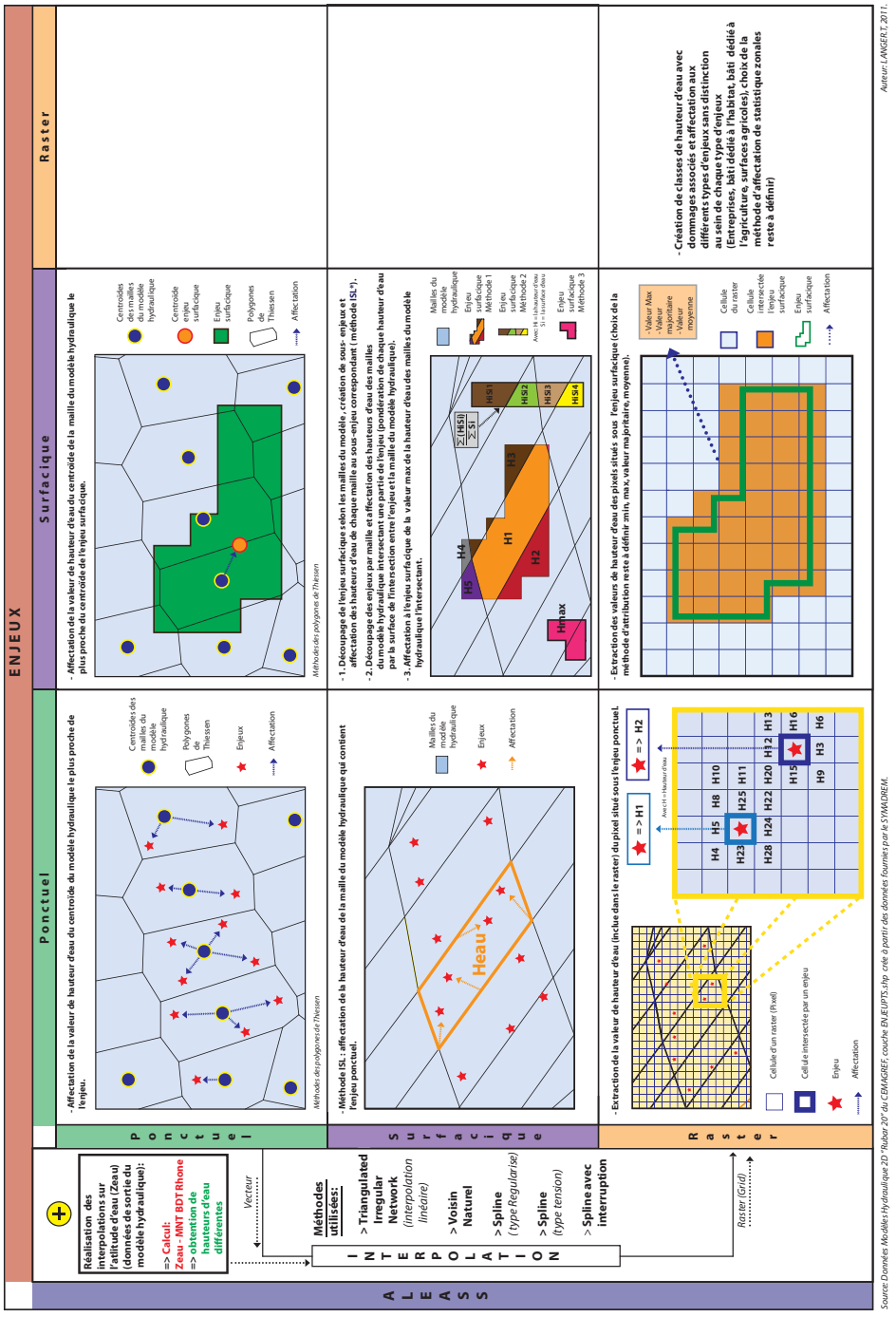

FIGURE A.3 – Méthodes de croisement aléa-enjeux selon le format des données « *aléa* » (en ligne) et le format des données « *enjeux* » (en colonne)
Source : Langer (2011)

Intégrer une incertitude de modèle à l'analyse de sensibilité Le choix d'une méthode de croisement aléa-enjeux peut être considéré comme une « *incertitude de modèle* ». Dans une optique de comparaison globale des sources d'incertitude dans le modèle NOE, il pourrait donc être intéressant de comparer la part de variabilité apportée par l'incertitude sur le choix d'une méthode de croisement aléa-enjeux avec la variabilité due aux incertitudes sur les données d'entrée du modèle. Pour cela, une possibilité est d'intégrer le choix de la méthode de croisement aléa-enjeu au sein de l'analyse de sensibilité du modèle NOE, en représentant ce choix par une variable aléatoire discrète (approche identique au « *map labelling* » décrit en §3.1.2.4 on page 84). L'objectif d'une telle analyse serait de déterminer si l'existence de plusieurs méthodes de croisement aléa-enjeux alternatives est une source d'incertitude importante ou non (relativement aux autres sources d'incertitude).

E Appendix to §2.3.3 : extra case studies

Dans cette annexe, nous présentons très succintement les trois études supplémentaires qui ont été traitées à l'aide du modèle NOE durant cette thèse, en plus de l'étude sur le delta de l'Orb. La Table 2.7 en page 71 résume les principales caractéristiques de ces trois études.

E.1 Plaine d'inondation de la Vilaine à Redon

(Passages extraits de Grelot et al. (2009)).

Le bassin versant de la confluence de la Vilaine et de l'Oust à Redon est d'une surface d'environ 8500 km^2. Il draine une bonne partie de la Bretagne centrale au travers de nombreux affluents (l'Arz, le Don, l'Aff, etc.). Le terrain étudié situé de part et d'autre de cette confluence est d'une surface de 17 km^2 (Figure $\mathcal{A}$.4). Il est composé des deux communes : Redon et Saint-Nicolas de Redon. Ce terrain a subi des inondations « *lentes* » (submersion d'une dizaine de jours) très importantes en 1995, 2000 et 2001. On dénombre sur ce terrain d'étude environ 500 enjeux (habitations, activités économiques et services publics) principalement situés au centre de la zone d'étude.

Cinq scénarios de crue ont été retenus pour calculer les dommages moyens annualisés : trois crues historiques (1995, 2000 et 2001) et deux crues de projet suivant des modèles pluie-débit appliqués à l'ensemble du bassin versant de la Vilaine. Ces deux crues dénommées GR2 et NO3 représentent pour GR2 une crue forte (période de retour de cinquante ans environ) et pour NO3, une crue dite « *à noyaux* », de fréquence de retour plus faible pour l'aval du bassin (Redon), de l'ordre de dix ans. Les simulations hydrauliques ont été réalisées par EGIS-Eau à l'aide du modèle Infoworks en mode filaire.

Les données sur les enjeux ont été collectées par un travail de terrain. Les types d'enjeux recensés sont : les habitations (528 entités), les activités économiques (113 entités) et les services publics (15 entités). Sur ce terrain les enjeux agricoles ont été considérés comme négligeables. Les enjeux ont été numérisés sur la base du cadastre pour créer une couche SIG vecteur (polygones). Une enquête exhaustive de terrain sur les hauteurs de seuil par rapport au terrain naturel des habitations a été également réalisée.

Il n'y a pas de station hydrologique située sur une entrée amont unique du terrain d'étude qui, au contraire, contient deux entrées principales (Oust et Vilaine). Le contexte hydrologique du terrain d'étude est donc particulièrement complexe (apports multiples, confluences, ouvrages, influuence de la marée) et les périodes de retour estimées pour les cinq crues modélisées en hydraulique ne peuvent que s'appuyer sur des données (pluie, débits, niveaux) à l'échelle du bassin versant.

Les courbes d'endommagement utilisées dans cette étude sont les mêmes que celles utilisées sur l'Orb pour le même sous-type d'enjeu. Elles s'appuient sur la typologie des enjeux proposée par Penning-Rowsell et al. (2005).

E.2 Projet de renforcement de la digue au Rhône entre Fourques et Beaucaire

(Passages extraits de Langer (2011)).

Ce terrain d'étude correspond à une zone de 125 km^2 située en Région Languedoc Roussillon dans le Département du Gard. Cette zone est localisée sur la rive droite (rive Ouest) du Rhône sur quatre communes : Bellegarde, Beaucaire, Fourques et Saint-Gilles situées à l'Ouest de la ville d'Arles (Figure $\mathcal{A}$.5).

Dix huit mille personnes habitent sur ce territoire dont douze mille cinq cent dans les zones PHEC (Plus Hautes Eaux Connues, ici indice de crue 1856 et 2003). Ce terrain d'étude a été choisi car une étude ACB a déjà été menée sur ce territoire dans le cadre d'un marché public mené entre le Syndicat Mixte Interrégional d'Aménagement des Digues du Delta du Rhône et de la Mer (SYMADREM, ici maître d'ouvrage) et le bureau d'étude ISL (ISL 2011). L'étude réalisée fait partie des actions de prévention du Plan Rhône et a pour objectif le confortement de la digue située en rive droite du Rhône allant de Beaucaire à Fourques. Les principaux enjeux de ce projet de confortement de digue sont la réduction de l'ensemble des dommages liés aux inondations de la plaine au travers de la mise hors d'eau de milliers d'habitants de la zone concernée ainsi que la mise en sécurité de certaines infrastructures importantes. En effet, le terrain d'étude situé sur les communes de Fourques, Beaucaire, Bellegarde et Saint-Gilles est un territoire vulnérable aux inondations qui a été atteint très fortement par le passé par des crues (1856, 2003). Dans le cadre du plan Rhône, aux vues des enjeux situés sur ce territoire, il a été décidé de mettre en place un nouveau tronçon résistant à la surverse, un renforcement du linéaire préexistant ainsi qu'un recul de certaines des digues.

E.3 Schéma d'optimisation des zones d'expansion des crues (ZEC) du Rhône entre Viviers et Beaucaire

(Passages extraits de Gilbert and Ledoux (2012)).

Le schéma d'optimisation des zones d'expansion des crues (ZEC) du Rhône entre Viviers et Beaucaire est un vaste projet de gestion du risque d'inondation, mené par le Plan Rhône, qui repose sur le constat suivant : l'inondation de certaines ZEC pour les crues moyennes n'est pas utile, dans la mesure où les volumes d'écrêtement disponibles sont utilisés trop tôt, et sont « *gaspillés* » avant l'arrivée de la pointe des crues majeures les plus dommageables. Le schéma d'optimisation des ZEC proposé consiste donc : d'une part à augmenter le niveau de protection de certaines ZEC, pour les solliciter moins fréquemment, mais de façon plus efficace pour les crues les plus dommageables ; d'autre part à remobiliser certaines plaines (Piolenc-Mornas) et îles (La Motte et l'Oiselet à Avignon) qui ont été soustraites aux inondations par les aménagements de la Compagnie Nationale du Rhône et qui pourraient apporter une contribution significative aux écrêtements.

Entre Viviers et Beaucaire, il mobilise ainsi les territoires suivants : Donzère Mondragon, Piolenc Mornas, Codolet l'Ardoise, Caderousse, Sauveterre, Iles de la Motte et de l'Oiselet, Ile de la Barthelasse, Monfrin Aramon, Boulbon Vallabrègues, Beaucaire (Figure $\mathcal{A}$.6).

Les modélisation hydrauliques ont été réalisées par la Compagnie Nationale du Rhône (CNR).

FIGURE A.4 – Plaine d'inondation de la Vilaine à Redon

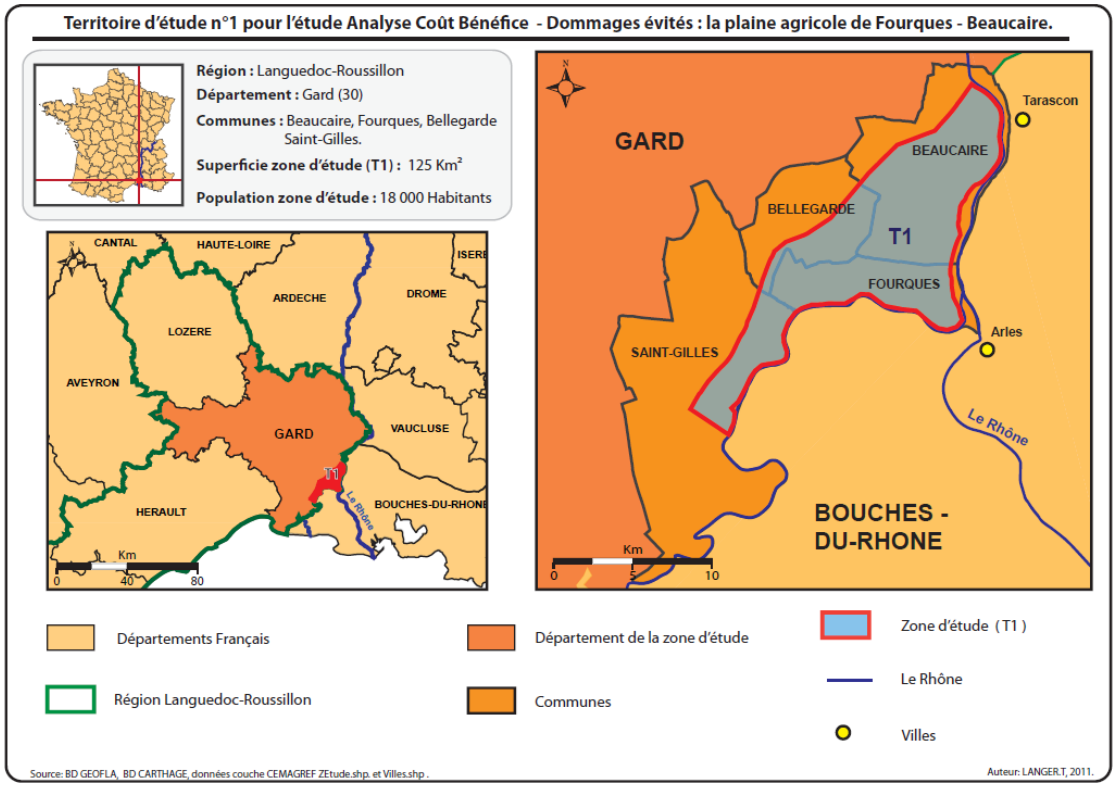

FIGURE A.5 – Le Rhône entre Fourques et Beaucaire

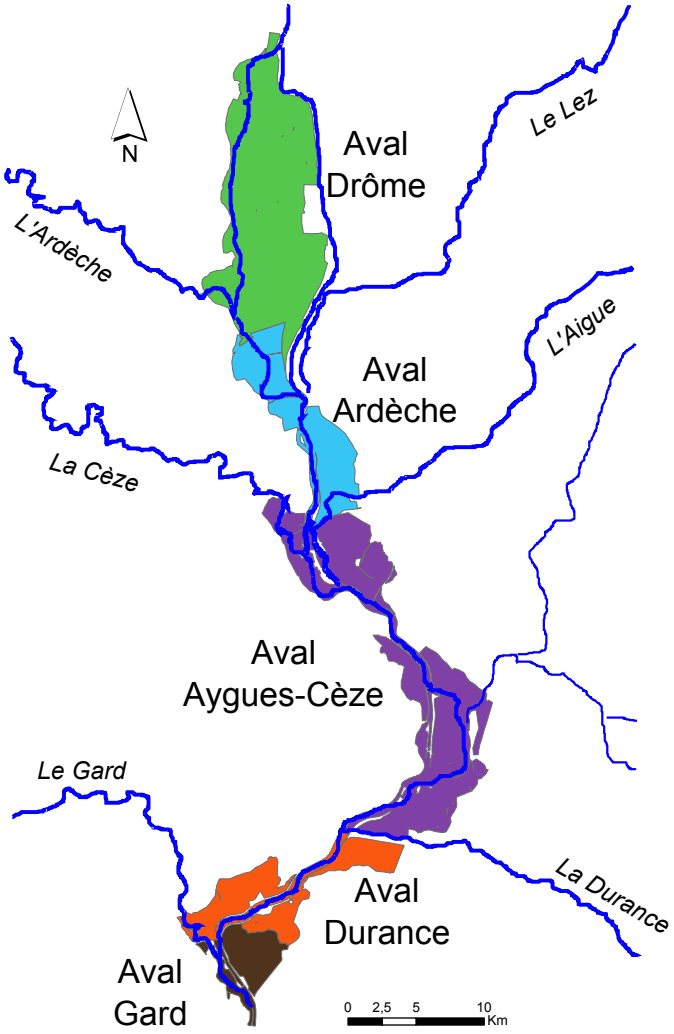

FIGURE $\mathcal{A}$.6 – Zone d'étude du projet ZEC

F Appendix to §3.1 : details on the numerical study

We give in this appendix some details on the numerical study presented in §3.1.2 on page 79. Please refer to this section for a general description of the goals of this study and a discussion of its results.

F.1 Analytical test cases

As mentioned in §3.1 on page 79, we studied three different numerical models, denoted by $\mathcal{F}_1$ to $\mathcal{F}_3$. They all have two independent scalar inputs U_1 and U_2, a single spatially distributed input Z and a scalar output Y_v. U_1 and U_2 are independent scalar inputs with a different uniform pdf for each test case. $\{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}$ is a 2D random field generated by a deterministic function denoted by $\phi_{\text{camp.}}$ (see §3.1.3.1 on page 88 for a detailed description of this function). The output of interest for all models $\mathcal{F}_1$ to $\mathcal{F}_3$ is the spatial average Y_v of a spatially distributed output of some local code $\mathcal{F}_{1,\text{loc}}$ to $\mathcal{F}_{3,\text{loc}}$ over a disk $v \subset \Omega$ of radius $r = 10$:

$$Y_v = \frac{1}{|v|} \int_{\mathbf{x} \in v} \mathcal{F}_{i,\text{loc}}\left[U_1, U_2, Z(\mathbf{x})\right] d\mathbf{x} \tag{$\mathcal{A}$.33}$$

We now give for each test case the expression of local code $\mathcal{F}_{i,\text{loc}}$, and the pdf of U_1 and U_2.

Analytical test case $\mathcal{F}_1$ Local code $\mathcal{F}_{1,\text{loc}}$ is the usual Ishigami function with parameters $A = 7$ and $B = 0.1$ (Homma and Saltelli 1996):

$$\mathcal{F}_{1,\text{loc}}\left[U_1, U_2, z\right] = \sin(z) + A \cdot \sin(U_1)^2 + B \cdot U_2^4 \cdot \sin(z) \tag{$\mathcal{A}$.34}$$

Scalar inputs U_1 and U_2 are independent and both follow the same uniform pdf $\mathcal{U}[-\pi, \pi]$. Spatial domain Ω is a disk of radius $r = 10$.

Analytical test case $\mathcal{F}_2$ Local code $\mathcal{F}_{2,\text{loc}}$ is the usual Sobol G function with parameters $a_0 = 0$, $a_1 = 1$ and $a_2 = 9$ (Archer et al. 1997):

$$\mathcal{F}_{2,\text{loc}}\left[U_1, U_2, z\right] = \frac{|4z - 2| + a_0}{1 + a_0} \cdot \frac{|4U_1 - 2| + a_1}{1 + a_1} \cdot \frac{|4U_2 - 2| + a_2}{1 + a_2} \tag{$\mathcal{A}$.35}$$

Scalar inputs U_1 and U_2 are independent and follow uniform pdf $U_1 \sim \mathcal{U}[-5, 5]$ and $U_2 \sim \mathcal{U}[-1, 5]$, respectively.

Analytical test case $\mathcal{F}_3$ Local code $\mathcal{F}_{3,\text{loc}}$ is a simple linear function of $Z(\mathbf{x})$:

$$\mathcal{F}_{3,\text{loc}}\left[U_1, U_2, z\right] = U_1 \cdot z + U_2 \tag{$\mathcal{A}$.36}$$

Scalar inputs U_1 and U_2 are independent and follow uniform pdf $U_1 \sim \mathcal{U}[0.5, 1.5]$ and $U_2 \sim \mathcal{U}[-8, 8]$, respectively.

F.2 Graphical results

The main results of this study are presented and discussed in §3.1.3.3 on page 90. Here we give some extra graphical outputs of the analysis. For each analytical test case $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$, we display:

- the empirical distribution of model output Y_v over $N_{\text{tot}} = (K + 2) \cdot N = 10\,240$ model runs, and sensitivity indices estimates obtained with the dimension reduction approach (with grouping) (Figure $\mathcal{A}$.7 on the next page, Figure $\mathcal{A}$.8 on page 222 and Figure $\mathcal{A}$.9 on page 223);
- scatterplots of model inputs U_1, U_2 and random label $\mathbf{L}$ against model output Y_v (Figure $\mathcal{A}$.10 on page 224, Figure $\mathcal{A}$.11 on page 224 and Figure $\mathcal{A}$.12 on page 224);
- plots of the contribution to the sample mean (Bolado-Lavin et al. 2009) of model inputs U_1, U_2 and random label $\mathbf{L}$ against model output Y_v (Figure $\mathcal{A}$.13 on page 225, Figure $\mathcal{A}$.14 on page 225 and Figure $\mathcal{A}$.15 on page 225).

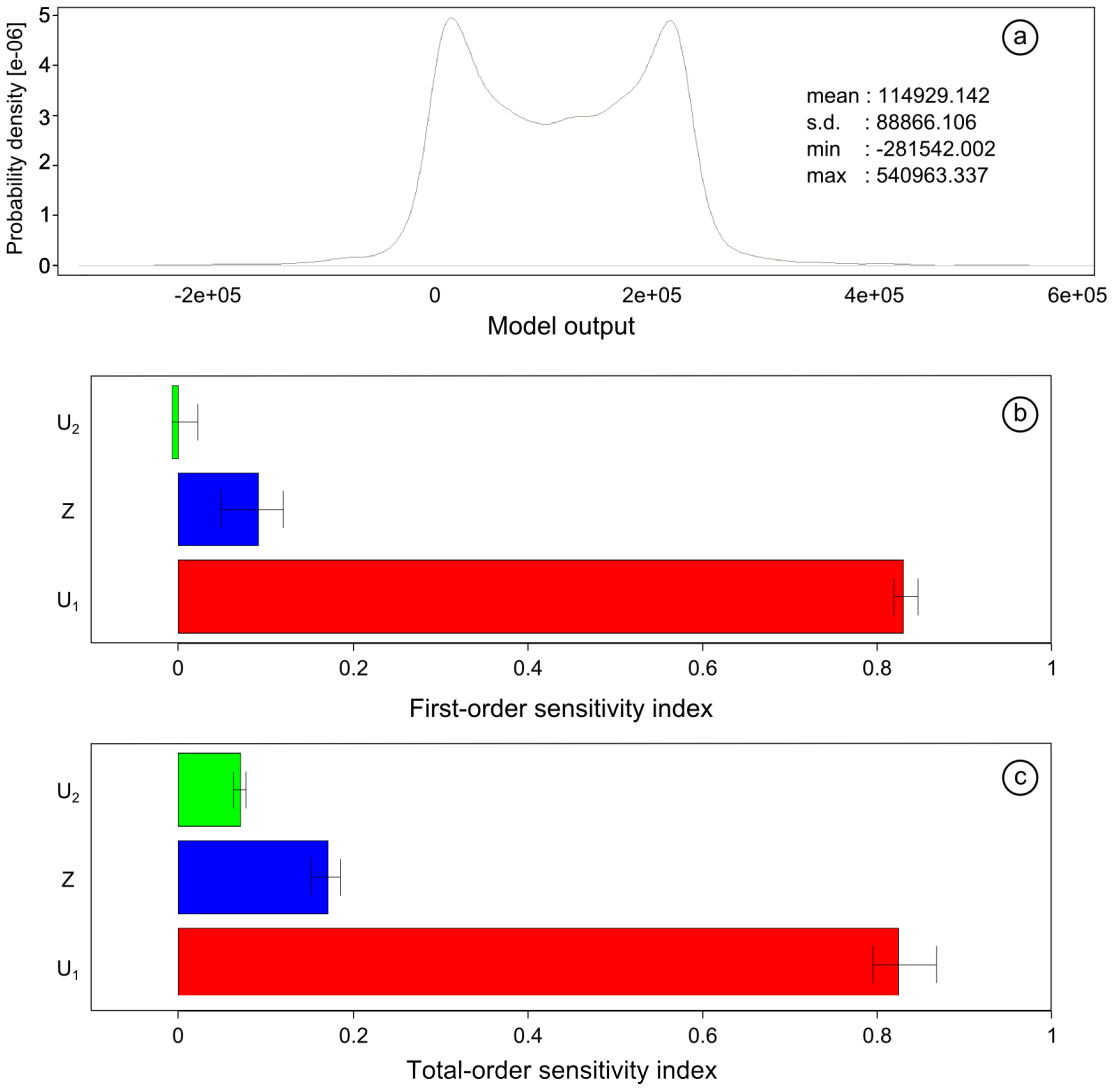

Figure $\mathcal{A}.7$: Test case $\mathcal{F}_1$. (a) empirical distribution of model output Y_v; (b) first-order sensitivity indices estimates; (c) total-order sensitivity indices estimates. 95% confidence interval computed by bootstrap (100 replicas)

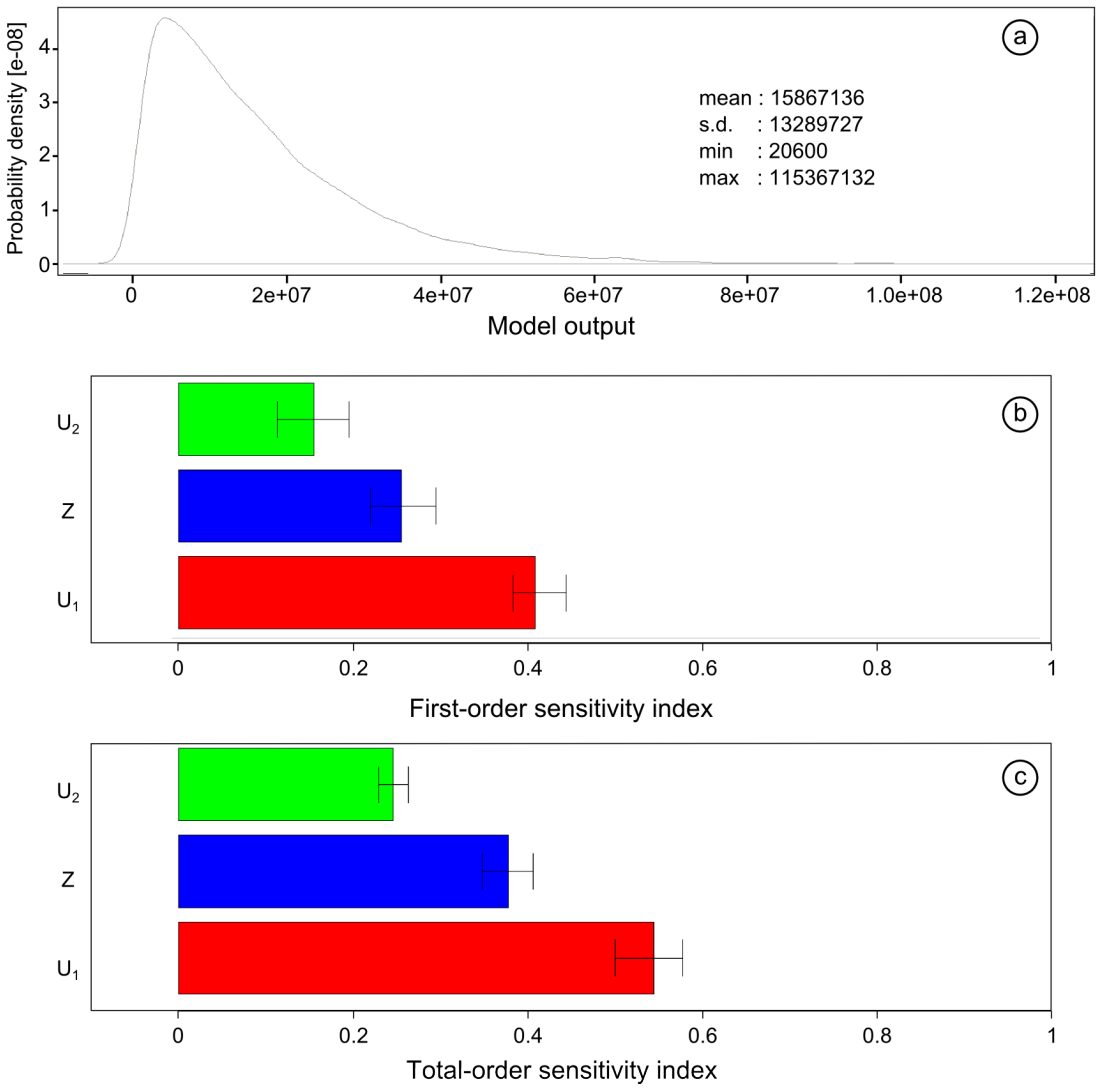

Figure $\mathcal{A}$.8: Test case $\mathcal{F}_2$. (a) empirical distribution of model output Y_v; (b) first-order sensitivity indices estimates; (c) total-order sensitivity indices estimates. 95% confidence interval computed by bootstrap (100 replicas)

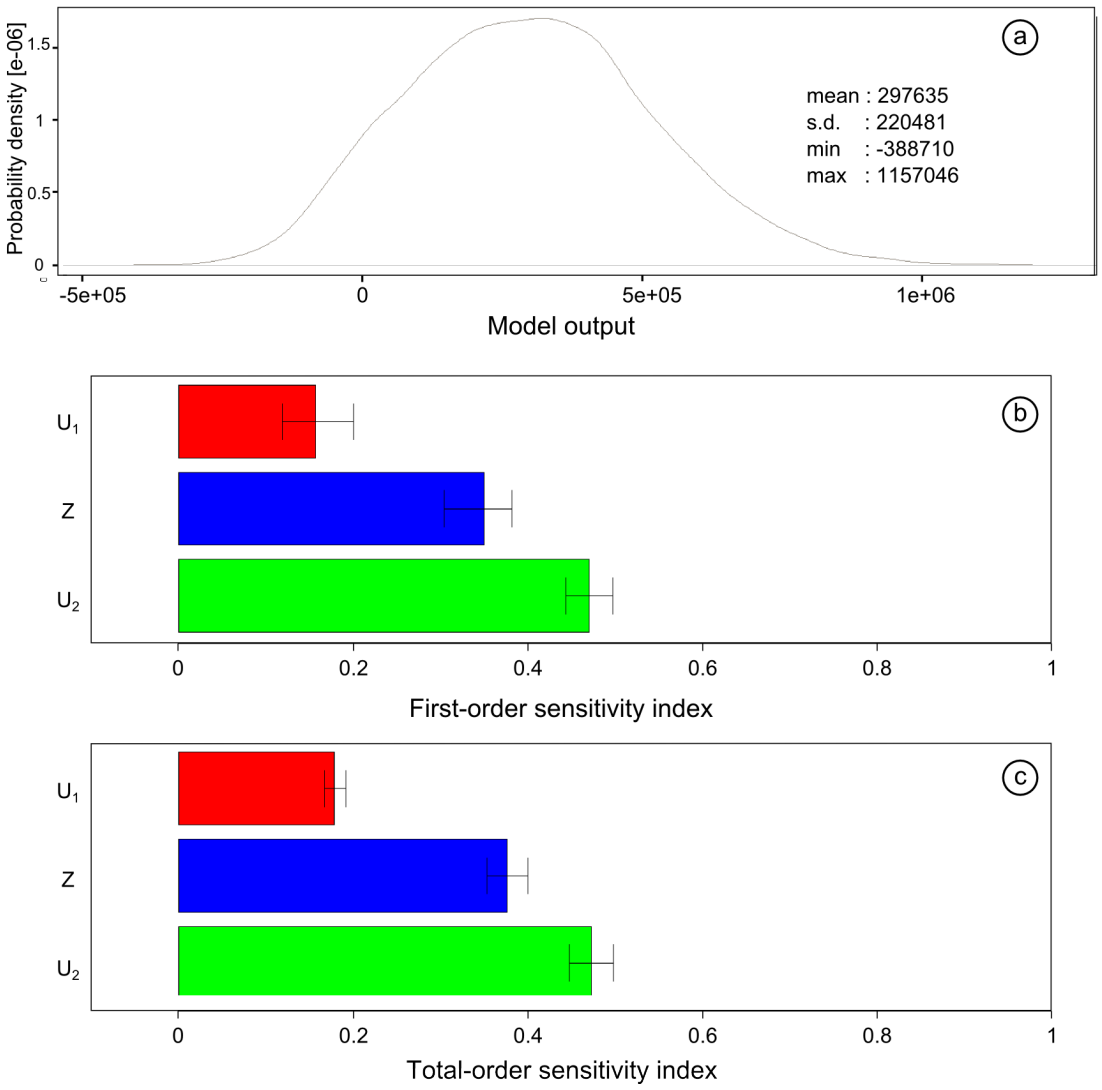

Figure $\mathcal{A}$.9: Test case $\mathcal{F}_3$. (a) empirical distribution of model output Y_v; (b) first-order sensitivity indices estimates; (c) total-order sensitivity indices estimates. 95% confidence interval computed by bootstrap (100 replicas)

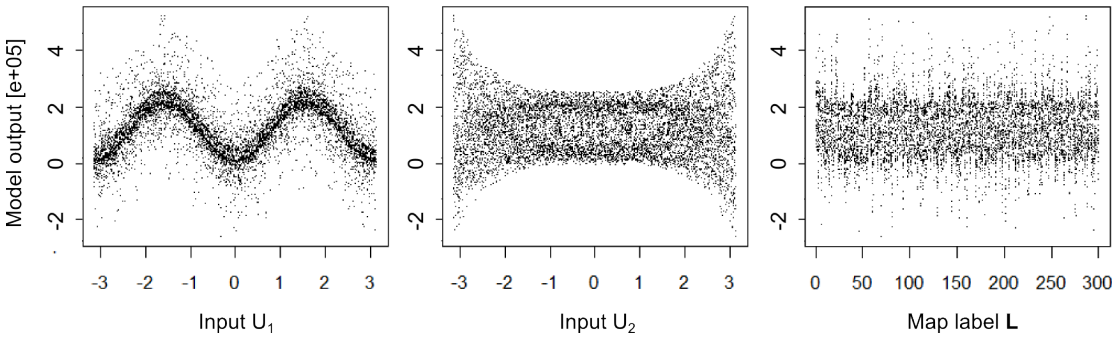

Figure $\mathcal{A}$.10: Test case $\mathcal{F}_1$: scatterplots of U_1, U_2 and $\mathbf{L}$ (random label) against model output Y_v

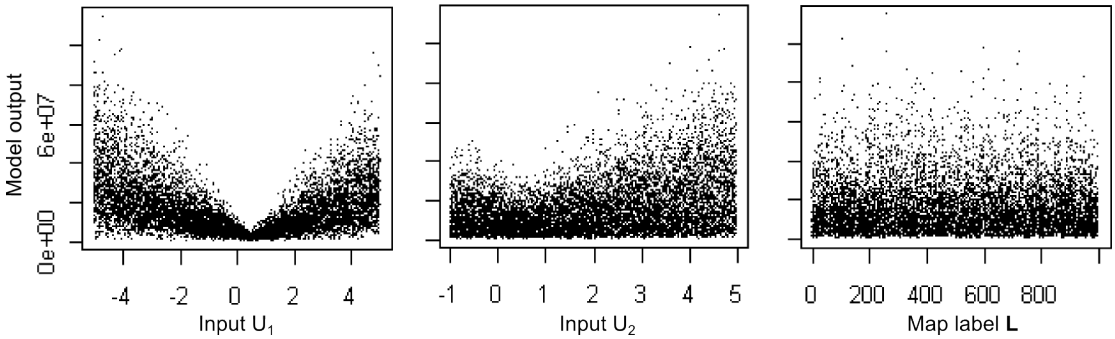

Figure $\mathcal{A}$.11: Test case $\mathcal{F}_2$: scatterplots of U_1, U_2 and $\mathbf{L}$ (random label) against model output Y_v

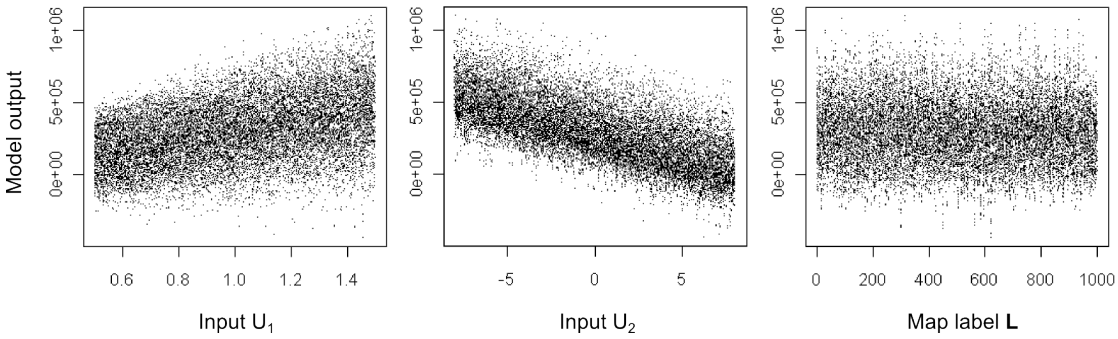

Figure $\mathcal{A}$.12: Test case $\mathcal{F}_3$: scatterplots of U_1, U_2 and $\mathbf{L}$ (random label) against model output Y_v

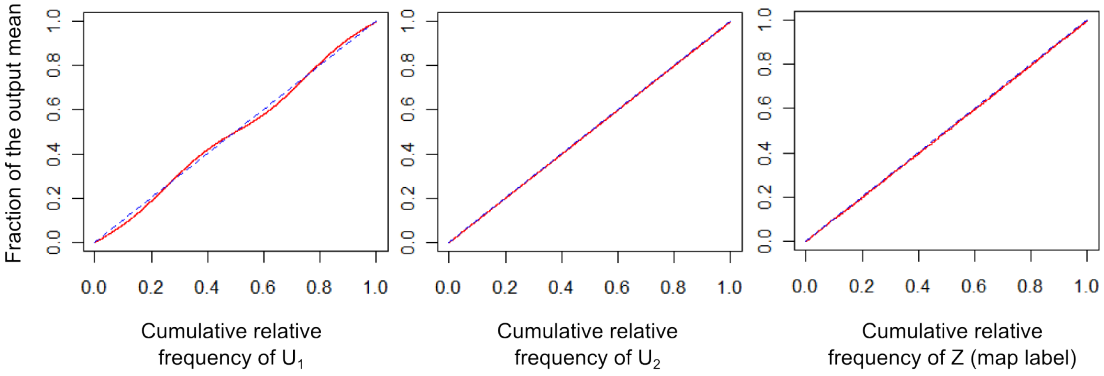

Figure $\mathcal{A}$.13: Test case $\mathcal{F}_1$: plots of contribution to the sample mean of U_1, U_2 and $\mathbf{L}$ (random label) with respect to the model output Y_v

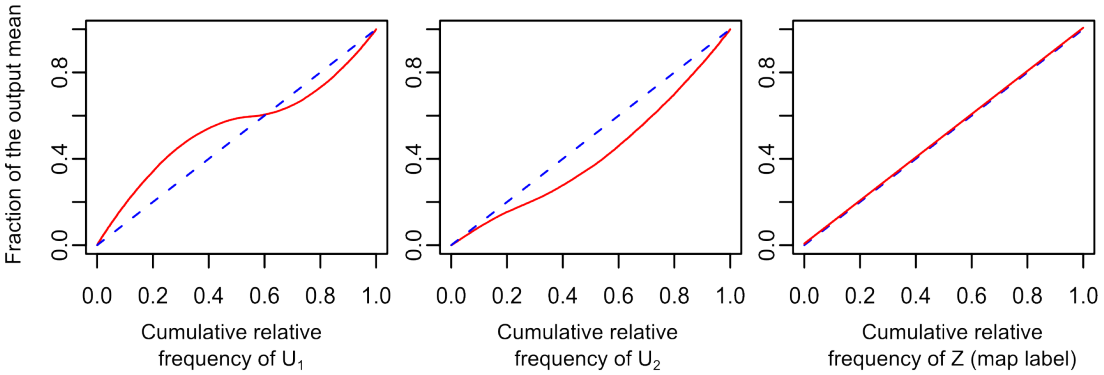

Figure $\mathcal{A}$.14: Test case $\mathcal{F}_2$: plots of contribution to the sample mean of U_1, U_2 and $\mathbf{L}$ (random label) with respect to the model output Y_v

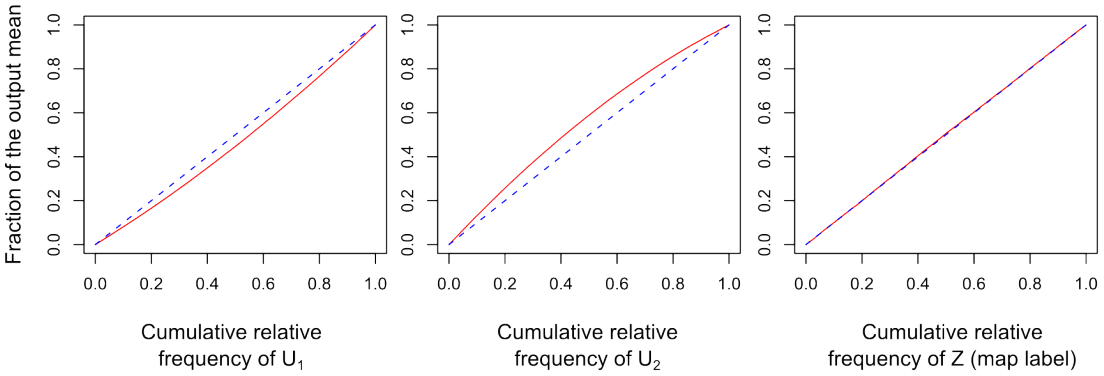

Figure $\mathcal{A}$.15: Test case $\mathcal{F}_3$: plots of contribution to the sample mean of U_1, U_2 and $\mathbf{L}$ (random label) with respect to the model output Y_v

G Appendix to §3.3: sensitivity indices with respect to the NPV indicator

On se propose dans cette annexe de donner l'expression exacte des indices de sensibilité relatifs à la valeur actuelle nette d'un aménagement de protection contre les crues (Net Present Value, NPV), lorsque cet indicateur est calculé selon l'équation Eqn. (2.5) on page 44 donnée dans le Chapitre 2 :

$$NPV = -CI + \sum_{i=0}^{R} (\Delta AAD - CM) \cdot \tau_i \tag{A.37}$$

G.1 Écriture générale

On commence par considérer le cas général d'une fonction ayant pour forme :

$$Y = \sum_{a \in \{0,1\}^K} \zeta_a \cdot \left(\prod_{i=1}^{K} U_i^{a_i}\right) \tag{A.38}$$

où les coefficient ζ_a sont des réels, et où les variables d'entrée U_i sont supposées être des variables aléatoires réelles indépendantes, d'espérance μ_i et de variance σ_i. On veut déterminer l'expression des indices de sensibilité de premier ordre et totaux de chacune des variables d'entrée U_i.

La variance de la sortie Y du modèle s'écrit :

$$\mathrm{var}(Y) = \sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K}} \zeta_a \zeta_b \cdot \mathrm{cov}\left(\prod_{i=1}^{K} U_i^{a_i}, \prod_{i=1}^{K} U_i^{b_i}\right) \tag{A.39}$$

qui se développe de la sorte :

$$\mathrm{var}(Y) = \sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K}} \zeta_a \zeta_b \cdot \left(\prod_{\substack{j=1 \\ a_j + b_j = 1}}^{K} \mu_j \right) \cdot \mathrm{var}\left(\prod_{\substack{j=1 \\ a_j = b_j = 1}}^{K} U_j \right) \tag{A.40}$$

d'où :

$$\mathrm{var}(Y) = \sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K}} \zeta_a \zeta_b \cdot \left(\prod_{\substack{j=1 \\ a_j + b_j = 1}}^{K} \mu_j \right) \cdot \left[\prod_{\substack{j=1 \\ a_j = b_j = 1}}^{K} (\sigma_j^2 + \mu_j^2) - \prod_{\substack{j=1 \\ a_j = b_j = 1}}^{K} \mu_j^2 \right] \tag{A.41}$$

L'espérance conditionnelle de la sortie Y sachant U_i s'écrit :

$$E(Y \mid U_i) = \sum_{a \in \{0,1\}^K} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} \mu_j^{a_j} \right) \cdot U_i^{a_i} \tag{A.42}$$

soit :

$$E(Y \mid U_i) = \left[\sum_{\substack{a \in \{0,1\}^K \\ a_i = 0}} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} \mu_j^{a_j} \right) \right] + \left[\sum_{\substack{a \in \{0,1\}^K \\ a_i = 1}} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} \mu_j^{a_j} \right) \right] \cdot U_i \tag{A.43}$$

donc :

$$\mathrm{var}\left[E(Y \mid U_i)\right] = \left[\sum_{\substack{a \in \{0,1\}^K \\ a_i = 1}} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} \mu_j^{a_j} \right) \right]^2 \cdot \sigma_i^2 \tag{A.44}$$

que l'on peut finalement écrire sous la forme :

$$\mathrm{var}\left[E(Y \mid U_i)\right] = \sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K \\ a_i = b_i = 1}} \zeta_a \zeta_b \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} \mu_j^{a_j + b_j} \right) \cdot \sigma_i^2 \tag{A.45}$$

Enfin, pour déterminer l'expression des indices de sensibilité totaux il nous faut calculer :

$$Y \mid U_{\sim i} = \left[\sum_{\substack{a \in \{0,1\}^K \\ a_i = 0}} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} U_j^{a_j} \right) \right] + \left[\sum_{\substack{a \in \{0,1\}^K \\ a_i = 1}} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} U_j^{a_j} \right) \right] \cdot U_i \tag{A.46}$$

donc :

$$\mathrm{var}(Y \mid U_{\sim i}) = \left[\sum_{\substack{a \in \{0,1\}^K \\ a_i = 1}} \zeta_a \cdot \left(\prod_{\substack{j=1 \\ j \neq i}}^{K} U_j^{a_j} \right) \right]^2 \cdot \sigma_i^2 \tag{A.47}$$

soit :

$$\mathrm{var}(Y \mid U_{\sim i}) = \left[\sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K \\ a_i = b_i = 1}} \zeta_a \zeta_b \cdot \prod_{\substack{j=1 \\ j \neq i}}^{K} U_j^{a_j + b_j} \right] \cdot \sigma_i^2 \tag{A.48}$$

d'où :

$$E\left[\mathrm{var}(Y \mid U_{\sim i})\right] = \left[\sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K \\ a_i = b_i = 1}} \zeta_a \zeta_b \cdot \prod_{\substack{j=1 \\ j \neq i}}^{K} E\left(U_j^{a_j + b_j}\right) \right] \cdot \sigma_i^2 \tag{A.49}$$

que l'on peut écrire :

$$E\left[\mathrm{var}(Y \mid U_{\sim i})\right] = \left[\sum_{\substack{a \in \{0,1\}^K \\ b \in \{0,1\}^K \\ a_i = b_i = 1}} \zeta_a \zeta_b \cdot \prod_{\substack{j=1 \\ j \neq i \\ a_j + b_j = 1}}^{K} \mu_j \cdot \prod_{\substack{j=1 \\ j \neq i \\ a_j = b_j = 1}}^{K} (\sigma_j^2 + \mu_j^2) \right] \cdot \sigma_i^2 \tag{A.50}$$

G.2 Application à l'indicateur NPV

On considère le modèle suivant :

$$NPV = -CI + (\Delta AAD - CM) \cdot \tau^* \tag{$\mathcal{A}$.51}$$

qui est dérivé de Eqn. ($\mathcal{A}$.37) on page 226 définissant l'indicateur NPV en posant $\tau^* = \sum_{i=0}^{R} \tau_i$. Les variables CI, ΔAAD, CM et τ^* sont supposées être des variables aléatoires réelles indépendantes.

On reconnaît la forme générale donnée dans Eqn. ($\mathcal{A}$.38) on page 226 avec :

$$\begin{cases} U_1 = CI \\ U_2 = \tau^* \\ U_3 = \Delta AAD \\ U_4 = CM \end{cases} \tag{$\mathcal{A}$.52}$$

et (seuls les coefficients ζ_a non nuls sont donnés) :

$$\begin{cases} \zeta_{1,0,0,0} = -1 \\ \zeta_{0,1,1,0} = 1 \\ \zeta_{0,1,0,1} = -1 \end{cases} \tag{$\mathcal{A}$.53}$$

D'après Eqn. ($\mathcal{A}$.41) on page 226, la variance totale de l'indicateur NPV s'écrit :

$$\begin{aligned} \text{var}(NPV) =& \sigma_{CI}^2 + \sigma_{\tau^*}^2 \cdot (\mu_{\Delta AAD} - \mu_{CM})^2 + \sigma_{\Delta AAD}^2 \cdot \mu_{\tau^*}^2 \\ &+ \sigma_{CM}^2 \cdot \mu_{\tau^*}^2 + \sigma_{\Delta AAD}^2 \cdot \sigma_{\tau^*}^2 + \sigma_{CM}^2 \sigma_{\tau^*}^2 \end{aligned}$$

Les indices de sensibilité de premier ordre s'obtiennent à partir de Eqn. ($\mathcal{A}$.45) on the preceding page :

$$\begin{cases} S_{CI} = \dfrac{\sigma_{CI}^2}{\text{var}(NPV)} \\ S_{\tau^*} = \dfrac{\sigma_{\tau^*}^2 \cdot (\mu_{\Delta AAD} - \mu_{CM})^2}{\text{var}(NPV)} \\ S_{\Delta AAD} = \dfrac{\sigma_{\Delta AAD}^2 \cdot \mu_{\tau^*}^2}{\text{var}(NPV)} \\ S_{CM} = \dfrac{\sigma_{CM}^2 \cdot \mu_{\tau^*}^2}{\text{var}(NPV)} \end{cases}$$

Les indices de sensibilité totaux s'obtiennent à partir de Eqn. ($\mathcal{A}$.50) on the previous page :

$$\begin{cases} ST_{CI} = \dfrac{\sigma_{CI}^2}{\text{var}(NPV)} \\ ST_{\tau^*} = \dfrac{\sigma_{\tau^*}^2 \cdot \left[\sigma_{\Delta AAD}^2 + \sigma_{CM}^2 + (\mu_{\Delta AAD} - \mu_{CM})^2\right]}{\text{var}(NPV)} \\ ST_{\Delta AAD} = \dfrac{\sigma_{\Delta AAD}^2 \cdot \left(\mu_{\tau^*}^2 + \sigma_{\tau^*}^2\right)}{\text{var}(NPV)} \\ ST_{CM} = \dfrac{\sigma_{CM}^2 \cdot \left(\mu_{\tau^*}^2 + \sigma_{\tau^*}^2\right)}{\text{var}(NPV)} \end{cases}$$

Et l'on peut aussi identifier les indices de second ordre (ceux non indiqués sont nuls) :

$$\begin{cases} S_{\Delta AAD,\tau^*} = \dfrac{\sigma^2_{\Delta AAD}\sigma^2_{\tau^*}}{\text{var}(NPV)} \\ S_{CM,\tau^*} = \dfrac{\sigma^2_{CM}\sigma^2_{\tau^*}}{\text{var}(NPV)} \end{cases}$$

On remarque que les valeurs des indices de sensibilité ne dépendent que des espérances et variances respectives des différents facteurs d'entrée, et pas de la forme exacte de leur densité de probabilité.

H Appendix to §4.1 : proofs

In this appendix, we give a number of detailed properties and proofs related to the first section §4.1 on page 140 of Chapter 4. We assume that function $\mathcal{F}_{\text{loc}}$ is such that the following random variables have finite expectation and finite variance, for any point $\mathbf{x} \in \Omega$: $Y(\mathbf{x})$, $\mathbb{E}[Y(\mathbf{x}) \mid Z(\mathbf{x})]$ and $\mathbb{E}[Y(\mathbf{x}) \mid U]$.

Definition ($\mathbb{E}_Z Y(\mathbf{x})$ random field)**.** $\mathbb{E}_Z Y(\mathbf{x})$ is defined as the conditionnal expectation of $Y(\mathbf{x})$ given $Z(\mathbf{x})$:

$$\forall \mathbf{x} \in \Omega, \quad \mathbb{E}_Z Y(\mathbf{x}) = \mathbb{E}[Y(\mathbf{x}) \mid Z(\mathbf{x})]$$

Property H.1. $\mathbb{E}_Z Y(\mathbf{x})$ *random field is a transformation of* $Z(\mathbf{x})$*:*

$$\boxed{\forall \mathbf{x} \in \Omega, \quad \mathbb{E}_Z Y(\mathbf{x}) = \bar{\mathcal{F}}_{loc}[Z(\mathbf{x})]}$$

where $\bar{\mathcal{F}}_{loc} : \mathbb{R} \to \mathbb{R}$ *is defined by* $\bar{\mathcal{F}}_{loc}(z) = \int_{u \in \mathbb{R}^k} \mathcal{F}_{loc}(u, z) \cdot p_U(u)\, du$*.*

Proof of property H.1. Let define function $\bar{\mathcal{F}}_{\text{loc}} : \mathbb{R} \to \mathbb{R}$ by:

$$\bar{\mathcal{F}}_{\text{loc}}(z) = \int_{u \in \mathbb{R}^k} \mathcal{F}_{\text{loc}}(u, z) \cdot p_U(u)\, du$$

($\bar{\mathcal{F}}_{\text{loc}}$ exists thanks to our hypotheses on function $\mathcal{F}_{\text{loc}}$). Now consider a point $\mathbf{x} \in \Omega$. We have:

$$\begin{aligned}
\bar{\mathcal{F}}_{\text{loc}}[Z(\mathbf{x})] &= \int_{u \in \mathbb{R}^k} \mathcal{F}_{\text{loc}}[u, Z(\mathbf{x})] \cdot p(u)\, du && \text{(definition of } \bar{\mathcal{F}}_{\text{loc}}\text{)} \\
&= \mathbb{E}(\mathcal{F}_{\text{loc}}[U, Z(\mathbf{x})] \mid Z(\mathbf{x})) \\
&= \mathbb{E}[Y(\mathbf{x}) \mid Z(\mathbf{x})] && \text{[Eqn. (4.2)] on page 142} \\
&= \mathbb{E}_Z Y(\mathbf{x}) && \text{by definition}
\end{aligned}$$

□

Property H.2 (Hermite orthonormal basis of $L^2(\mathcal{N})$)**.**
Consider the sequence of Hermite polynomials $(\chi_j)_{j \in \mathbb{N}}$ *defined by:*

$$\forall j \in \mathbb{N}, \quad \chi_j(z) = \frac{1}{\sqrt{j!}} \cdot \frac{1}{n(z)} \cdot \frac{\partial^j}{\partial z^j} n(z)$$

in which $n(\cdot)$ *is the normal pdf of mean 0 and variance 1. The sequence* $(\chi_j)_{j \in \mathbb{N}}$ *forms an orthonormal basis of Hilbert space* $L^2(\mathcal{N})$ *defined by:*

$$L^2(\mathcal{N}) = \left\{ f : \mathbb{R} \to \mathbb{R} \quad \textit{such that} \int_{-\infty}^{\infty} f^2(z) \cdot n(z)\, dz < \infty \right\}$$

Proof of property H.2. See Journel and Huijbregts (1978), §VI.B.3 for details on Hermite polynomials.
□

Property H.3 (Hermite expansion of $\bar{\mathcal{F}}_{\text{loc}}$)**.** *Function $\bar{\mathcal{F}}_{loc}$ can be expanded on the basis of Hermite polynomials:*

$$\boxed{\bar{\mathcal{F}}_{loc} = \sum_{j=0}^{\infty} \pi_j \cdot \chi_j \quad \text{with } \pi_j = \int_{-\infty}^{+\infty} \chi_j(z) \cdot \bar{\mathcal{F}}_{loc}(z) \cdot n(z)\, dz}$$

Proof of property H.3. We assumed that random variable $\mathbb{E}[Y(\mathbf{x}) \mid Z(\mathbf{x})]$ has finite expectation and variance, thus function $\bar{\mathcal{F}}_{\text{loc}}$ has the following property:

$$\int_{-\infty}^{\infty} \bar{\mathcal{F}}^2_{\text{loc}}(z) \cdot n(z)\, dz < \infty$$

Thus function $\bar{\mathcal{F}}_{\text{loc}}$ is an element of Hilbert space $L(\mathcal{N})$. As a consequence, $\bar{\mathcal{F}}_{\text{loc}}$ can be expressed as a linear combination of the vectors χ_j of the Hermitian basis. The expression of coefficients π_i of this expansion is given in Journel and Huijbregts (1978), §VI.B.3, Eqn. (VI.23). □

Property H.4 (Hermite expansion of $\mathbb{E}_Z Y$)**.**
The following properties hold:

- $\mathbb{E}_Z Y(\mathbf{x})$ *random field is order 2 stationary*
- *it can be expanded as follows:*

$$\boxed{\forall \mathbf{x} \in \Omega, \quad \mathbb{E}_Z Y(\mathbf{x}) = \sum_{k=0}^{\infty} \pi_k \cdot \chi_k[Z(\mathbf{x})] \qquad (a)}$$

- *its mean is* π_0
- *its covariance function $C^*(h)$ is given by:*

$$\boxed{\forall h \geq 0, \quad C^*(h) = \sum_{j=0}^{\infty} \pi_j^2 \cdot \frac{C^j(h)}{C^j(0)} \qquad (b)}$$

Proof of property H.4. Consider a point $\mathbf{x} \in \Omega$. According to property H.1 on the preceding page, we have:

$$\mathbb{E}_Z Y(\mathbf{x}) = \bar{\mathcal{F}}_{\text{loc}}[Z(\mathbf{x})]$$

Using the Hermite expansion of $\bar{\mathcal{F}}_{\text{loc}}$ given in property H.3, we can write:

$$\forall \mathbf{x} \in \Omega, \quad \mathbb{E}_Z Y(\mathbf{x}) = \sum_{j=0}^{\infty} \pi_j \cdot \chi_j[Z(\mathbf{x})]$$

$\mathbb{E}_Z Y(\mathbf{x})$ is assumed to have a finite expectation value. We can write:

$$\begin{aligned} \forall \mathbf{x} \in \Omega, \quad \mathbb{E}[\mathbb{E}_Z Y(\mathbf{x})] &= \mathbb{E}\left(\sum_{j=0}^{\infty} \pi_j \cdot \chi_j[Z(\mathbf{x})]\right) \\ &= \sum_{j=0}^{\infty} \pi_j \cdot \mathbb{E}(\chi_j[Z(\mathbf{x})]) \end{aligned}$$

Besides, Hermite polynomials have the following properties (see Journel and Huijbregts (1978), §VI.B.3, between Eqn. (VI.21) and Eqn. (VI.22)):

$$\int_{-\infty}^{+\infty} \chi_i(z) \cdot \chi_j(z) \cdot n(z)\, dz = \delta_{ij}$$

Thus, using $\chi_0(z) = 1$:

$$\forall \mathbf{x} \in \Omega, \quad \mathbb{E}\left(\chi_j\left[Z(\mathbf{x})\right]\right) = \int_{-\infty}^{+\infty} \chi_0(z) \cdot \chi_j(z) \cdot n(z)\, dz = \delta_{0j}$$

So:

$$\forall \mathbf{x} \in \Omega, \quad \mathbb{E}\left[\mathbb{E}_Z Y(\mathbf{x})\right] = \sum_{k=0}^{\infty} \pi_k \cdot \delta_{0j} = \pi_0$$

The expression of the covariance function C^* is taken from Allard (2010), or Chilès and Delfiner (1999 p.396, Eqn. (6.23) and p.399, Eqn. (6.25)): Finally we have:

$$\forall \mathbf{x} \in \Omega, \quad \begin{cases} \mathbb{E}\left[\mathbb{E}_Z Y(\mathbf{x})\right] = \pi_0 \\ \operatorname{cov}\left[\mathbb{E}_Z Y(\mathbf{x}),\, \mathbb{E}_Z Y(\mathbf{x}+\mathbf{h})\right] = C^*(h) \end{cases}$$

These two properties are the definition of order 2 stationarity of $\mathbb{E}_Z Y(\mathbf{x})$ random field. □

Property H.5 (Expression of block variance V_v^*)**.**
Consider a block $v \subset \Omega$. Let denote by $\overline{C^*}(v,v)$ the mean value of the covariance $C^*(h)$ when the two extremities of the distance vector h describe the domain v:

$$\overline{C^*}(v,v) = \frac{1}{|v|^2} \iint_{(\mathbf{x},\mathbf{x}')\in v^2} C^*(\mathbf{x}-\mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}'$$

Block variance V_v^* is equal to:

$$\boxed{V_v^* = \overline{C^*}(v,v)}$$

Proof of property H.5. Consider a block $v \subset \Omega$. Block variance V_v^* is given by:

$$\begin{aligned}
V_v^* &= \operatorname{var}\left[\frac{1}{|v|}\int_{\mathbf{x}\in v} \mathbb{E}_Z Y(\mathbf{x})\, d\mathbf{x}\right] && \text{(by definition, see §4.1.3.1 on page 144)} \\
&= \frac{1}{|v|^2} \iint_{(\mathbf{x},\mathbf{x}')\in v^2} \operatorname{cov}\left[\mathbb{E}_Z Y(\mathbf{x}), \mathbb{E}_Z Y(\mathbf{x}')\right]\, d\mathbf{x}\, d\mathbf{x}' \\
&= \frac{1}{|v|^2} \iint_{(\mathbf{x},\mathbf{x}')\in v^2} C^*(\mathbf{x}-\mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}' && \text{(property H.4 on page 231)} \\
&= \overline{C^*}(v,v) && \text{(definition of } \overline{C^*}(v,v)\text{)}
\end{aligned}$$

□

Property H.6 (Change of support)**.** *Consider $v \subset \Omega$ and $\mathcal{V} \subset \Omega$ two blocks such that $v \ll \mathcal{V}$. Let denote by $D^2(v \mid \mathcal{V})$ the dispersion variance of v within $\mathcal{V}$ for the $\mathbb{E}_Z Y(\mathbf{x})$ random field:*

$$D^2(v \mid \mathcal{V}) = \overline{C^*}(v, v) - \overline{C^*}(\mathcal{V}, \mathcal{V})$$

The ratio of block sensitivity indices of U and Z verifies:

$$\frac{S_Z(\mathcal{V})}{S_{\mathbf{U}}(\mathcal{V})} = \frac{S_Z(v)}{S_{\mathbf{U}}(v)} - \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{D^2(v \mid \mathcal{V})}{V^*}$$

thus:

$$\boxed{\frac{S_Z(\mathcal{V})}{S_{\mathbf{U}}(\mathcal{V})} \leq \frac{S_Z(v)}{S_{\mathbf{U}}(v)}}$$

Proof of property H.6. Consider $v \subset \Omega$ and $\mathcal{V} \subset \Omega$ two blocks such that $v \ll \mathcal{V}$. Let denote by $D^2(v \mid \mathcal{V})$ the dispersion variance of v within $\mathcal{V}$ for $\mathbb{E}_Z Y(\mathbf{x})$ random field. We have:

$$D^2(v \mid \mathcal{V}) = \overline{C^*}(v, v) - \overline{C^*}(\mathcal{V}, \mathcal{V})$$

(see Journel and Huijbregts (1978) §II.C.2 Eqn. (II.35)). Using this expression of $D^2(v \mid \mathcal{V})$ with the equality $V^*_{\mathcal{V}} = \overline{C^*}(\mathcal{V}, \mathcal{V})$ (property H.5 on the preceding page) and the expression of ratio $S_Z(\mathcal{V})/S_{\mathbf{U}}(\mathcal{V})$ [Eqn. (4.8) on page 145], we obtain:

$$\frac{S_Z(\mathcal{V})}{S_{\mathbf{U}}(\mathcal{V})} = \frac{S_Z(v)}{S_{\mathbf{U}}(v)} - \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{D^2(v \mid \mathcal{V})}{V^*}$$

We know that dispersion variance $D^2(v \mid \mathcal{V})$ is positive (Journel and Huijbregts (1978), Eqn. (II.33)), thus:

$$\frac{S_Z(\mathcal{V})}{S_{\mathbf{U}}(\mathcal{V})} \leq \frac{S_Z(v)}{S_{\mathbf{U}}(v)}$$

□

Property H.7 (Influence of covariance range)**.** *Consider a block $v \subset \Omega$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property (a denotes the covariance range):*

$$\forall h \geq 0, \quad \frac{\partial C}{\partial a}(h) \geq 0$$

Then the ratio of block sensitivity indices of $\mathbf{U}$ and Z verifies:

$$\boxed{\frac{\partial}{\partial a}\left[\frac{S_Z(v)}{S_{\mathbf{U}}(v)}\right] \geq 0}$$

Proof of property H.7. Consider a block $v \subset \Omega$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ is differentiable with respect to parameter a (range) and has the following property:

$$\forall h \geq 0, \quad \frac{\partial C}{\partial a}(h) \geq 0$$

Then, using the expression of covariance function C^* (property H.4 on page 231), we find that covariance function C^* is also differentiable with respect to a and:

$$\forall h \geq 0, \quad \frac{\partial C^*}{\partial a}(h) \geq 0$$

v^2 is a finite domain, thus using the expression of $\overline{C^*}(v,v)$ (property H.5 on page 232), we obtain that $\overline{C^*}(v,v)$ is differentiable with respect to a and that:

$$\frac{\partial \overline{C^*}(v,v)}{\partial a} \geq 0$$

Using the equality $V_v^* = \overline{C^*}(v,v)$ (property H.5 on page 232), we show that block variance V_v^* is differentiable with respect to a and that:

$$\frac{\partial V_v^*}{\partial a} \geq 0$$

Finally, using the expression of ratio $S_Z(v)/S_{\mathbf{U}}(v)$ given in Eqn. (4.8) on page 145, we obtain the expected property. □

Property H.8 (Influence of nugget parameter)**.** *Consider a block $v \subset \Omega$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property (η denotes the covariance nugget):*

$$\forall h \geq 0, \quad \frac{\partial C}{\partial \eta}(h) \leq 0$$

Then the ratio of block sensitivity indices of $\mathbf{U}$ and Z verifies:

$$\boxed{\frac{\partial}{\partial \eta}\left[\frac{S_Z(v)}{S_{\mathbf{U}}(v)}\right] \leq 0}$$

Proof of property H.8. The proof is the same as the proof of property H.7 on the previous page. □

Property H.9 (Case of a pure-nugget random field)**.** *Consider a block $v \subset \Omega$. Let assume that $Z(\mathbf{x})$ is a random field without spatial auto-correlation, that is, that its covariance function $C(\mathbf{h})$ is null except for $\mathbf{h} = 0$:*

$$\forall \mathbf{h} \neq \mathbf{0}, \quad C(\mathbf{h}) = 0$$

Then the ratio of block sensitivity indices of $\mathbf{U}$ and Z verifies:

$$\boxed{\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = 0}$$

Proof of property H.9. If the covariance function $C(\cdot)$ of $Z(\mathbf{x})$ is null except for $\mathbf{h} = 0$, then the covariance function $C^*(\cdot)$ of $\mathbb{E}_Z Y(\mathbf{x})$ has the same property [H.4 on page 231 b)]. Then the average value $\overline{C^*}(v,v)$ of the covariance $C^*(h)$ over domain v is equal to zero. Finally, using the expression of ratio $S_Z(v)/S_{\mathbf{U}}(v)$ given in Eqn. (4.8) on page 145, we obtain the expected property. □

Property H.10 (Case of a totally correlated random field)**.** *Consider a block $v \subset \Omega$. Let assume that the covariance function $C(\cdot)$ of random field $Z(\mathbf{x})$ is constant:*

$$\forall \mathbf{h} \in \Omega, \quad C(\mathbf{h}) = C(0)$$

Then the ratio of block sensitivity indices of $\mathbf{U}$ and Z is equal to the ratio of site sensitivity indices:

$$\boxed{\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{S_Z}{S_{\mathbf{U}}}}$$

Proof of property H.10 on the preceding page. If the covariance function $C(\cdot)$ of $Z(\mathbf{x})$ is constant, then the covariance function $C^*(\cdot)$ of $\mathbb{E}_Z Y(\mathbf{x})$ has the same property [H.4 on page 231 b)]:

$$\forall \mathbf{h} \in \Omega, \quad C^*(\mathbf{h}) = C^*(0) = \sum_{j=0}^{\infty} \pi_j^2 \tag{A.54}$$

Then the average value $\overline{C^*}(v, v)$ of the covariance $C^*(\cdot)$ over domain v is simply equal to $C^*(0)$. Finally, using the expression of ratio $S_Z(v)/S_{\mathbf{U}}(v)$ given in Eqn. (4.8) on page 145, we obtain the expected property. □

Property H.11 (Change of support effect - case of an homothety)**.** *Consider a block $v \subset \Omega$. Let denote by $v' \subset \Omega$ the block obtained from v by a homothety of center $O \in \Omega$ and ratio $\tau > 1$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property:*

$$\forall h \geq 0, \quad \frac{\partial C}{\partial h}(h) \leq 0$$

Then block sensitivity indices verify:

$$\boxed{\frac{S_Z(v')}{S_{\mathbf{U}}(v')} < \frac{S_Z(v)}{S_{\mathbf{U}}(v)}}$$

Proof of property H.11. Consider a block $v \subset \Omega$. Let denote by $v' \subset \Omega$ the block obtained from v by a homothety of center $O \in \Omega$ and ratio $\tau > 1$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field is differentiable and has the following property:

$$\forall h \geq 0, \quad \frac{\partial C}{\partial h}(h) \leq 0$$

Using the expression of covariance function $C^*(h)$ (property H.4 on page 231) and the hypothesis made on $C(h)$, we deduce that $C^*(h)$ is differentiable with respect to h and that:

$$\forall h > 0, \quad \frac{\partial C^*}{\partial h}(h) \leq 0$$

Block variance $V_{v'}^*$ writes:

$$\begin{aligned} V_{v'}^* &= \operatorname{var}\left[\frac{1}{v'} \int_{\mathbf{x} \in v'} \mathbb{E}_Z Y(\mathbf{x})\, d\mathbf{x}\right] \\ &= \frac{1}{v'^2} \iint_{(\mathbf{x},\mathbf{x}') \in v'^2} \operatorname{Cov}\left[\mathbb{E}_Z Y(\mathbf{x}), \mathbb{E}_Z Y(\mathbf{x}')\right] d\mathbf{x}\, d\mathbf{x}' \\ &= \frac{1}{v'^2} \iint_{(\mathbf{x},\mathbf{x}') \in v'^2} C^*(|\mathbf{x} - \mathbf{x}'|)\, d\mathbf{x}\, d\mathbf{x}' \quad \text{property H.5 on page 232} \end{aligned}$$

We note $\mathbf{x} = \tau \mathbf{y}$ and $\mathbf{x}' = \tau \mathbf{y}'$, and we use $v' = \tau^2 \cdot v$ and $(\mathbf{x} - \mathbf{x}') = \tau \cdot (\mathbf{y} - \mathbf{y}')$ to obtain:

$$V_{v'}^* = \frac{1}{v^2} \iint_{(\mathbf{y},\mathbf{y}') \in v^2} C^*(\tau \cdot |\mathbf{y} - \mathbf{y}'|)\, d\mathbf{y}\, d\mathbf{y}'$$

We have $\tau > 1$ and $\partial C^* / \partial h \leq 0$ (hypothesis), thus $C^*(\tau \cdot |\mathbf{y} - \mathbf{y}'|) \leq C^*(|\mathbf{y} - \mathbf{y}'|)$ so:

$$V^*_{v'} \leq V^*_v$$

Using the expression of block sensitivity indices [Eqn. (4.8) on page 145], we finally find the expected property.

□

I Appendix to §4.1: case of a linear model

In this appendix, we display a detailed study of the spatial model $\mathcal{F}$ presented in Chapter 4 (§4.1.2.1 on page 141), for the specific case in which the following conditions are verified:

- function $\mathcal{F}_{\text{loc}}$ defined in §4.1.2.1 on page 141 is linear:

$$\forall \mathbf{x} \in \Omega, \quad Y(\mathbf{x}) = \mathcal{F}_{\text{loc}}[\mathbf{U}, Z(\mathbf{x})] = U_1 \cdot Z(\mathbf{x}) + U_2 \qquad (a) \tag{$\mathcal{A}$.55}$$

- U_1 and U_2 are independant random variables with normal distributions:

$$U_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad \text{and} \quad U_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \qquad (b)$$

Please refer to Chapter 4 for the definitions of all variables.

I.1 Properties of $\mathbb{E}_Z Y(\mathbf{x})$ random field

Definition . Consider a block $v \subset \Omega$. We will denote by Z_v the mean value of $Z(\mathbf{x})$ over v:

$$Z_v = \frac{1}{|v|} \int_{\mathbf{x} \in v} Z(\mathbf{x})\, d\mathbf{x}$$

We will denote by V_v the block variance of $Z(\mathbf{x})$ random field over v:

$$V_v = \text{var}[Z_v]$$

Property I.1. *The following properties hold:*

$$\begin{aligned} \mathbb{E}_Z Y(\mathbf{x}) &= \mu_1 \cdot Z(\mathbf{x}) + \mu_2 && (a) \\ \mathbb{E}_Z Y_v &= \mu_1 \cdot Z_v + \mu_2 && (b) \\ V^* &= \mu_1^2 \cdot V && (c) \\ V_v^* &= \mu_1^2 \cdot V_v && (d) \end{aligned}$$

Proof of property I.1. These equations are derived from the definition of model $\mathcal{F}$ and from the definitions of random field $\mathbb{E}_Z Y(\mathbf{x})$ [Eqn. (4.7) on page 144]. □

I.2 Expression of sensitivity indices

Property I.2 (Expression of sensitivity indices for a linear model)**.**
Point-based sensitivity indices are given by:

$$\boxed{S_Z(\mathbf{x}) = \frac{\mu_1^2 V}{\varsigma_1 \cdot V + \varsigma_2} \qquad S_{\mathbf{U}}(\mathbf{x}) = \frac{\varsigma_2}{\varsigma_1 \cdot V + \varsigma_2} \qquad S_{\mathbf{U},Z}(\mathbf{x}) = \frac{\sigma_1^2 \cdot V}{\varsigma_1 \cdot V + \varsigma_2}}$$

and block sensitivity indices are given by:

$$\boxed{S_Z(v) = \frac{\mu_1^2 V_v}{\varsigma_1 \cdot V_v + \varsigma_2} \qquad S_{\mathbf{U}}(v) = \frac{\varsigma_2}{\varsigma_1 \cdot V_v + \varsigma_2} \qquad S_{\mathbf{U},Z}(v) = \frac{\sigma_1^2 \cdot V_v}{\varsigma_1 \cdot V_v + \varsigma_2}}$$

with:

$$\begin{cases} \varsigma_1 = \sigma_1^2 + \mu_1^2 & \geq 0 \\ \varsigma_2 = \sigma_1^2 \cdot \mu_Z^2 + \sigma_2^2 & \geq 0 \end{cases}$$

Proof of property I.2 on the previous page. Consider a point $\mathbf{x} \in \Omega$. Variance of $Y(\mathbf{x})$ writes:

$$\begin{aligned} \text{var}[Y(\mathbf{x})] &= \text{var}(U_1 \cdot Z(\mathbf{x}) + U_2) \\ &= \sigma_1^2\sigma_Z^2 + \mu_1^2\sigma_Z^2 + \sigma_1^2\mu_Z^2 + \sigma_2^2 \\ &= \varsigma_1 \cdot \sigma_Z^2 + \varsigma_2 \end{aligned}$$

with:

$$\begin{cases} \varsigma_1 = \sigma_1^2 + \mu_1^2 & \geq 0 \\ \varsigma_2 = \sigma_1^2 \cdot \mu_Z^2 + \sigma_2^2 & \geq 0 \end{cases}$$

Besides, conditional expectation of $Y(\mathbf{x})$ given $\mathbf{U}$ writes:

$$\begin{aligned} \mathbb{E}(Y(\mathbf{x}) \mid \mathbf{U}) &= \mathbb{E}(U_1 \cdot Z(\mathbf{x}) + U_2 \mid \mathbf{U}) \\ &= U_1 \cdot \mu_Z + U_2 \end{aligned}$$

thus $\text{var}\left[\mathbb{E}(Y(\mathbf{x}) \mid \mathbf{U})\right] = \sigma_1^2 \cdot \mu_Z^2 + \sigma_2^2 = \varsigma_2$. Using the expression of $\text{var}\left[Y(\mathbf{x}^*)\right]$, property I.1 on the preceding page (b) and the expression of point-based sensitivity indices [Eqn. (4.5) on page 143], we find the expected expression for point-based sensitivity indices for a linear model.

Using the definition of model $\mathcal{F}$ and the definition of Y_v (§4.1.2.1 on page 141), we have $Y_v = U_1 \cdot Z_v + U_2$. Thus, block variance $\text{var}\left[Y_v\right]$ writes:

$$\begin{aligned} \text{var}\left[Y_v\right] &= \text{var}(U_1 Z_v + U_2) \\ &= \sigma_1^2 V_v + \mu_1^2 V_v + \sigma_1^2 \mu_Z^2 + \sigma_2^2 \\ &= \varsigma_1 \cdot V_v + \varsigma_2 \end{aligned}$$

Using property I.1 on the previous page (d) and the definition of block sensitivity indices [Eqn. (4.5) on page 143], we find the expected expression for block sensitivity indices for a linear model. □

I.3 Influence of covariance range a

Property I.3 (Influence of covariance range)**.** *Consider a block $v \subset \Omega$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property (a denotes the covariance range):*

$$\forall h \geq 0, \quad \frac{\partial C}{\partial a}(h) \geq 0$$

Then, block sensitivity indices of $\mathbf{U}$ and Z verify:

$$\boxed{\frac{\partial S_Z(v)}{\partial a} > 0 \quad \frac{\partial S_{\mathbf{U}}(v)}{\partial a} < 0 \quad \frac{\partial S_{\mathbf{U},Z}(v)}{\partial a} > 0}$$

Property I.4 (Limits when $a \to +\infty$)**.** *Consider a block $v \ll \Omega$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property:*

$$\forall h > 0, \quad \begin{cases} \lim\limits_{a \to +\infty} C(h) = (1 - \eta) \cdot V \\ \lim\limits_{a \to 0} C(h) = 0 \end{cases}$$

Then, the limits of block sensitivity indices are given by:

$$\boxed{\begin{array}{ll} \lim\limits_{a\to+\infty} S_Z(v) = \dfrac{\mu_1^2 \cdot V}{\varsigma_1 V + \frac{\varsigma_2}{1-\eta}} & \lim\limits_{a\to 0} S_Z(v) = 0 \\ \lim\limits_{a\to+\infty} S_{\mathbf{U}}(v) = \dfrac{\frac{\varsigma_2}{1-\eta}}{\varsigma_1 V + \frac{\varsigma_2}{1-\eta}} & \lim\limits_{a\to 0} S_{\mathbf{U}}(v) = 1 \\ \lim\limits_{a\to+\infty} S_{\mathbf{U},Z}(v) = \dfrac{\sigma_1^2 \cdot V}{\varsigma_1 V + \frac{\varsigma_2}{1-\eta}} & \lim\limits_{a\to 0} S_{\mathbf{U},Z}(v) = 0 \end{array}}$$

Proof of properties I.3 to I.4 on the preceding page. Consider a block $v \subset \Omega$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field is differentiable with respect to parameter a (range) and has the following property:

$$\forall h \geq 0, \quad \frac{\partial C}{\partial a}(h) \geq 0$$

Let denote by $\overline{C}(v,v)$ the mean value of covariance $C(h)$ when the two extremities of distance vector h describe the block v. We know that block variance V_v can be written (same proof as property H.5 on page 232):

$$V_v = \overline{C}(v,v)$$

Using this expression and the hypothesis on $C(h)$, we can conclude that V_v is differentiable with respect to parameter a and that:

$$\frac{\partial V_v}{\partial a} \geq 0$$

Using the expression of block sensitivity indices given in property I.2 on page 237, we obtain that block sensitivity indices are differentiable with respect to a and that:

$$\begin{cases} \dfrac{\partial S_Z(\mathcal{V})}{\partial a} = \dfrac{\varsigma_2 \cdot \mu_1^2}{(\varsigma_1 \cdot V_v + \varsigma_2)^2} \cdot \dfrac{\partial V_v}{\partial a} & \geq 0 \\ \dfrac{\partial S_{\mathbf{U}}(\mathcal{V})}{\partial a} = \dfrac{-\varsigma_1 \cdot \varsigma_2}{(\varsigma_1 \cdot V_v + \varsigma_2)^2} \cdot \dfrac{\partial V_v}{\partial a} & \leq 0 \\ \dfrac{\partial S_{\mathbf{U},Z}(\mathcal{V})}{\partial a} = \dfrac{\varsigma_2 \cdot \sigma_1^2}{(\varsigma_1 \cdot V_v + \varsigma_2)^2} \cdot \dfrac{\partial V_v}{\partial a} & \geq 0 \end{cases}$$

Now, let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property:

$$\forall h > 0, \quad \begin{cases} \lim\limits_{a\to+\infty} C(h) = (1-\eta) \cdot V \\ \lim\limits_{a\to 0} C(h) = 0 \end{cases}$$

Then the mean value $\overline{C}(v,v)$ (over a finite domain v) has the following limits:

$$\begin{cases} \lim\limits_{a\to+\infty} \overline{C}(v,v) = (1-\eta) \cdot V \\ \lim\limits_{a\to 0} \overline{C}(v,v) = 0 \end{cases}$$

Using the equality $V_v = \overline{C}(v,v)$ and the expression of block sensitivity indices given in property I.2 on page 237, we obtain the expected limits. □

I.4 Change of support effect

Property I.5 (Change of support)**.** *Consider $v \subset \Omega$ and $\mathcal{V} \subset \Omega$ two blocks such that $v \ll \mathcal{V}$. Let denote by $D^2(v \mid \mathcal{V})$ the dispersion variance of v within $\mathcal{V}$ for $Z(\mathbf{x})$ random field. Block-based sensitivity indices of $\mathbf{U}$ and Z verify:*

$$\boxed{\begin{aligned} S_Z(\mathcal{V}) - S_Z(v) &= -\frac{\mu_1^2 \cdot \varsigma_2}{\varrho} \cdot D^2(v \mid \mathcal{V}) \quad \leq 0 \\ S_{\mathbf{U}}(\mathcal{V}) - S_{\mathbf{U}}(v) &= +\frac{\varsigma_1 \cdot \varsigma_2}{\varrho} \cdot D^2(v \mid \mathcal{V}) \quad \geq 0 \\ S_{\mathbf{U},Z}(\mathcal{V}) - S_{\mathbf{U},Z}(v) &= -\frac{\sigma_1^2 \cdot \varsigma_2}{\varrho} \cdot D^2(v \mid \mathcal{V}) \quad \leq 0 \end{aligned}}$$

with:

$$\varrho = [\varsigma_1 \cdot V_v + \varsigma_2] \cdot [\varsigma_1 \cdot V_v + \varsigma_2]$$

Proof of property I.5. Consider $v \subset \Omega$ and $\mathcal{V} \subset \Omega$ two blocks such that $v \ll \mathcal{V}$. Let denote by $\overline{C}(v,v)$ the mean value of covariance $C(h)$ when the two extremities of distance vector h describe the block v. We know that block variance V_v can be written (same proof as property H.5 on page 232):

$$V_v = \overline{C}(v,v)$$

Let denote by $D^2(v \mid \mathcal{V})$ the dispersion variance of v within $\mathcal{V}$ for $Z(\mathbf{x})$ random field. We have:

$$D^2(v \mid \mathcal{V}) = \overline{C}(v,v) - \overline{C}(\mathcal{V},\mathcal{V})$$

(see Journel Journel and Huijbregts (1978), section II.C.2 equation II.35). Using this expression of $D^2(v \mid \mathcal{V})$ with the equality $V_v = \overline{C}(v,v)$ and the expression of ratio $S_Z(\mathcal{V})/S_{\mathbf{U}}(\mathcal{V})$ for a linear model (property I.2 on page 237), we obtain:

$$\begin{aligned} S_Z(\mathcal{V}) &= S_Z(v) - \frac{\mu_1^2 \cdot \varsigma_2}{\varrho} \cdot D^2(v \mid \mathcal{V}) \\ S_{\mathbf{U}}(\mathcal{V}) &= S_{\mathbf{U}}(v) + \frac{\varsigma_1 \cdot \varsigma_2}{\varrho} \cdot D^2(v \mid \mathcal{V}) \\ S_{\mathbf{U},Z}(\mathcal{V}) &= S_{\mathbf{U},Z}(v) - \frac{\sigma_1^2 \cdot \varsigma_2}{\varrho} \cdot D^2(v \mid \mathcal{V}) \end{aligned}$$

with:

$$\varrho = (\varsigma_1 \cdot V_{\mathcal{V}} + \varsigma_2) \cdot (\varsigma_1 \cdot V_v + \varsigma_2)$$

By noting that $\varsigma_1 > 0$, $\varsigma_2 > 0$ (see property I.2 on page 237), $\mu_1^2 > 0$, $\sigma_1^2 > 0$ and $V_{\mathcal{V}} > 0$, we have $\varrho > 0$. Noting that the dispersion variance is positive (see Journel Journel and Huijbregts (1978), equation II.33), we can conclude that:

$$S_Z(\mathcal{V}) < S_Z(v) \qquad S_{\mathbf{U}}(\mathcal{V}) > S_{\mathbf{U}}(v) \qquad S_{\mathbf{U},Z}(\mathcal{V}) < S_{\mathbf{U},Z}(v)$$

□

Property I.6 (Change of support effect - bis)**.** *Consider a block $v \subset \Omega$. Let denote by $v' \subset \Omega$ the block obtained from v by a homothety of center $O \in \Omega$ and ratio $\tau > 1$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field has the following property:*

$$\forall h \geq 0, \quad \frac{\partial C}{\partial h}(h) \leq 0$$

Then , block sensitivity indices verify:

$$\boxed{S_Z(v') < S_Z(v) \quad S_{\mathbf{U}}(v') > S_{\mathbf{U}}(v) \quad S_{U,Z}(v') < S_{\mathbf{U},Z}(v)}$$

Proof of property I.6 on the preceding page. Consider a block $v \subset \Omega$. Let denote by $v' \subset \Omega$ the block obtained from v by a homothety of center $O \in \Omega$ and ratio $\tau > 1$. Let assume that the covariance function $C(h)$ of $Z(\mathbf{x})$ random field is differentiable and has the following property:

$$\forall h \geq 0, \quad \frac{\partial C}{\partial h}(h) \leq 0$$

We deduce from this property that $V'_v < V_v$ (see H.11 on page 235 for a detailed proof). Using the expression of block sensitivity indices for a linear model (property I.2 on page 237), we find the expected property. □

J Appendix to §4.2: proofs

In this appendix, we give a number of proofs related to the second section §4.2 on page 153 of Chapter 4. Please refer to this section for the definitions of all variables.

Property J.1 (Ratio of proxies for block sensitivity indices)**.** *The ratio of proxies for block sensitivity indices can be written as:*

$$\boxed{\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_{\mathbf{U}}(\Omega)} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{\tilde{V}^*_\Omega}{V^*}} \tag{$\mathcal{A}$.56}$$

Proof. The proof of property J.1 is very similar to the proof given in §4.1.6 on page 151 for the expression of the ratio of block sensitivity indices. Here, the ratio of proxies of block sensitivity indices $\tilde{S}_Z(\Omega)$ and $\tilde{S}_{\mathbf{U}}(\Omega)$ defined with respect to the approximated model output $\tilde{Y}_\Omega$ is equal to:

$$\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_{\mathbf{U}}(\Omega)} = \frac{\mathrm{var}\left(\mathbb{E}\left[\tilde{Y}_\Omega \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}\right]\right)}{\mathrm{var}\left(\mathbb{E}\left[\tilde{Y}_\Omega \mid \mathbf{U}\right]\right)} \tag{$\mathcal{A}$.57}$$

The conditional expectation of the approximated aggregated model output $\tilde{Y}_\Omega$ given $Z(\mathbf{x})$ gives:

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_\Omega \mid Z\right] &= \mathbb{E}\left[\left(\frac{1}{G}\sum_{i=1}^{G} Y(\mathbf{x}_i)\right) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}\right] \quad \text{(definition of } \tilde{Y}_\Omega\text{)} \\
&= \frac{1}{G}\sum_{i=1}^{G} \mathbb{E}\left[Y(\mathbf{x}_i) \mid Z(\mathbf{x}_i)\right] \quad \text{(for a point-based model)} \\
&= \frac{1}{G}\sum_{i=1}^{G} \mathbb{E}_Z Y(\mathbf{x}_i) \quad \text{(definition of } \mathbb{E}_Z Y(\mathbf{x})\text{)}
\end{aligned} \tag{$\mathcal{A}$.58}$$

Thus we have $\mathrm{var}\left(\mathbb{E}\left[\tilde{Y}_\Omega \mid Z\right]\right) = \mathrm{var}\left(1/G \sum_{i=1}^{G} \mathbb{E}_Z Y(\mathbf{x}_i)\right) = \tilde{V}^*_\Omega$ (definition of $\tilde{V}^*_\Omega$). Moreover, the conditional expectation of the approximated model output $\tilde{Y}_\Omega$ given input $\mathbf{U}$ gives:

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_\Omega \mid \mathbf{U}\right] &= \mathbb{E}\left[\left(\frac{1}{G}\sum_{i=1}^{G} Y(\mathbf{x}_i)\right) \mid \mathbf{U}\right] \quad \text{(definition of } \tilde{Y}_\Omega\text{)} \\
&= \frac{1}{G}\sum_{i=1}^{G} \mathbb{E}\left[Y(\mathbf{x}_i) \mid \mathbf{U}\right]
\end{aligned} \tag{$\mathcal{A}$.59}$$

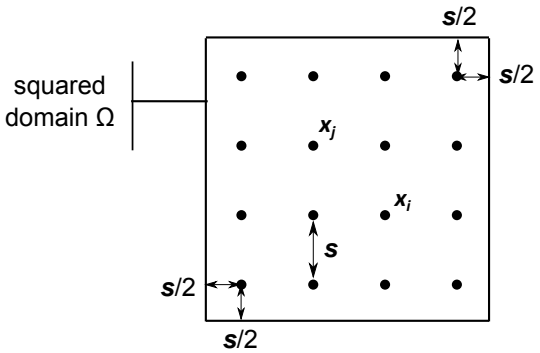

Figure $\mathcal{A}$.16: Spatial domain Ω, regularly positionned points $\mathbf{x}_i$, spacing s (total number of points: $G = |\Omega|/s^2$)

$\mathbb{E}[Y(\mathbf{x}_i) \mid \mathbf{U}]$ does not depend on site $\mathbf{x}_i$ under the stationarity of SRF $Z(\mathbf{x})$; thus, we have in particular $\mathbb{E}\left[\tilde{Y}_\Omega \mid \mathbf{U}\right] = \mathbb{E}[Y(\mathbf{x}^*) \mid \mathbf{U}]$ in which $\mathbf{x}^* \in \Omega$ is some specific site, and $\text{var}\left(\mathbb{E}\left[\tilde{Y}_\Omega \mid \mathbf{U}\right]\right) = \text{var}\left(\mathbb{E}[Y(\mathbf{x}^*) \mid \mathbf{U}]\right)$. Combining these expressions with Eqn. ($\mathcal{A}$.57) on the preceding page yields:

$$\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_\mathbf{U}(\Omega)} = \frac{\tilde{V}^*_\Omega}{\text{var}(\mathbb{E}[Y(\mathbf{x}^*) \mid \mathbf{U}])} \tag{$\mathcal{A}$.60}$$

Besides, the ratio of site sensitivity indices gives [Eqn. (4.4) on page 143]:

$$\cdot\frac{S_Z}{S_\mathbf{U}} = \frac{\text{var}(\mathbb{E}[Y(\mathbf{x}^*) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}])}{\text{var}(\mathbb{E}[Y(\mathbf{x}^*) \mid \mathbf{U}])} \tag{$\mathcal{A}$.61}$$

We notice that for point-based models $\text{var}[\mathbb{E}(Y(\mathbf{x}^*) \mid \{Z(\mathbf{x}) : \mathbf{x} \in \Omega\}] = \text{var}[\mathbb{E}_Z Y(\mathbf{x}^*)] = V^*$ (definition of $\mathbb{E}_Z Y(\mathbf{x})$ [Eqn. (4.7) on page 144]). Finally, it follows from Eqn. ($\mathcal{A}$.60) and Eqn. ($\mathcal{A}$.61) that:

$$\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_\mathbf{U}(\Omega)} = \frac{S_Z}{S_\mathbf{U}} \cdot \frac{\tilde{V}^*_\Omega}{V^*} \tag{$\mathcal{A}$.62}$$

□

Property J.2 (Block variance approximation error)**.** *Let denote by $\epsilon = V^*_\Omega - \tilde{V}^*_\Omega$ the error made when approximating V^*_Ω with $\tilde{V}^*_\Omega$. We have:*

$$\boxed{\epsilon = \frac{C^{*,+}(0) - C^*(0)}{G} + \left[\overline{C^*}(\Omega,\Omega) - \frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right]} \tag{$\mathcal{A}$.63}$$

where:

$$C^{*,+}(h) = \begin{cases} C^*(h) & \forall h > 0 \\ \lim\limits_{h \to 0^+} C^*(h) & \text{if } h = 0 \end{cases} \tag{$\mathcal{A}$.64}$$

The first part of the equation is known as the "zero-effect": each time two points coincide ($\mathbf{x}_i = \mathbf{x}_j$) in the estimation of $\overline{C^*}(\Omega,\Omega)$, a greater importance is given to the value $C^*(0)$. Is causes the mean value $\overline{C^*}(\Omega,\Omega)$ to be systematically over-estimated (see Journel and Huijbregts (1978), p.96 for more details). The second part of the equation is the error of a Riemann approximation of an integral over Ω^2.

Proof of property J.2. Variance $\tilde{V}^*_\Omega$ is equal to:

$$\begin{aligned}
\tilde{V}^*_\Omega &= \text{var}\left[\frac{1}{G}\sum_{i=1}^{G} \mathbb{E}_Z Y(\mathbf{x}_i)\right] && \text{(definition)} \\
&= \frac{1}{G^2}\sum_{i,j=1}^{G} \text{cov}\left[\mathbb{E}_Z Y(\mathbf{x}_i), \mathbb{E}_Z Y(\mathbf{x}_j)\right] \\
&= \frac{1}{G^2}\sum_{i,j=1}^{G} C^*(\mathbf{x}_i - \mathbf{x}_j) && \text{(property H.4)} \\
&= \frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j) + \frac{1}{G^2}\sum_{i=1}^{G}\left[C^*(0) - C^{*,+}(0)\right] && \text{(definition of } C^{*,+}\text{)} \\
&= \frac{C^*(0) - C^{*,+}(0)}{G} + \frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)
\end{aligned}$$

Let $\epsilon = V_\Omega^* - \tilde{V}_\Omega^*$. Using the equality $V_\Omega^* = \overline{C^*}(\Omega,\Omega)$ (property H.5 on page 232), we obtain:

$$\epsilon = \frac{C^{*,+}(0) - C^*(0)}{G} + \left[\overline{C^*}(\Omega,\Omega) - \frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] \tag{A.65}$$

□

Property J.3 (Convergence)**.** *The ratio of proxies of sensitivity indices $\tilde{S}_Z(\Omega)/\tilde{S}_\mathbf{U}(\Omega)$ converges toward the ratio of exact sensitivity indices when the number of points $\mathbf{x}_i$ grow, that is, when the spacing s of the set of points tends to zero (Figure A.16 on page 242):*

$$\boxed{\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_\mathbf{U}(\Omega)} \underset{s\to 0}{=} \frac{S_Z(\Omega)}{S_\mathbf{U}(\Omega)} + \mathcal{O}(s)} \tag{A.66}$$

Proof of property J.3. Let denote $\epsilon = V_\Omega^* - \tilde{V}_\Omega^*$. Using the expression of the ratio of block sensitivity indices [Eqn. (4.8) on page 145] and the expression of the ratio of proxies for these indices (property J.1 on page 242), we have:

$$\frac{S_Z(\Omega)}{S_\mathbf{U}(\Omega)} - \frac{\tilde{S}_Z(\Omega)}{\tilde{S}_\mathbf{U}(\Omega)} = \frac{S_Z}{S_\mathbf{U}} \cdot \frac{\epsilon}{V^*} \tag{A.67}$$

Using property J.2 on the preceding page, we have:

$$\epsilon = \frac{C^{*,+}(0) - C^*(0)}{G} + \left[\overline{C^*}(\Omega,\Omega) - \frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] \tag{A.68}$$

The second part of the right member of this equation, that we denote by A, is the error on the Riemann approximation of an integral over domain Ω^2:

$$A = \overline{C^*}(\Omega,\Omega) - \left[\frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] \tag{A.69}$$

A converges to 0 when $G \to +\infty$. We will prove now that it is a $\mathcal{O}\left(\frac{1}{\sqrt{G}}\right)$. Let denote by c_i the square cell of width s, centered on the point $\mathbf{x}_i$. Each cell c_i has a surface area $|c_i| = s^2 = |\Omega|/G$ (Figure A.16 on page 242). We can write:

$$\begin{aligned}
A &= \frac{1}{|\Omega|^2}\iint_{\substack{\mathbf{x}\in\Omega\\ \mathbf{y}\in\Omega}} C^*(\mathbf{x}-\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} - \left[\frac{1}{G^2}\sum_{i,j=1}^{G} C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] \\
&= \frac{1}{|\Omega|^2}\cdot\sum_{i,j=1}^{G}\left[\iint_{\substack{\mathbf{x}\in c_i\\ \mathbf{y}\in c_j}} C^*(\mathbf{x}-\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} - \frac{|\Omega|^2}{G^2}C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] \\
&= \frac{1}{|\Omega|^2}\sum_{i,j=1}^{G}\iint_{\substack{\mathbf{x}\in c_i\\ \mathbf{y}\in c_j}} \left[C^*(\mathbf{x}-\mathbf{y}) - C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right]\,d\mathbf{x}\,d\mathbf{y} \quad \text{because} \quad |c_i|\cdot|c_j| = s^4 = \frac{|\Omega|^2}{G^2}
\end{aligned} \tag{A.70}$$

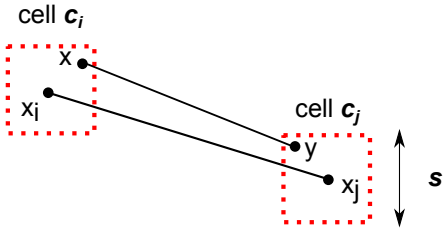

Figure $\mathcal{A}$.17: Two cells c_i and c_j centered on points $\mathbf{x}_i$ and $\mathbf{x}_j$, with two points $\mathbf{x} \in c_i$ and $\mathbf{y} \in c_j$

We need to find the maximum gap between $||\mathbf{x}_i - \mathbf{x}_j||$ (distance between the two centers of cells c_i and c_j) and $||\mathbf{x} - \mathbf{y}||$ (distance between two points $\mathbf{x} \in c_i$ and $\mathbf{y} \in c_j$) (Figure $\mathcal{A}$.17).

Consider two points $\mathbf{x} \in c_i$ and $\mathbf{y} \in c_j$:

$$\overrightarrow{\mathbf{xy}} = \overrightarrow{\mathbf{xx}_i} + \overrightarrow{\mathbf{x}_i\mathbf{x}_j} + \overrightarrow{\mathbf{x}_j\mathbf{y}}$$

Thus:

$$-||\overrightarrow{\mathbf{xx}_i}|| - ||\overrightarrow{\mathbf{x}_j\mathbf{y}}|| \le ||\overrightarrow{\mathbf{xy}}|| - ||\overrightarrow{\mathbf{x}_i\mathbf{x}_j}|| \le ||\overrightarrow{\mathbf{xx}_i}|| + ||\overrightarrow{\mathbf{x}_j\mathbf{y}}|| \tag{$\mathcal{A}$.71}$$

As $\mathbf{x} \in c_i$ and $\mathbf{y} \in c_j$, the distances $||\overrightarrow{\mathbf{xx}_i}||$ and $||\overrightarrow{\mathbf{x}_j\mathbf{y}}||$ are lower than $s/\sqrt{2}$ (Figure $\mathcal{A}$.17). Thus, we obtain:

$$|\, ||\overrightarrow{\mathbf{xy}}|| - ||\overrightarrow{\mathbf{x}_i\mathbf{x}_j}|| \,| \le \sqrt{2} \cdot s \tag{$\mathcal{A}$.72}$$

so:

$$|C^*(\mathbf{x}-\mathbf{y}) - C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)| \le \Delta_{\max} \quad \text{with} \quad \Delta_{\max} = \sqrt{2} \cdot s \cdot \sup_{h \in \mathbb{R}^+} \left|\frac{dC^*(h)}{dh}\right| \tag{$\mathcal{A}$.73}$$

Thus the following inequalities hold:

$$\begin{aligned}
A &= \frac{1}{|\Omega|^2} \sum_{i,j=1}^{G} \iint_{\substack{\mathbf{x}\in c_i \\ \mathbf{y}\in c_j}} \left[C^*(\mathbf{x}-\mathbf{y}) - C^{*,+}(\mathbf{x}_i - \mathbf{x}_j)\right] d\mathbf{x}\, d\mathbf{y} \\
&\le \frac{1}{|\Omega|^2} \sum_{i,j=1}^{G} \iint_{\substack{\mathbf{x}\in c_i \\ \mathbf{y}\in c_j}} \Delta_{\max}\, d\mathbf{x}\, d\mathbf{y} \\
&\le \frac{1}{|\Omega|^2} \sum_{i,j=1}^{G} |c_i| \cdot |c_j| \cdot \Delta_{\max} \\
&\le \frac{G^2 \cdot s^4}{|\Omega|^2} \cdot \Delta_{\max} \quad \text{because} \quad |c_i| = |c_j| = s^2 \\
&\le \Delta_{\max} \quad \text{because} \quad s^2 = |\Omega|/G
\end{aligned} \tag{$\mathcal{A}$.74}$$

Finally, using the expression of ϵ (property J.2 on page 243) and the expression of $\Delta_{\max}$ given above, we obtain:

$$|\epsilon| \le \frac{C^*(0) - C^{*,+}(0)}{G} + \sqrt{2} \cdot s \cdot \sup_{h \in \mathbb{R}^+} \left|\frac{dC^*(h)}{dh}\right| \tag{$\mathcal{A}$.75}$$

using $1/G = s^2/|\Omega|$, we obtain:

$$|\epsilon| \le \frac{C^*(0) - C^{*,+}(0)}{|\Omega|} \cdot s^2 + \sqrt{2} \cdot \sup_{h \in \mathbb{R}^+} \left|\frac{dC^*(h)}{dh}\right| \cdot s \tag{$\mathcal{A}$.76}$$

Thus:

$$\epsilon \underset{s\to 0}{=} \mathcal{O}(s) \tag{$\mathcal{A}$.77}$$

Using Eqn. ($\mathcal{A}$.67) on page 244, we find:

$$\frac{\tilde{S}_Z(\Omega)}{\tilde{S}_{\mathbf{U}}(\Omega)} \underset{s\to 0}{=} \frac{S_Z(\Omega)}{S_{\mathbf{U}}(\Omega)} + \mathcal{O}(s) \tag{$\mathcal{A}$.78}$$

□

Sensitivity analysis of spatial models using geostatistical simulation

Nathalie SAINT-GEOURS[1], Christian LAVERGNE[2], Jean-Stéphane BAILLY[3] & Frédéric GRELOT[4]

[1] AgroParisTech, UMR TETIS, I3M, France,
saintge@teledetection.fr

[2] I3M, Université Montpellier III, France,
christian.lavergne@univ-montp3.fr

[3] AgroParisTech, UMR TETIS, France,
bailly@teledetection.fr

[4] Cemagref, UMR G-EAU, France,
grelot@cemagref.fr

Peer-reviewed IAMG 2011 publication

doi:10.5242/iamg.2011.0172

Abstract

Geostatistical simulations are used to perform a global sensitivity analysis on a model $Y = f(X_1 ... X_k)$ where one of the model inputs X_i is a continuous 2D-field. Geostatistics allow specifying uncertainty on X_i with a spatial covariance model and generating random realizations of X_i. These random realizations are used to propagate uncertainty through model f and estimate global sensitivity indices. Focusing on variance-based global sensitivity analysis (GSA), we assess in this paper how sensitivity indices vary with covariance parameters (range, sill, nugget). Results give a better understanding on how and when to use geostatistical simulations for sensitivity analysis of spatially distributed models.

1 Introduction

Numerous spatial models are developed to support decision making in various fields of environmental management. These models use environmental data that is spatially distributed, including maps derived from sampled data (e.g. digital elevation model, soil map, etc.). These spatial inputs are always partly uncertain, due to measurement errors, lack of knowledge, aleatory variability (see Refsgaard et al., 2007 for a discussion on the various sources of uncertainty in model inputs). In order to provide confidence in these models, uncertainty analysis (UA) and sensitivity analysis (SA) are increasingly recognized as important steps in the modelling process. They allow robustness of model predictions to be checked and help identifying the input factors that account for most of model output variability (Saltelli et al., 2008).

Geostatistical simulation has an important role to play in UA/SA of models $Y = f(X_1 \dots X_k)$ when some model input X_i is a continuous 2D-field. Geostatistics first offers a way to describe the uncertainty on spatial input X_i with a spatial covariance model. Then, random realizations of X_i can be generated through geostatistical simulation (Journel and Huijbregts, 1978). These random realizations can be used to propagate uncertainty through model f and discuss the resulting uncertainty on model output Y (Aerts et al., 2003 - on a problem of optimal location of a ski run; Ruffo et al., 2006 - on hydrocarbon exploration risk evaluation). Within variance-based global sensitivity analysis (GSA) framework, these random realizations can also be sampled alongside with other scalar model inputs to estimate sensitivity indices for each model input (Lilburne & Tarantola, 2009).
Still, a practical problem remains for modellers who intend to use geostatistical simulations in UA/SA of a spatially distributed model: covariance parameters which describe uncertainty on input 2D-field X_i must be estimated carefully, but there is usually few data to support this estimation. At the same time, UA/SA results are known to depend heavily on the specification of uncertainty on model inputs. Thus, the following questions arise: to what extent are UA/SA results influenced by spatial covariance parameters? In which cases the uncertainty on input 2D-field X_i accounts for a large or a small part of total variability of model output?
To answer these questions, this article aims at determining, in the context of spatial GSA, how sensitivity indices depend on the covariance parameters which describe uncertainty on spatially distributed model inputs. We first describe a simple spatial model $Y = f(X, Z)$ with two inputs: a scalar input X and a 2D spatially distributed input $Z(u)$ (section 2). Then we present variance-based global sensitivity analysis (section 3), and show into details how to estimate sensitivity indices on model M using geostatistical simulations of 2D-field $Z(u)$ (section 4).We finally assess the impact of the three usual covariance parameters (range, sill, nugget) on sensitivity indices in model M (section 5). Our results might well prove useful in better understanding the results of a spatial GSA and in deciding whether it is necessary to carefully estimate spatial covariance parameters to describe uncertainty on input 2D-fields.

2 A simple spatially distributed model *M*

For sake of clarity, we will base our paper on a simple case-study. We describe in this section an example of a spatially distributed model M.

2.1 Description of model M

Consider a spatial domain $D \subset R^2$. For numerical application, we represent domain D by a regular square grid G of size 50×50. We will study in the following sections a model M with two inputs:

$$Y = M(X, Z)$$

where:

- $X = (X_1, X_2)$ is a vector of two scalars
- *Z(u)* is a 2D continuous field defined on domain *D*.
- model output *Y(u)* is also a 2D continuous field defined by :

$$\forall u \in D, \quad Y(u) = f(X, Z(u))$$

Function *f(.,.)* can be any mapping from $\mathbb{R}^3$ to $\mathbb{R}$. For numerical application, we arbitrarily choose the following mapping:

$$f(x_1, x_2, z) = 10^{-3} \cdot \left(x_1^2 + x_2 \cdot e^{0.036 \cdot Z(u)} + 40 \cdot Z(u)\right)$$

Model *M* is a "point-based model": the value of model output *Y(u)* at any point $u \in D$ only depends on the scalar inputs (X_1, X_2) and on the value of *Z(u)* at the same point *u*. Point-based models are encountered in many environmental applications. For example, *M* could be a spatially distributed model used for economic assessment of flood risk: in this case, model input *Z(u)* could be a map of the maximal water levels reached during a flood event over a given area *D*, $X = (X_1, X_2)$ would be a set of economic parameters, and model output *Y(u)* would be the map of expected damages due to the flood over the area.

2.2 Output of interest

In order to perform sensitivity analysis of model *M*, we need to consider a single scalar quantity of interest derived from model output *Y(u)*. In most applications, the output of interest is either the value of 2D-field *Y(u)* at some specific point *u* of the study area, or the mean (or total) value of *Y(u)* over a given zone within the study area. Here we define the output of interest Y_D as the mean value of field *Y(u)* over spatial domain *D*:

$$Y_D = \frac{1}{|D|} \cdot \int_{u \in D} Y(u) \cdot du$$

In the following sections we will use variance-based global sensitivity analysis to assess the variability of Y_D due to the uncertainty on model inputs *X* and *Z(u)*.

3 Variance-based global sensitivity analysis

Sensitivity analysis (SA) aims at a studying how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model inputs. Among the various available SA techniques (see Helton and Davis, 2006 for a review), variance-based global sensitivity analysis (GSA) has several advantages: it explores widely the space of uncertain input factors and is suitable for complex models with non-linear effects and interactions among factors.

GSA is based on the decomposition of the variance of model output *Y* in conditional variances. It leads to the definition of two importance measures for each input factor X_i of a model: first-order sensitivity index S_i and total-order sensitivity index ST_i. First-order sensitivity index of input factor X_i is defined by:

$$S_i = \frac{Var\left(E\left[Y \middle| X_i\right]\right)}{Var(Y)}$$

S_i measures the main effect contribution of input factor X_i to the variance of model output Y. It is the expected part of output variance $Var(Y)$ that could be reduced if input factor X_i was perfectly known. Total order sensitivity index ST_i of input factor X_i is defined as:

$$ST_i = \frac{E\left(Var\left[Y \middle| X_{\sim i}\right]\right)}{Var(Y)}$$

where $X_{\sim i}$ denotes all input factors but X_i. ST_i measures the contribution of input factor X_i and all its interactions with other input factors X_j to the variance of model output Y. It is the expected part of output variance $Var(Y)$ that would remain if all input factors but X_i were perfectly known.

Sensitivity indices can be used to identify the model inputs that account for most of model output variability (input factors X_i with high first order indices S_i); it may lead to model simplification by identifying model inputs that have little influence on model output variance (input factors X_i with low total order sensitivity indices ST_i); it also allows discussing the contribution of interactions between input factors to the model output variance (comparison between first and total order sensitivity indices). For more details on GSA basics, see Saltelli et al., 2008.

4 Estimating sensitivity indices using geostatistical simulations

GSA was initially designed to study models with scalar inputs only. Some authors have suggested solutions to handle spatially distributed inputs as well (Volkova et al., 2008 ; Iooss & Ribatet, 2009; Ruffo et al., 2006; Lilburne & Tarantola, 2009). We describe in this section how to estimate sensitivity indices in model M by associating randomly generated realizations of uncertain 2D-field $Z(u)$ to scalar values, according to the approach developed by Lilburne and Tarantola.

Three steps are needed to apply GSA on model M (Figure 1):

1. modelling uncertainty on model inputs X and $Z(u)$
2. propagating input uncertainty through model M
3. estimating sensitivity indices

Each step is described in details in the following subsections.

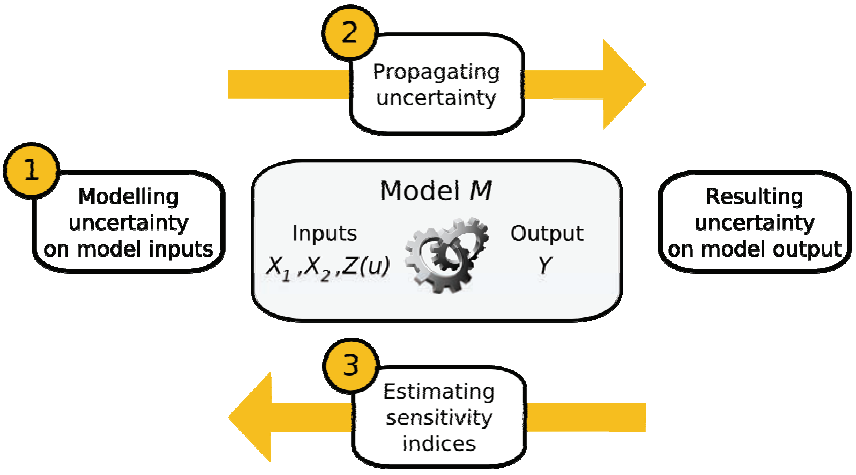

Figure 1: Steps of sensitivity analysis of model *M*

4.1 Modelling uncertainty using geostatistical simulations

The values of model inputs are always partly uncertain, due to measurement errors, lack of knowledge, natural variability, modelling errors... Within the GSA method, uncertainty on model inputs is described using a probabilistic framework (Table 1).

Table 1: Specification of uncertainty on model inputs

Model input	Model of uncertainty
$X=(X_1, X_2)$	X_1 and X_2 independent random variables: $X_1 \sim \mathcal{N}(12,24)$ and $X_2 \sim \mathcal{N}(1,20)$
$Z(u)$	Gaussian Random Field of mean $\mu=1$ and covariance $\rho_\theta(h)$

4.1.1 Uncertainty on model input *X*

Model input $X=(X_1,X_2)$ is a vector of two scalar factors. X_1 and X_2 are supposed to be independent random variables following Gaussian distribution $\mathcal{N}(12,24)$ and $\mathcal{N}(1,20)$ respectively (Gaussian distribution parameters were chosen arbitrarily).

4.1.2 Uncertainty on model input *Z(u)*

2D-field $Z(u)$ is supposed to be a Gaussian random field. It is assumed to be order 2 stationary with mean $\mu=1$. Its covariance function is denoted by $\rho_\theta(h)$:

$$\forall u \in D, \quad \forall h \geq 0, \quad \mathrm{cov}\left[Z(u), Z(u+h)\right] = \rho_\theta(h)$$

For numerical application, covariance function $\rho_\theta(h)$ is supposed to be exponential:

$$\forall h \geq 0, \quad \rho_\theta(h) = \sigma^2 \cdot \left[\eta \cdot \delta_0(h) + (1-\eta) \cdot e^{-\frac{h}{l}}\right]$$

Parameter $\theta = (l, \sigma^2, \eta)$ describes the covariance parameters: l is the practical range of covariance, σ^2 the sill and η the nugget.

In order to represent the uncertainty on 2D-field $Z(u)$, a set of $n=100$ random realizations is sampled. These random realizations are generated with Simple Random Sampling using LU decomposition of the covariance matrix (Journel and Huijbregts, 1978). These n random realizations are considered as equiprobable, and each realization is labelled with a unique integer in the set $\{1, ..., n\}$ (Figure 2).

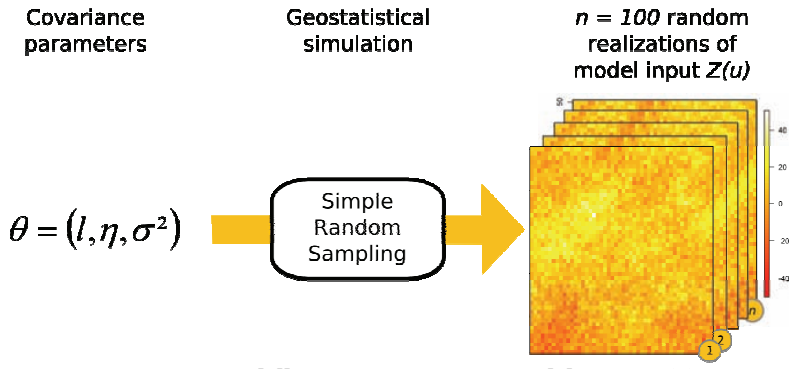

Figure 2: Modelling uncertainty on model input *Z(u)*

4.2 Propagating uncertainty through model M

Input uncertainty is propagated through model *M* using a sampling-based approach, according to "spatial GSA„ method (Lilburne and Tarantola, 2009).

4.2.1 Sampling of model inputs

Spatial GSA uses two quasi-random independent samples *A* and *B* of length *N=4096*, combined through several permutations, to explore the uncertainty domain of input factors *X* and *Z(u)*. The i^{th} line of sample *A* or *B* is a set $\left(X_1^{(i)}, X_2^{(i)}, z^{(i)}\right)$ where:

- $X_1^{(i)}$ is a random value drawn from pdf of input factor X_1
- $X_2^{(i)}$ is a random value drawn from pdf of input factor X_2
- $z^{(i)}$ is a random integer sampled from a discrete uniform distribution in *{1, ..., n}*. Each discrete level in *{1, ..., n}* is associated with a single random realization of *Z(u)* from the set of *n* maps previously generated (see 4.1.2). The value of $z^{(i)}$ indicates which random realization of *Z(u)* sould be used to evaluate model *M* for the i^{th} line of the sample.

4.2.2 Permutations

In order to estimate sensitivity indices for model inputs *X=(X_1, X_2)* and *Z(u)*, we must evaluate model *M* at points $\left(X_1^{(i)}, X_2^{(i)}, z^{(i)}\right)$ where only one of the three factors changes from a previous line $\left(X_1^{(j)}, X_2^{(j)}, z^{(j)}\right)$ where model *M* has already been evaluated. Thus, new samples are created by combining original samples *A* and *B*. For *j* = 1 to 3, a new sample $A_B^{(j)}$ is created: it is equal to sample *A*, except for the j^{th} column which is taken from sample *B* (Figure 3).

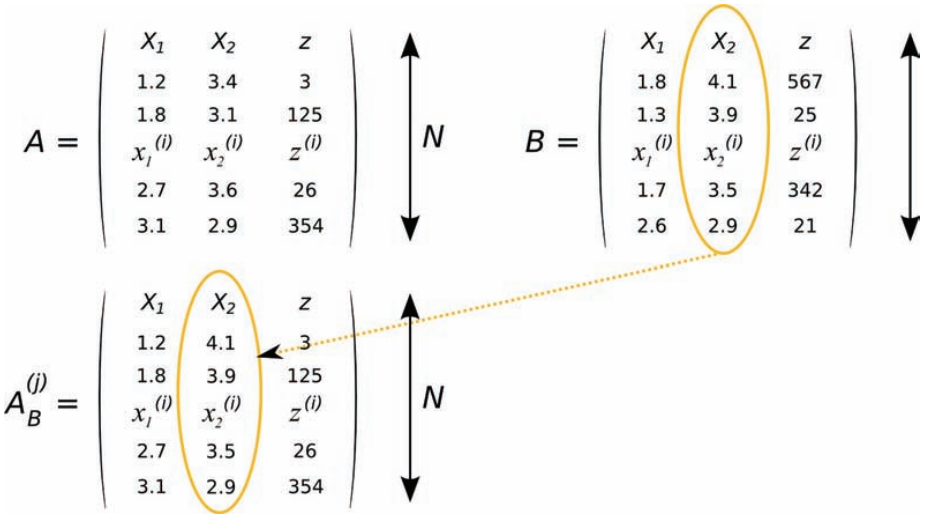

Figure 3: Creating sample $A_B^{(j)}$

4.2.3 Model runs

Model *M* is finally evaluated for each line of samples *A*, *B* and $A_B^{(j)}$ for *j=1..3*. Total number of model runs is $C = 5 \cdot N = 20480$. Each model run gives a value for the output of interest Y_D. We denote by Y_A, Y_B and $Y_{A_B^{(j)}}$ the vectors of length *N* giving the value of Y_D for each line of samples *A*, *B* and $A_B^{(j)}$.

4.3 Estimating sensitivity indices

First and total order sensitivity indices of the j^{th} input factor are estimated using expressions (1) and (2) given in (Saltelli et al., 2008). Model input $X=(X_1,X_2)$ is treated as a "group of factors"; components X_1 and X_2 were sampled independently from their pdf, but first order and total order sensitivity indices are estimated globally for the group $X=(X_1,X_2)$ (see section 1.2.15 of Saltelli et al., 2008 for a complete discussion on grouping model inputs in GSA).

$$S_j = \frac{\frac{1}{N}\cdot\sum_{i=1}^{N} Y_B^{(i)}\cdot Y_{A_B^{(j)}}^{(i)} - \frac{1}{N}\cdot\sum_{i=1}^{N} Y_B^{(i)}\cdot Y_A^{(i)}}{\frac{1}{N}\cdot\sum_{i=1}^{N} Y_A^{(i)}\cdot Y_A^{(i)} - \left(\frac{1}{N}\cdot\sum_{i=1}^{N} Y_A^{(i)}\right)\cdot\left(\frac{1}{N}\cdot\sum_{i=1}^{N} Y_B^{(i)}\right)} \qquad (1)$$

$$ST_j = \frac{\frac{1}{2N}\cdot\sum_{i=1}^{N}\left(Y_A^{(i)} - Y_{A_B^{(j)}}^{(i)}\right)^2}{\frac{1}{N}\cdot\sum_{i=1}^{N} Y_A^{(i)}\cdot Y_A^{(i)} - \left(\frac{1}{N}\cdot\sum_{i=1}^{N} Y_A^{(i)}\right)\cdot\left(\frac{1}{N}\cdot\sum_{i=1}^{N} Y_B^{(i)}\right)} \qquad (2)$$

We finally obtain four different sensitivity indices: first and total order sensitivity indices of model input $X=(X_1,X_2)$, denoted by S_X and ST_X ; first and total order sensitivity indices of model input $Z(u)$, denoted by S_Z and ST_Z. In the current case of a model with only two inputs (X and $Z(u)$), the following properties hold:

$$ST_X = S_X + S_{X,Z} \qquad \text{and} \qquad ST_Z = S_Z + S_{X,Z}$$

where $S_{X,Z} = 1 - S_X - S_Z$ is a second order sensitivity index which accounts for the contribution of the interaction between X and $Z(u)$ to the variance of model output Y_D. Thus, we will only pay attention in the following sections to first order indices S_X and S_Z.

5 Influence of covariance parameters on sensitivity indices

In this section, we want to assess how GSA results on model M are influenced by covariance parameters (l,σ^2,η). 26 different sets $\theta_k = (l_k,\sigma_k{}^2,\eta_k)$ of covariance range, sill and nugget are defined (Table 2). For each set θ_k of covariance parameters, GSA is performed as follows:

- a set of $n=100$ random realizations of input random field $Z(u)$ is generated using geostatistical simulation as described in 4.1
- uncertainty is propagated through model M as described in 4.2
- total variance of model output Y_D is computed
- first order sensitivity indices S_X and S_Z are estimated as described in 4.3.

The whole procedure is replicated 100 times. Then, for each set of covariance parameters, mean value of $Var(Y_D)$, S_X and S_Z and their 95% confidence interval over the 100 replicas are computed.

Table 2: Sets of covariance parameters

Set name	Covariance parameters		
	Range *l*	**Sill σ^2** (square root)	**Nugget η**
θ_1 to θ_8	5 to 40 (step 5)	70	0.1
θ_9 to θ_{16}	60	20 to 55 (step 5)	0.1
θ_{17} to θ_{26}	60	70	0.1 to 1 (step 0.1)

5.1 Influence of the ratio covariance range *l* / size of domain *D*

Fig 4. shows output variance $Var(Y_D)$ and sensitivity indices S_X and S_Z for increasing covariance range l (sets θ_1 to θ_8). It appears that the absolute contribution of model input $Z(u)$ to total output variance $Var(Y_D)$ increases with covariance range l, while absolute contribution of model input X remains constant. Accordingly, sensitivity index of model input $Z(u)$ increases with covariance range l, while sensitivity index of X decreases when covariance range l increases.

Let define the ratio r of covariance range l compared to the size of domain D: $r = l/|D|$. This numerical case-study illustrates the following property: the larger the ratio r, the larger the part of output variance $Var(Y_D)$ explained by the uncertainty on $Z(u)$. For a low ratio (i.e. when range l is small compared to the size of domain D), variability of $Z(u)$ is mainly "local", and spatial correlation of $Z(u)$ variability over domain D is weak. This "local" variability averages over domain D when model output Y_D is computed. Thus the uncertainty on input 2D-field $Z(u)$ has a small influence on output variance $Var(Y_D)$.

On the contrary, for a greater ratio r (i.e. when range l is large compared to the size of domain D), spatial correlation of $Z(u)$ variability over domain D is strong. The averaging effect of "local" variability of $Z(u)$ over domain D is weaker. Thus the uncertainty on input 2D-field $Z(u)$ has a larger influence on output variance $Var(Y_D)$.

5.2 Influence of covariance sill

Fig 5. shows output variance $Var(Y_D)$ and sensitivity indices S_X and S_Z for increasing covariance sill σ^2 (sets θ_9 to θ_{16}). It appears that the absolute contribution of model input $Z(u)$ to total output variance $Var(Y_D)$ increases with covariance sill σ^2, while absolute contribution of model input X remains constant. Accordingly, sensitivity index of model input $Z(u)$ increases with covariance sill σ^2, while sensitivity index of X decreases when covariance sill σ^2 increases.

This numerical case-study illustrates the following straightforward property: the larger the covariance sill σ^2 in random field $Z(u)$, the larger the part of output variance $Var(Y_D)$ explained by the uncertainty on $Z(u)$. Covariance sill σ^2 controls the overall variability of model input $Z(u)$, thus sensitivity index of $Z(u)$ with respect to model output Y_D is a monotonically increasing function of sill σ^2.

5.3 Influence of covariance nugget

Fig 6. shows output variance $Var(Y_D)$ and sensitivity indices S_X and S_Z for increasing covariance nugget h (sets θ_{17} to θ_{26}). It appears that the absolute contribution of model input $Z(u)$ to total output variance $Var(Y_D)$ decreases when covariance nugget h increases, while absolute contribution of model input X remains constant. Accordingly, sensitivity index of model input $Z(u)$ decreases when covariance nugget h increases.

Nugget parameter h controls the intensity of "noise" in Gaussian random field $Z(u)$. When h is close to 1, the largest part of $Z(u)$ variability is due to the "nugget effect", i.e. to "local" noise at each point $u \in D$ with no spatial correlation. This local noise averages over domain D when

model output Y_D is computed. Thus the uncertainty on input 2D-field *Z(u)* has a small influence on output variance $Var(Y_D)$. On the contrary, for a lower value of nugget parameter, (h close to 0), most of the uncertainty in random field *Z(u)* is spatially correlated, and local noise plays a small part. The averaging effect of uncorrelated variability of *Z(u)* over domain *D* is weaker. Thus the uncertainty on input 2D-field *Z(u)* has a larger influence on output variance $Var(Y_D)$.

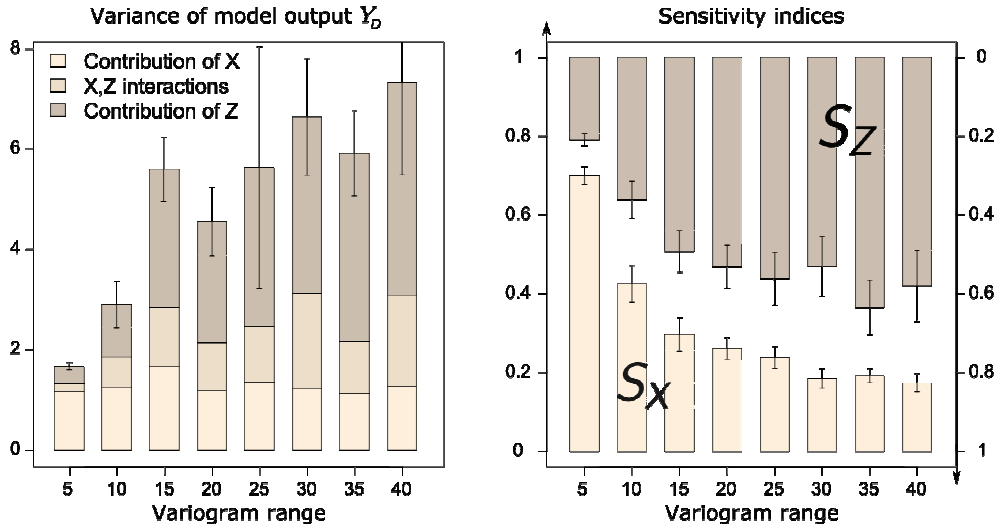

Figure 4: Influence of covariance range *l* on GSA results. (left) Total variance of model output Y_D and contribution of model inputs. (right) First order sensitivity indices S_X and S_Z. (error bars show 95% confidence interval over 100 replicas).

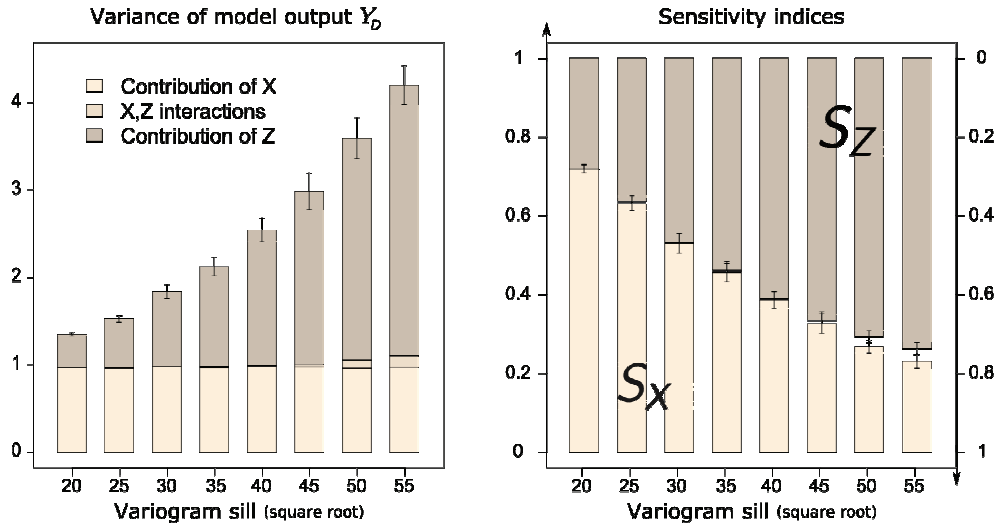

Figure 5: Influence of covariance sill σ^2 on GSA results. (left) Total variance of model output Y_D and contribution of model inputs. (right) First order sensitivity indices S_X and S_Z (error bars show 95% confidence interval over 100 replicas).

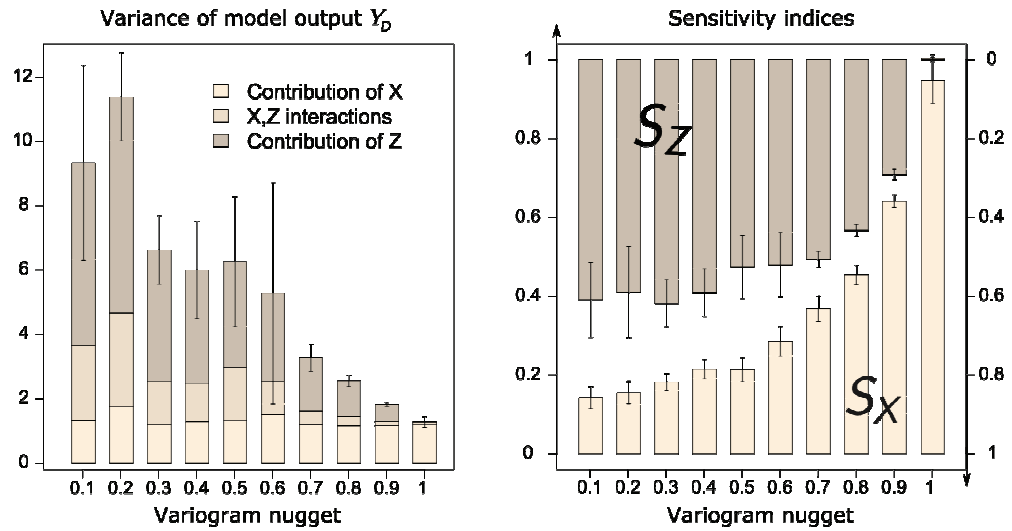

Figure 6: Influence of covariance nugget η on GSA results. (left) Total variance of model output Y_D and contribution of model inputs. (right) First order sensitivity indices S_X and S_Z (error bars show 95% confidence interval over 100 replicas).

6 Discussion

This research sought to illustrate on a simple case-study how to use geostatistical simulation to perform variance-based global sensitivity analysis (GSA) on a spatially distributed model. We also aimed at exploring how GSA results depend on covariance parameters chosen to describe uncertainty on spatially distributed model inputs.

6.1 Using geostatistical simulation for spatial GSA

We demonstrated on a simple case-study the suitability of "spatial GSA" approach (Lilburne & Tarantola, 2009) to perform sensitivity analysis on a spatially distributed model with continuous 2D-fields inputs. Geostatistical simulation was used to generate a set of *n* random realizations of continuous 2D-field *Z(u)* and estimate sensitivity indices of uncertain model inputs *Z(u)* and *X* through a sampling-based approach. Spatial GSA makes it possible to account for the relative contribution of each uncertain model input to the total variance of model output. It helps assessing model robustness and should be systematically performed when developing a model with uncertain spatial inputs. Nevertheless, two limits of this approach must be highlighted:

- spatial GSA is a sampling-based approach which needs lots of model runs to estimate sensitivity indices. As a consequence, it is limited to models with low CPU-cost. For high CPU-cost models, other sensitivity analysis methods such as Elementary Effects or One-At-a-Time should be applied (see Saltelli et al., 2008).
- spatial GSA uses a set of *n* random realizations to represent the uncertainty on spatial input *Z(u)* (assumed to be a Gaussian Random Field). When *n* is too low, the small set of map simulations fails to capture the overall variability of *Z(u)*, and sensitivity indices estimates S_X and S_Z are biased. Previous work had been carried out to compare the use of two different geostatistical simulation algorithms (Simple Random Sampling and Latin Hypercube Sampling) to generate realizations of spatial input *Z(u)* for GSA (Kyriakydis, 2005; Saint-Geours et al., 2010), but no optimal sampling strategy was found to reduce this bias.

6.2 Impact of spatial covariance parameters on spatial GSA

The influence of covariance range, sill and nugget on sensitivity indices was assessed on a simple case-study. It was initially suggested that covariance parameters chosen to describe uncertainty on spatial input $Z(u)$ would influence GSA results. Our results prove such to be the case. On our case-study, it appears that first order sensitivity index S_Z of model input $Z(u)$ is a monotically increasing function of both covariance range l and covariance sill σ^2, and a decreasing function of covariance nugget η.

These properties were only illustrated on a simple case-study with a specific model M and an exponential covariance function. Nevertheless, it can be analytically shown (on-going work) that these properties are actually verified for any monotically increasing covariance function and for any point-based model M where mapping f is square-integrable.

Results of this study may well help modellers when estimating spatial covariance parameters to describe uncertainty on a spatial input $Z(u)$ for sensitivity analysis of a spatially distributed model. When field data is lacking to carefully estimate covariance parameters, at least the *a-priori* impact of giving wrong values to these parameters will be known: over-estimating covariance range l or covariance sill σ^2 wil result in over-estimating sensitivity indices of spatial input $Z(u)$ and under-estimating sensitivity indices of scalar inputs X_i. On the contrary, over-estimating covariance nugget η will result in under-estimating sensitivity indices of $Z(u)$.

Conclusion

Variance-based global sensitivity analysis (GSA) was performed on a simple example of a spatially distributed model $Y=M(X,Z)$ with two inputs: a scalar input X and a spatial input $Z(u)$. In order to represent the variability on uncertain spatial input Z(u), it was assumed to be a Gaussian Random Field, and random realizations were generated using geostatistical simulation. These random realizations were used to propagate input uncertainty through model M. Sensitivity indices of model inputs X and $Z(u)$ were estimated with a sampling-based approach. The influence of spatial covariance parameters on GSA results was assessed by estimating sensitivity indices for different sets of covariance range, sill and nugget.

Results show that (1) first order sensitivity index S_Z of spatial input $Z(u)$ is a monotically increasing function of covariance range l (2) first order sensitivity index S_Z of spatial input $Z(u)$ is a monotically increasing function of covariance sill σ^2 (3) first order sensitivity index S_Z of spatial input $Z(u)$ is a monotically decreasing function of covariance nugget η.

These empirical results may be of importance when setting covariance parameters to describe uncertainty in spatial inputs for sensitivity analysis of a spatial model. Yet further research is needed to prove analytically that these properties hold for a large range of point-based models and monotonic covariance functions. Such study may help promoting the use of geostatistical simulation to perform sensitivity analysis of spatially distributed models.

References

AERTS, J. C. J. H., HEUVELINK, G. B. M., GOODCHILD, M. F. (2003): Accounting for spatial uncertainty in optimization with spatial decision support systems. Transactions in GIS, Vol. 7, 211-230

HELTON, J. C., DAVIS, F. J. (2006): Sampling-based methods for uncertainty and sensitivity analysis. Multimedia Environmental Models, Vol. 32, 135-154

IOOSS, B., RIBATET, M. (2009): Global sensitivity analysis of computer models with functional inputs. Reliability Engineering and System Safety, Vol. 94, 1194 - 1204

JOURNEL, A.G., HUIJBREGTS, C.J. (1978): Mining geostatistics. The Blackburn Press

KYRIAKIDIS, P. C. (2005): Sequential spatial simulation using Latin Hypercube Sampling. Geostatistics Banff 2004, 65-74

LILBURNE, L., TARANTOLA, S. (2009): Sensitivity analysis of spatial models. International Journal of Geographical Information Science, Vol. 23, No. 2, 151-168

REFSGAARD, J. C., VAN DER SLUIJS, J. P., HØJBERG, A. L., VANROLLEGHEM, P. A. (2007): Uncertainty in the environmental modelling process: a framework and guidance. Environmental Modelling and Software, Vol. 22, 1543-1556

RUFFO, P., BAZZANA, L., CONSONNI, A., CORRADI, A., SALTELLI, A., TARANTOLA, S. (2006): Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques. Reliability Engineering and System Safety, Vol. 91, 1155-1162

SAINT-GEOURS, N., BAILLY, J.-S., GRELOT, F., LAVERGNE, C. (2010): Is there room to optimise the use of geostatistical simulations for sensitivity analysis of spatially distributed models ? Accuracy 2010 : proceedings of the ninth International Symposium on Spatial Accuracy Assesment *in* Natural Resources and Environmental Sciences, 81-84

SALTELLI, A., RATTO, M., ANDRES, T., CAMPOLONGO, F., CARIBONI, J., GATELLI, D., SAISANA, M., TARANTOLA, S. (2008): Global Sensitivity Analysis - The Primer. Wiley, New-York.

VOLKOVA, E., IOOSS, B., VAN DORPE, F. (2008): Global sensitivity analysis for a numerical model of radionuclide migration from the RRC Kurchatov Institute radwaste disposal site. Stochastic Environmental Research and Risk Assessment, Vol. 22, 17-31

Accuracy 2010 Symposium, July 20-23, Leicester, UK

Comparison of three spatial sensitivity analysis techniques

Nathalie Saint-Geours
AgroParisTech, UMR TETIS,
F-34035, Montpellier,
France
nathalie.saint-geours@teledetection.fr

Linda Lilburne
Landcare Research
PO Box 40, Lincoln 7640, Canterbury
New Zealand
lilburnel@landcareresearch.co.nz

Abstract—**This paper compares the spatial Sobol' sensitivity analysis approach to two other sensitivity analysis techniques on a model with spatially distributed inputs. The comparison is performed on AquiferSim, a model that simulates groundwater flow and nitrate transport from paddock (i.e. field) to aquifer. Some of the input layers have considerable uncertainty. Alternative soil and land-use layers were simulated through Monte Carlo simulation based on expert-derived confusion matrices. Uncertainty of the raster rainfall layer was simulated via geostatistical unconditional simulation of error fields. The three sensitivity techniques are: (1) the spatial Sobol' technique, (2) one-at-a-time (OAT) variation around base sample points, and (3) the Elementary Effects method. The results show that the spatial Sobol' approach gives the best insight on AquiferSim behavior. OAT local variations of inputs around some sample points allow checking of the robustness of model predictions around those points, but give no insight on the relative importance of inputs. The Elementary Effects method shows that land use layer is the most influential input factor, but fails to capture interactions between input factors. The spatial Sobol' approach identifies the land use layer as being the most influential. It shows that strong interactions occur between most of the inputs, explaining 43% of the output variability.**

Keywords: sensitivity analysis, Sobol, Elementary Effects

I. INTRODUCTION

Sensitivity analysis (SA) techniques can be used to study how uncertainty in model inputs influences uncertainty in model predictions (Saltelli et al., 2000). Various techniques are available to perform sensitivity analysis from a set of model evaluations (Helton, 1993; Saltelli et al., 2004). They differ mainly in how the uncertainty of the model inputs is sampled, and in how the sensitivity measures are calculated. Some of these various techniques are suitable for models with spatially distributed inputs, others, like regression-based approaches, are not (Lilburne and Tarantola, 2009).

A spatial SA has been carried out on AquiferSim, a model that simulates groundwater flow and nitrate transport from paddock to aquifer. The AquiferSim model has been designed for analysis of cumulative effects of leaching from agricultural non-point sources at a regional scale on alluvial plains in New Zealand. Some of the inputs for AquiferSim are categorical vector data (soil, land use and climate zone layers); others are continuous raster data (aquifer transmissivity and recharge). These spatial inputs, along with other scalar inputs, are subject to various sources of uncertainty (measurement errors, interpolation errors, missing data, etc.). Sensitivity analysis (SA) is needed to study how these uncertainties influence the variability of AquiferSim predictions, to check the robustness of the model predictions, and to identify the factors that account for most of the model output variability. Three different spatial SA techniques were used: one-at-a-time (OAT) local variations around some base points, the Elementary Effects method, and the variance-based spatial Sobol' method. This paper compares the results of the three methods.

II. METHODS

A. AquiferSim Model

AquiferSim is a steady-state model of groundwater flow and contaminant transport. It has been designed for analysis of cumulative effects of leaching from agricultural non-point sources at a regional scale under a range of land use scenarios (Bidwell et al., 2005). AquiferSim takes various GIS layers as inputs, including maps of land use, soil type and climate zone (categorical vector data) and maps of river recharge and transmissivity of the aquifer (continuous raster data). AquiferSim models the three-dimensional concentration of nitrate in the groundwater. One output is vertical 1-D profile graphs of nitrate concentration below point locations chosen by the user.

B. Target Function of Study

AquiferSim was used to simulate a region of 3500 km^2 with a cell size of 100 m. The objective function Y for the sensitivity analysis was chosen as the average of the maximum nitrate concentrations (mg/L) below a set of 10 points distributed evenly around the study area.

C. Simulating Uncertainty of the Model Inputs

Most SA methods need the model to be evaluated on multiple points of the space of the input factors. Six input factors were considered in this analysis, the first is a lookup table of nitrate and drainage values, the other five are spatial inputs: land use, soil, climate, transmissivity of the aquifer, and river recharge maps. Uncertainty for each input factor has been simulated using a specific model (Table 1).

Accuracy 2010 Symposium, July 20-23, Leicester, UK

The nitrate and drainage lookup table is derived from outputs from a paddock-scale model. The table lists nitrate concentration and vertical drainage values per unit land area for every possible combination of land use type, soil type and rainfall class. In addition to the original table, two alternatives were generated by applying global –5% and +5% variations to the nitrate concentration values only, leaving the drainage values unchanged.

Soil and land use are categorical vector layers. Derived from a soil survey, and a farmer questionnaire respectively, they identify 7 key soil and 25 land-use types over the study area. Uncertainty was modeled using two expert-derived confusion matrices. Nine alternative soil layer realizations and nine alternative land-use layer realizations were generated through random simulation of the soil and land use uncertainty by applying probabilities derived from the confusion matrices.

The climate map comes from an interpolated layer of mean annual rainfall. Three rainfall layer realizations were simulated by a geostatistic method, adding a stochastic error field to the original data. This error field was generated using the circulant embedding technique (Chan and Wood, 1997), as a zero mean stationary Gaussian field with some spatial structure, described by a variance and a semi-variogram model.

Maps of river recharge and transmissivity of the aquifer come from calibrated runs of the Modflow groundwater model (Harbaugh et al., 2000) where hydraulic conductivity has been varied to match groundwater levels. Three water-budget scenarios, based on expert opinion of the ratio of river recharge to land-surface recharge and the proportion of ocean discharge, were simulated in Modflow, resulting in paired river recharge and transmissivity layers for each of the three scenarios.

D. *Multiple Evaluations of the AquiferSim Model*

One run of the AquiferSim model takes several minutes to complete. In order to run multiple evaluations of the model, a parallel computing framework has been designed in Python to use a cluster of 10 powerful nodes, thus reducing computational time considerably. In this configuration, a Monte Carlo type approach to sensitivity analysis is practical.

TABLE I. INPUT VARIABLES OF AQUIFERSIM MODEL

Variable	Model of Uncertainty
Soil map	9 alternative layers derived from a Monte Carlo simulation using a confusion matrix
Land-use map	9 alternative layers derived from a Monte Carlo simulation using a confusion matrix
Nirate/drainage table	3 alternative tables : original table, –5%, +5%
Climate map	3 alternative layers of rainfall zone from geostatistical simulation
River recharge map and aquifer transmissivity map	3 sets of linked pairs of maps – based on three water-budget scenarios with higher/lower proportions of ocean discharge

E. *Sensitivity Analysis*

Three different approaches have been considered to perform spatial sensitivity analysis of the AquiferSim model: OAT local variations around some base points, the Elementary Effects method (Morris, 1991; Saltelli et al., 2008) and the variance-based spatial Sobol' method (Lilburne and Tarantola, 2009).

The same principle was used to handle spatially distributed inputs in the three approaches: the i^{th} spatial input factor X_i is described as a discrete uniform distribution in $\{0 ; 1 ; .. ; n_i - 1\}$ where n_i represents the total number of realizations for the i^{th} input factor. Each discrete level is associated with a single realization of the spatial input, generated from a specific spatial uncertainty model. One must note that SA techniques based on regression and the Fourier-based techniques cannot be applied in this case, as the order in which spatial realizations are ranked has no meaning.

The three different approaches considered differ on various points, including design of the sampling of input factors, measures of sensitivity, and computational cost (e.g. number of model simulation runs). The first two are mainly local, quantifying the variation of the model output due to small variations in the uncertain model inputs around a base point. The spatial Sobol' method is global, exploring more of the multidimensional input space.

1) *OAT local variations around some base points*

A first rough approach is to analyze the local influence of the variation of each input factor X_i around a few random base points $P^j = \left(X_1^j, X_2^j, ..., X_k^j\right)$ where X_i^j is the j^{th} realization (layer) of input factor X_i for k input factors. The following steps were followed for $j = 0, 1, 2$:

- select the realizations for the base point P_*^j from the space of the input factors
- for each input factor X_i:
  - calculate the model output for each spatial realization of X_i, with the other input factors (layers) remaining unchanged from the base point.
  - compute S_i^j, the range of model output Y for this set of runs (there are not enough runs to calculate variance).

S_i^j can be seen as a sensitivity measure of input factor X_i around base point P^j.

The total number of model evaluations is $C = p \times \sum n_i$ where n_i is the number of possible values for input factor X_i and p the number of base points studied. Here $C = 81$.

2) *Elementary Effects method*

The Elementary Effects method is a technique for sensitivity analysis that overcomes some of the limitations of the previous technique, whose results depend heavily on a single base point. It also belongs to the class of OAT

sampling designs. In this study r = 100 different trajectories in the input space were generated from randomly selected base points. Each trajectory has (k+1) points, where k is the number of input factors, and has two key properties: two consecutive points differ in only one input factor, and each input factor has been varied exactly once in the trajectory. This approach can be adapted for the spatial case by using another realization to vary each each input factor (layer) in turn. From each trajectory t, the elementary effect EE_i^t of input factor X_i can be computed as:

$$EE_i^t = Y\left(X_1,...,X_{i-1},X_i^{\sim j},X_{i+1},...,X_k\right) - Y\left(X_1,X_2,...X_i^j...,X_k\right)$$

where $X_i^{\sim j}$ is a different layer realization to X_i^j:

For each input factor X_i, a sensitivity measure μ_i^* can be calculated. It is defined as the mean of the absolute values of the computed elementary effects EE_i^t over the r different trajectories:

$$\mu_i^* = \frac{1}{r}\sum_{t=1}^{r}\left|EE_i^t\right|$$

The total number of model evaluations is $C = r\times(k+1)$. Here C = 600.

3) Variance-based spatial Sobol' method

The spatial Sobol' method is a generalization of the methods of Sobol' (1993) and Saltelli et al. (2000) to spatially dependent models (Lilburne and Tarantola, 2009). Briefly, it is based on the decomposition of the output variance in conditional variances. It uses two quasi-random sample matrices to explore the space of the input factors. Each spatial factor (layer) X_i is sampled from a discrete uniform distribution in {0 ; … ; n_i−1 }. Numerous permutations between the two sample matrices allow the computation of Sobol' first-order and total-order sensitivity indices S_i and S_{Ti}, for each input factor. Empirical 90% confidence intervals for the Sobol' indices have been computed from the same set of model evaluations using 100 bootstrap replicas.

The computational cost of the spatial Sobol' method is considerably higher than that of the other two techniques: the number of model evaluations needed is $C = 2\times N\times(k+1)$ where N is the size of the quasi-random sample matrices (here N = 512 rows) and k is the number of input factors (here k = 5). Here C = 6144.

III. RESULTS

The results from OAT local variations of model inputs around some base points show that this type of analysis is heavily dependent on the selected base point. Local sensitivity measures of the input factors are different for each base point and no general ranking can be inferred (Table 2).

The Elementary Effects method gives more valuable information: land use layer is the most influential input (Table 3), while no input can clearly be discarded as having little influence.

The spatial Sobol' method leads to similar conclusions: the land use layer also appears to be the most influential input (Table 4). The sum of first-order effects is only 57%, showing that there are significant interactions between the inputs. The results also show that all of the inputs have some interactions with the other inputs.

IV. DISCUSSION

Associating randomly generated map and tabular realizations to scalar values sampled from discrete uniform distributions makes it possible to use various sensitivity analysis techniques on the spatial AquiferSim model.

Complex descriptions of spatial uncertainty can be used to generate map realizations, including geostatistical techniques and simulation from confusion matrices.

OAT variations of each input factor around base sample points depends heavily on the selected sample point and it fails to identify the factors that account for most of the model output variability. Yet it is a relatively costless method to implement, and provides valuable information on the local behavior of the model around some specific points.

The most influential input, which is the land use layer, was clearly identified by the Elementary Effects method, with a quite low computational cost (Table 5). Yet this technique does not explore the whole space of the input factors, and some specific interaction effects could have been missed by the analysis. Moreover, the adaptation of this method to a spatial context raises some difficulties: specifically, as the integer index associated with each spatial realization has no meaning, the computation of sensitivity measures μ and σ (Morris, 1991) is not possible, therefore limiting the usefulness of the Elementary Effects method.

TABLE II. RESULTS OF OAT LOCAL VARIATION ABOUT BASE SAMPLE POINTS

	Base Point 1		Base Point 2		Base Point 3	
	Rank	*Range of Y*	*Rank*	*Range of Y*	*Rank*	*Range of Y*
Land use	1	1.73	3	0.87	2	0.81
River recharge & transmissivity	2	1.23	2	1.23	5	0.65
Soil	3	1.03	1	1.71	3	0.75
Nitrate/drainage table	4	0.73	4	0.67	4	0.70
Climate	5	0.09	5	0.45	1	1.29

Accuracy 2010 Symposium, July 20-23, Leicester, UK

TABLE III. RESULTS OF ELEMENTARY EFFECTS SENSITIVITY ANALYSIS

	Rank	μ*
Land use	1	68.57
River recharge & transmissivity	2	57.04
Climate	3	56.74
Nitrate/drainage table	5	49.51
Soil	6	44.66

TABLE IV. RESULTS OF SPATIAL SOBOL' SENSITIVITY ANALYSIS

	First-Order Sensitivity Index (90% Confidence Interval)	Total-Order Sensitivity Index (90% Confidence Interval)
Land use	15% – 28%	49% – 60%
River recharge & transmissivity	6% – 19%	22% – 29%
Nitrate/drainage table	5% – 19%	12% – 18%
Climate	0.5% – 12%	32% – 44%
Soil	0% – 11%	26% – 31%

The Spatial Sobol' method has none of those shortcomings, but at the price of a much higher computational cost (in this case nearly a week). It ensures that the entire model input space is explored. Land use was identified by the Spatial Sobol' method as being the input that accounts for most of the output variability. This method gives valuable insight on the interactions between input factors, which account for 43% of the output variance. Finally it shows that no inputs can be considered as negligible. With a number of sample rows of $N = 512$, confidence intervals on the sensitivity indices are small enough to draw firm conclusions.

V. CONCLUSION

By treating spatial layers as unique random variables, various sensitivity analysis techniques have been applied to the AquiferSim spatial model. Specific and complex descriptions of spatial uncertainty were used for each spatial input, to generate sets of random realizations. OAT local variations of inputs around some sample points allow the robustness of model predictions to be checked around those points, but give no insight on the relative importance of inputs. The Elementary Effects method shows that the land use layer is the most influential input factor: it would be worthwhile to obtain better land use information in order to reduce the output variability. The spatial Sobol' method gives the most valuable information: it confirms that the land use layer is the most influential input; it also shows that strong interactions occur between most of the inputs, explaining around 40% of the output variability.

TABLE V. SUMMARY COMPARISON BETWEEN THE THREE TECHNIQUES

	Model Evaluations	Ranking of Inputs	Inter-actions
Local OAT variations around base point	81	No	No
Elementary Effects	600	Yes	No
Spatial Sobol'	6144	Yes	Yes

ACKNOWLEDGEMENTS

We thank David Pairman, Stephen McNeill and Christine Bezar for their helpful comments. Nathalie's visit to New Zealand was supported by the research program "Risque, Décision, Territoire" (EPI) funded by MEEDEM, France. Funding for this work was provided by Foundation for Research Science and Technology (NZ), contract CO9X0306.

REFERENCES

Bidwell, V.J., Lilburne, L.R. and Good, J.M. (2005). Strategy for developing GIS-based tools for management of the effects on groundwater of nitrate leaching from agricultural land use. In: A. Zerger and R.M. Argent (Eds) *MODSIM 2005 International Congress on Modelling and Simulation.* Melbourne, Australia: Modelling and Simulation Society of Australia and New Zealand.

Chan, G. and Wood, A.T.A. (1997). An algorithm for stimulating stationary Gaussian random fields. *Applied Statistics.* 46, 171–181.

Harbaugh, A.W., Banta, E.R., Hill, M.C. and Mcdonald, M.G. (2000). *MODFLOW-2000, the U.S. Geological Survey modular ground-water model: User guide to modularization concepts and the ground-water flow process.* U.S. Geological Survey.

Helton, J.C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety.* 42, 327–367.

Lilburne, L. & Tarantola, S. (2009). Sensitivity analysis of spatial models. *International Journal of Geographical Information Science.* 23, 151–168.

Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics.* 33, 161–174.

Saltelli, A., Chan, K. and Scott, M. (Eds.) (2000). *Sensitivity analysis.* Chichester, UK: John Wiley.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004). *Sensitivity analysis in practice. A guide to assessing scientific models.* Chichester, UK: John Wiley.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global sensitivity analysis. The primer.* Chichester, UK: John Wiley.

Sobol', I.M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment.* 1, 407–414.

ANALYSE DE SENSIBILITÉ GLOBALE D'UN MODÈLE D'ÉVALUATION ÉCONOMIQUE DU RISQUE D'INONDATION

Nathalie Saint-Geours[1,2], Christian Lavergne[1], Jean-Stéphane Bailly [2] & Frédéric Grelot[3]

[1] *UMR I3M, Université Montpellier 2, France*
[2] *AgroParisTech, UMR TETIS, F-34093, Montpellier, France*
[3] *Cemagref, UMR G-EAU, F-34093, Montpellier, France*

Mots-clés : analyse de sensibilité globale, indices de Sobol, spatial, inondation
Keywords: global sensitivity analysis, variance-based, Sobol' indices, spatial, flood

Résumé[1] : L'analyse de sensibilité globale peine à se développer dans le champ de la modélisation environnementale. Dans sa formulation initiale, elle est limitée à l'étude de modèles $Y = f(X_1, \ldots, X_p)$ où les variables d'entrée X_j et la sortie Y sont scalaires, alors que nombre de modèles environnementaux incluent une dimension spatiale marquée, soit qu'ils fassent appel à des cartes comme variables d'entrée, soit que leurs sorties soient distribuées spatialement. Au travers d'une étude de cas détaillée, nous présentons dans cet article une extension de l'analyse de sensibilité globale à l'étude de modèles spatialisés. Le modèle étudié, nommé ACB-DE, est un outil d'évaluation économique du risque d'inondation. La réalisation de cartes d'indices de sensibilité permet d'étudier les sorties spatialisées du modèle ACB-DE et de rendre compte de la variabilité spatiale des indices de Sobol. L'influence relative des variables d'entrée à différentes échelles d'étude est analysée par la réalisation de cartes d'indices de sensibilité de résolution croissante. L'analyse réalisée permet d'identifier les variables d'entrée incertaines qui expliquent la plus grande part de la variabilité de l'indicateur économique fourni par le modèle ACB-DE ; elle apporte un éclairage nouveau sur le choix de l'échelle adéquate de représentation spatialisée de cet indicateur selon la précision des variables d'entrée.

Abstract: Variance-based Sobol' global sensitivity analysis (GSA) was initially designed for the study of models with scalar inputs and outputs, while many models in the environmental field are spatially explicit. As a result, GSA is not a common practise in environmental modelling. In this paper we describe a detailed case study where GSA is performed on a spatially dependent model for flood risk economic assessment on the Orb valley (southeast France). The realisations of random input maps can be generated by any method including geostatistical simulation techniques, allowing for spatial structure and auto-correlation to be taken into account. The estimation of sensitivity indices on ACB-DE spatial outputs makes it possible to produce maps of sensitivity indices. These maps describe the spatial variability of Sobol' indices. Sensitivity maps of different resolutions are then compared to discuss the relative influence of uncertain input factors at different scales.

[1]Un article présentant ce travail plus en détails a été soumis au Journal de la SFdS pour un numéro spécial *Analyse de sensibilité*.

Introduction

L'analyse de sensibilité d'un modèle $Y = f(X_1, \ldots, X_p)$ vise à quantifier l'impact de l'incertitude pesant sur les p variables d'entrée X_i du modèle sur la variabilité de sa sortie Y. L'analyse de sensibilité globale basée sur la variance proposée par Sobol (1993) [7] peut être utilisée pour étudier des codes numériques de type *boîte noire* peu coûteux en temps de calcul. Elle s'appuie sur la décomposition de la variance de Y en variances conditionnelles ; on y définit des indices de sensibilité S_i (parfois appelés indices de Sobol), compris entre 0 et 1, qui traduisent la part de chaque variable d'entrée X_i dans la variance de Y :

$$S_i = \frac{\mathrm{Var}\left[E(Y \mid X_i)\right]}{\mathrm{Var}(Y)}$$

Dans sa formulation initiale, l'analyse de sensibilité globale basée sur la variance est limitée à l'étude de modèles où les variables d'entrée X_i sont scalaires, tout comme la sortie Y. Or, dans le vaste champ de la recherche en environnement, notamment dans l'étude des risques naturels, nombre de modèles incluent une dimension spatiale marquée, soit qu'ils fassent appel à des *cartes* comme variables d'entrée, soit que leurs sorties soient spatialisées (cartes de risque par exemple). De ce fait, l'analyse de sensibilité peine à se développer dans ces champs thématiques, comme l'a montré Delgado (2004) [2]. Nous présentons ici une application de l'analyse de sensibilité globale basée sur la variance à un modèle d'évaluation économique du risque d'inondation (modèle ACB-DE), dont les variables d'entrée et les sorties sont distribuées spatialement.

1 Modèle ACB-DE

Le modèle ACB-DE est un outil d'évaluation économique du risque d'inondation développé par Erdlenbruch et al. (2008) [3]. Il vise à caractériser l'exposition d'un territoire au risque d'inondation, au moyen d'un indicateur spatialisé décrivant les dommages liés aux crues potentielles. Les données utilisées dans ce travail sont issues d'une étude de cas sur la basse vallée de l'Orb, dans l'Hérault (superficie de 100 km^2).

Le modèle ACB-DE fait appel à des modules de natures différentes (hydrologique, hydraulique, économique), et combine six variables d'entrée notées X_0 à X_5 : X_2, X_3 et X_5 appartiennent respectivement à $\mathbb{R}^5$, $\mathbb{R}^{20}$ et $\mathbb{R}$. Les trois autres variables d'entrée sont distribuées spatialement et sont représentées sous forme de couches dans un logiciel de Système d'Information Géographique : X_0 est une carte décrivant l'étendue des plaines d'inondation, X_1 est un Modèle Numérique de Terrain et X_2 une carte d'occupation du sol. La résolution horizontale de ces cartes est de 5 m.

La sortie Y du modèle ACB-DE est une carte donnant en tout point u du territoire les dommages évités $Y(u)$ en ce point. Sa résolution horizontale est de 5 m. A partir de cette sortie Y, on crée quatre autres cartes plus grossières notées $Y^{(200)}$, $Y^{(400)}$, $Y^{(800)}$ et $Y^{(1600)}$, de résolution respectives $s = 200$, 400, 800 et 1600 m, à destination des gestionnaires de

bassins versants (Figure 1). Plus précisement, pour une résolution s donnée, $Y^{(s)}$ désigne la carte donnant en toute cellule C de côté s la somme Y_C des dommages évités sur cette cellule : $Y_C = \int_{u \in C} Y(u)\, du$.

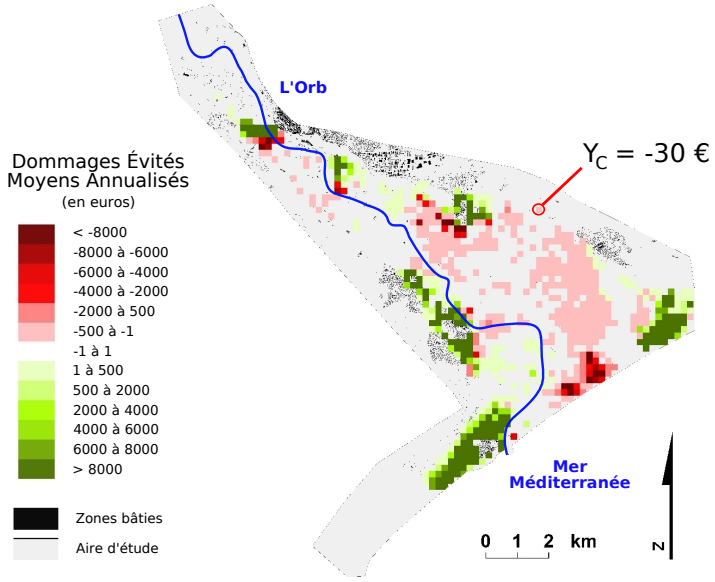

Figure 1: $Y^{(200)}$, carte des Dommages Évités Moyens Annualisés sur une grille régulière constituée de cellules de côté $s = 200$ m. En chaque cellule C est représentée la somme Y_C des dommages évités sur cette cellule.

2 Analyse de sensibilité globale spatialisée

Plusieurs travaux ont déjà apporté des éléments de réponse aux problèmes posés par l'analyse de sensibilité globale d'un modèle spatialisé ; ils se sont intéressés à l'intégration dans l'analyse de variables d'entrée spatialisées et à l'étude de sorties spatialisées. On renvoie à Lilburne et Tarantola (2009) [4] pour une liste de références sur ce sujet.

On propose ici de combiner deux approches pour mener à bien l'analyse de sensibilité du modèle ACB-DE : on associe le traitement des variables d'entrée spatialisées selon la méthode de Lilburne et Tarantola (2009) [4] et l'analyse de sorties spatialisées via des cartes de sensibilité comme proposé par Marrel et al (2009) [5]. On explore de plus un point non abordé à notre connaissance, qui est l'impact sur la valeur des indices de Sobol de la résolution s à laquelle est représentée la variable de sortie Y spatialisée.

Description des incertitudes Dans un premier temps, on spécifie l'incertitude pesant sur chaque variable d'entrée (Table 1). Des distributions de probabilité sont identifiées pour les variables X_2, X_3 et X_5. Afin de rendre compte de l'incertitude sur les deux variables spatialisées (X_1, carte d'occupation du sol et X_4, Modèle Numérique de Terrain), deux jeux de $n_1 = 1000$ et $n_4 = 100$ réalisations aléatoires de ces variables ont été générés : un champ d'erreur Gaussien présentant une structure spatiale de covariance modélisée par un variogramme est ajouté au Modèle Numérique de Terrain initial selon un algorithme de simulation Gaussienne séquentielle proposé par Castrignano (2006) [1] ; des confusions probabilistes entre classes sont introduites dans la carte d'occupation du sol.

	Variable d'entrée	**Spécification de l'incertitude**
X_0	Carte de la plaine d'inondation	Incertitudes non prises en compte.
X_1	Carte d'occupation du sol	Matrice de confusion
$X_2 \in \mathbb{R}^5$	Périodes	Composantes indépendantes de lois uniformes.
$X_3 \in \mathbb{R}^{20}$	Endommagement	Composantes indépendantes de loi uniforme.
X_4	Modèle Numérique de Terrain	Modélisation et simulation géostatistique.
$X_5 \in \mathbb{R}$	C_∞	Distribution triangulaire symétrique.

Table 1: Spécification de l'incertitude sur les variables d'entrée du modèle ACB-DE.

Estimation des indices de sensibilité Dans un second temps, un échantillon de taille $N = 28\,672$ est constitué afin d'explorer l'espace des variables incertaines X_1 à X_5, puis le modèle ACB-DE est simulé en chaque point de cet échantillon. La variable spatialisée X_1 (resp. X_4) est intégrée dans cet échantillon en associant un entier compris entre 1 et n_1 (resp. n_4) à chaque réalisation aléatoire préalablement générée de cette variable, ces réalisations étant considérées équiprobables. On estime alors les indices de sensibilité de premier ordre de chacune des variables X_j par rapport à Y, notés $S_j(Y)$, selon les expressions proposées par Saltelli et al (2008) [6]. Des intervalles de confiance à 90 % sont estimés pour chacun des indices de sensibilité par bootstrap (nombre d'échantillons bootstrap $n_{\text{boot}} = 100$).

Réalisation de cartes d'indices de sensibilité Pour chacune des quatre sorties spatialisées $\mathrm{Y}^{(200)}$ à $\mathrm{Y}^{(1600)}$, on établit des *cartes de sensibilité* pour chaque variable d'entrée X_1 à X_5. Ces cartes sont définies sur la même grille régulière que la sortie $\mathrm{Y}^{(s)}$ considérée. La valeur de la carte de sensibilité de la variable X_j en une cellule C est égale à l'indice de sensibilité $S_j(Y_C)$ de cette variable par rapport à la somme Y_C des dommages évités sur cette cellule. On obtient ainsi un jeu de 20 cartes de sensibilité qui traduisent l'influence de l'incertitude des différentes variables d'entrée X_1 à X_5 sur la variabilité de la carte Y des dommages évités à différentes résolutions spatiales.

3 Résultats

La Figure 2 présente les *cartes d'indices de sensibilité* de trois variables d'entrée (X_1, X_2 et X_3) par rapport à la carte Y des dommages évités à deux résolutions différentes, $s = 200$ et $s = 1600$ m. Sur chacune de ces cartes, on observe une variabilité spatiale des indices de sensibilité. À une résolution s donnée, l'influence d'une variable incertaine X_i sur la variance de la sortie Y_C (somme des dommages évités sur la cellule C) varie d'une zone à l'autre du territoire. On peut identifier deux éléments de structuration spatiale majeurs : les effets de bord et la présence de villes (bâti urbain).

Par ailleurs, on observe que les cartes de sensibilité par rapport à la sortie $\mathrm{Y}^{(200)}$ sont différentes des cartes réalisées par rapport à la sortie $\mathrm{Y}^{(1600)}$. Pour permettre une comparaison entre les cartes d'indices de sensibilité aux différentes échelles, on a calculé pour chaque variable d'entrée X_j et chaque résolution $s = 200, 400, 800$ et 1600 m la valeur moyenne $\mu_j^{(s)}$ de l'indice de sensibilité $S_j(Y_C)$ sur l'ensemble des cellules C de la grille.

On montre que pour chaque variable d'entrée X_j, l'indice moyen $\mu_j^{(s)}$ dépend de manière monotone de la résolution s. L'indice de sensibilité des variables *distribuées spatialement* (ici X_1 et X_4) diminue à mesure qu'augmente le côté s des cellules sur lesquelles sont calculés les dommages évités. Un effet de compensation des incertitudes en est responsable : les incertitudes qui affectent les variables spatialisées sont *locales* ; lorsqu'une large zone est considérée, ces incertitudes locales se compensent et leur impact relatif par rapport aux autres variables diminue. Symétriquement, l'indice de sensibilité des variables *non distribuées spatialement* (ici X_2, X_3 et X_5) augmente avec le côté s de la cellule C sur laquelle on calcule les dommages évités. Selon la résolution s à laquelle on représente le champ Y des dommages évités, les variables d'entrée qui expliquent sa variabilité ne sont donc pas les mêmes.

Conclusion

L'approche proposée est applicable à un large éventail de modèles spatialisés, à la condition qu'ils soient peu coûteux en temps de calcul. Elle offre de nouvelles perspectives dans la compréhension des liens entre effets d'échelle, incertitudes et sensibilité, questions qui sont propres à ce type de modèles. Des travaux complémentaires sont à mener pour préciser les effets de biais induits par la méthode de représentation de l'incertitude sur les variables spatialisées, pour optimiser l'échantillonnage de réalisations aléatoires de ces variables, ou encore pour mieux comprendre et prédire la structure spatiale des cartes de sensibilité.

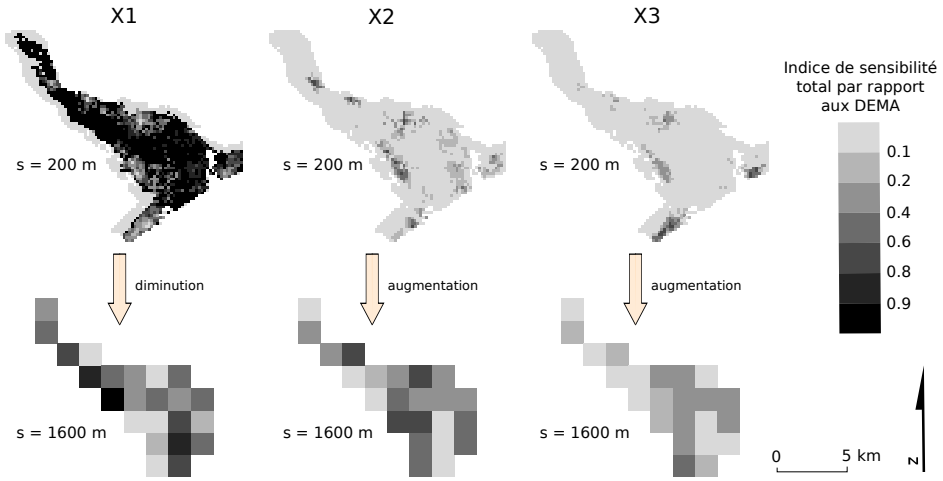

Figure 2: Cartes de sensibilité. Pour chacune des trois variables X_1 à X_3, la carte du haut (resp. du bas) présente la *carte de sensibilité* de la variable X_j par rapport à la sortie spatialisée $Y^{(200)}$ (resp. $Y^{(1600)}$) : sur chaque cellule C de côté $s = 200$ m (resp. 1600 m), on représente $S_j(Y_C)$, indice de sensibilité de la variable X_j par rapport à la somme Y_C des dommages évités sur la cellule.

Bibliographie

[1] Castrignano, A. (2006) Accuracy assessment of Digital Elevation Model using stochastic simulation. *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*.

[2] Delgado, M. et Sendra, J. (2004) Sensitivity analysis in multicriteria spatial decision-making: a review. *Human and Ecological Risk Assessment*, 10(6), 1173-1187.

[3] Erdlenbruch, K., Gilbert, E., Grelot, F. et Lescoulier, C. (2008) Une analyse coût-bénéfice spatialisée de la protection contre les inondations : Application de la méthode des dommages évités à la basse vallée de l'Orb. *Ingénieries Eau-Agriculture-Territoires*, 53, 3-20.

[4] Lilburne, L. et Tarantola, S. (2009) Sensitivity analysis of spatial models. *International Journal of Geographical Information Science*, 23(2), 151-168.

[5] Marrel, A., Iooss, B., Jullien, M., Laurent, B. and Volkova, E. Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, in press

[6] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. et Tarantola, S. (2008) *Global Sensitivity Analysis - The Primer*, Wiley, Ltd.

[7] Sobol', I. M. (1993) Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1, 407-414.