

Département de géomatique appliquée
Faculté des lettres et sciences humaines
Université de Sherbrooke

**Approche géomatique de la variabilité spatio-temporelle de la contamination
microbienne des eaux récréatives**

Francine Nzang Essono

Thèse présentée pour l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en télédétection

Mars, 2017

© **Francine Nzang Essono, 2017**

Cette thèse a été évaluée par un jury composé des personnes suivantes :

M. Richard Fournier, directeur de recherche, Université de Sherbrooke, Sherbrooke, Canada

M. Pascal Michel, Agence de la santé publique du Canada, Saint-Hyacinthe, Canada

M. Goze Bertin Béné, examinateur interne, Université de Sherbrooke, Sherbrooke, Canada

M. Karem Chokmani, examinateur externe, Institut national de la recherche scientifique, Centre Eau Terre Environnement, Québec, Canada

M. Robert Lecomte, examinateur externe, Université de Sherbrooke, Sherbrooke, Canada

La présidence a été assurée par Madame Christine Hudon, doyenne de la Faculté des lettres et sciences humaines de l'université de Sherbrooke.

La soutenance publique s'est tenue à l'Université de Sherbrooke en Novembre 2016

Résumé

L'objectif général de cette thèse est de caractériser la dynamique des transferts des bactéries fécales à l'aide d'une modélisation spatio-temporelle, à l'échelle du bassin versant (BV) dans une région agricole et à l'échelle événementielle. Ce projet vise à mieux comprendre l'influence des processus hydrologiques, les facteurs environnementaux et temporels impliqués dans l'explication des épisodes de contamination microbienne des eaux récréatives.

Premièrement, un modèle bayésien hiérarchique a été développé pour quantifier et cartographier les niveaux de probabilité des eaux à être contaminées par des effluents agricoles, sur la base des données spectrales et des variables géomorphologiques. Par cette méthode, nous avons pu calculer les relations pondérées entre les concentrations d'*Escherichia coli* et la distribution de l'ensemble des paramètres agro-pédo-climatiques qui régissent sa propagation. Les résultats ont montré que le modèle bayésien développé peut être utilisé en mode prédictif de la contamination microbienne des eaux récréatives. Ce modèle avec un taux de succès de 71 % a mis en évidence le rôle significatif joué par la pluie qui est la cause principale du transport des polluants.

Deuxièmement, le modèle bayésien a fait l'objet d'une analyse de sensibilité liée aux paramètres spatiaux, en utilisant les indices de Sobol. Cette démarche a permis (i) la quantification des incertitudes sur les variables pédologiques, d'occupation du sol et de la distance et (2) la propagation de ces incertitudes dans le modèle probabiliste c'est-à-dire le calcul de l'erreur induite dans la sortie par les incertitudes des entrées spatiales. Enfin, une analyse de sensibilité des simulations aux différentes sources d'incertitude a été effectuée pour évaluer la contribution de chaque facteur sur l'incertitude globale en prenant en compte leurs interactions. Il apparaît que sur l'ensemble des scénarios, l'incertitude de la contamination microbienne dépend directement de la variabilité des sols argileux. Les indices de premier ordre de l'analyse de Sobol ont montré que parmi les facteurs les plus susceptibles d'influer la contamination microbienne, la superficie des zones agricoles est le premier facteur important dans l'évaluation du taux de coliformes. C'est donc sur ce paramètre que l'attention devra se porter dans le contexte de prévision d'une contamination microbienne. Ensuite, la deuxième variable la plus importante est la zone urbaine avec des parts de sensibilité d'environ 30 %. Par ailleurs, les estimations des indices totaux sont meilleures que celles des indices de premier ordre, ce qui

signifie que l'impact des interactions paramétriques est nettement significatif pour la modélisation de la contamination microbienne

Enfin, troisièmement, nous proposons de mettre en œuvre une modélisation de la variabilité temporelle de la contamination microbiologique du bassin versant du lac Massawippi, à partir du modèle AVSWAT. Il s'agit d'une modélisation couplant les composantes temporelles et spatiales qui caractérisent la dynamique des coliformes. La synthèse des principaux résultats démontrent que les concentrations de coliformes dans différents sous-bassins versants se révèlent influencées par l'intensité de pluie. La recherche a également permis de conclure que les meilleures performances en calage sont obtenues au niveau de l'optimisation multi-objective. Les résultats de ces travaux ouvrent des perspectives encourageantes sur le plan opérationnel en fournissant une compréhension globale de la dynamique de la contamination microbienne des eaux de surface.

Mots clefs : Contamination microbienne, eaux récréatives, approche probabiliste, indices de Sobol; AVSWAT, cartographie, modélisation spatio-temporelle.

Abstract

The aim of this study was to predict water faecal contamination from a bayesian probabilistic model, on a watershed scale in a farming area and on a factual scale. This project aims to better understand the influence of hydrological, environmental and temporal factors involved in the explanation of microbial contamination episodes of recreational waters.

First, a bayesian probabilistic model: Weight of Evidence was developed to identify and map the probability of water levels to be contaminated by agricultural effluents, on the basis of spectrals data and geomorphologic variables. By this method, we were able to calculate weighted relationships between concentrations of *Escherichia coli* and distribution of key agronomic, pedologic and climatic parameters that influence the spread of these microorganisms. The results showed that the Bayesian model that was developed can be used as a prediction of microbial contamination of recreational waters. This model, with a success rate of 71%, highlighted the significant role played by the rain, which is the main cause of pollution transport.

Secondly, the Bayesian probabilistic model has been the subject of a sensitivity analysis related to spatial parameters, using Sobol indications. This allowed (1) quantification of uncertainties on soil variables, land use and distance and (2) the spread of these uncertainties in the probabilistic model that is to say, the calculation of induced error in the output by the uncertainties of spatial inputs. Lastly, simulation sensitivity analysis to the various sources of uncertainty was performed to assess the contribution of each factor on the overall uncertainty taking into account their interactions. It appears that of all the scenarios, the uncertainty of the microbial contamination is directly dependent on the variability of clay soils. Sobol prime indications analysis showed that among the most likely to influence the microbial factors, the area of farmland is the first important factor in assessing the coliforms. Importance must be given on this parameter in the context of preparation for microbial contamination. Then, the second most important variable is the urban area with sensitivity shares of approximately 30%. Furthermore, estimates of the total indications are better than those of the first order, which means that the impact of parametric interaction is clearly significant for the modeling of microbial contamination.

Thirdly, we propose to implement a temporal variability model of microbiological contamination on the watershed of Lake Massawippi, based on the AVSWAT model. This is a

model that couples the temporal and spatial components that characterize the dynamics of coliforms. The synthesis of the main results shows that concentrations of *Escherichia coli* in different sub-watersheds are influenced by rain intensity. Research also concluded that best performance is obtained by multi-objective optimization. The results of these studies show the prospective of operationally providing a comprehensive understanding of the dynamics of microbial contamination of surface water.

Keywords Faecal contamination; recreational waters; probabilistic modelling; Sobol's sensitivity indices; predictive mapping; AVSWAT; spatio-temporal model.

Tables des matières

| | |
|---|----|
| Chapitre 1. Introduction générale | 1 |
| 1.1. Contexte | 1 |
| 1.2. Objectifs de la thèse | 6 |
| 1.3. Organisation de la thèse | 7 |
| Chapitre 2. Revue de Littérature..... | 8 |
| 2.2. Revue de littérature | 8 |
| 2.2.1. <i>Escherichia coli</i> et sa survie dans les eaux | 8 |
| 2.2.2. La distribution des bactéries dans l'eau | 9 |
| 2.2.3. Transport des bactéries par les matières en suspension dans l'eau | 11 |
| 2.2.4. Corrélation entre <i>E. coli</i> et les matières en suspension..... | 12 |
| 2.2.3. Facteurs influant la contamination des eaux de surface | 13 |
| 2.2.3.1. Les variables géomorphologiques..... | 13 |
| 2.2.3.2. Les variables pédologiques | 14 |
| 2.2.3.3. Les variables climatiques : les précipitations..... | 15 |
| 2.2.3.4. Les variables liées à l'occupation du sol..... | 15 |
| 2.2.4. Méthode de cartographie de la qualité des eaux de surface..... | 16 |
| 2.2.4.1. Les méthodes qualitatives | 16 |
| 2.2.4.2. Les méthodes quantitatives | 17 |
| Chapitre 3. Modélisation probabiliste de la contamination microbienne des eaux récréatives | 21 |
| 3.1. Introduction..... | 21 |
| 3.2. Objectifs..... | 22 |
| 3.3. Principe du modèle <i>PdIP</i> | 23 |
| 3.4. Justification du choix du <i>PdIP</i> | 25 |
| 3.5. Zone d'étude | 26 |
| 3.5.1. Aperçu des bassins hydrographiques du lac Massawippi et de la Yamaska | 27 |
| 3.5.2. Choix des sous-bassins versants de l'étude | 29 |
| 3.6. Matériels | 30 |
| 3.6.1. Base de données de la variable prédictive : la teneur en coliformes fécaux..... | 30 |
| 3.6.2. Base de données nécessaires à l'extraction des variables explicatives | 32 |
| 3.7. Méthodologie | 34 |
| 3.7.1. Choix des variables explicatives..... | 34 |
| 3.7.2. L'évaluation de l'indépendance conditionnelle entre les variables explicatives | 38 |
| 3.7.3. Application de l'approche <i>PdIP</i> | 39 |

| | |
|--|----|
| 3.7.4. Calage | 40 |
| 3.7.5. Validation | 41 |
| 3.8. Résultats et leur interprétation | 42 |
| 3.8.1. Analyse de la dépendance conditionnelle | 42 |
| 3.8.2. Attribution des poids | 45 |
| 3.9. Conclusion partielle | 52 |
| Chapitre 4. Analyse spatiale de la sensibilité du modèle <i>PdIP</i> | 54 |
| 4.1. Introduction..... | 54 |
| 4.2. Objectifs..... | 57 |
| 4.3. Cadre mathématique : principe d'analyse de sensibilité de Sobol..... | 57 |
| 4.4. L'échantillonnage..... | 60 |
| 4.5. L'estimation des indices de sensibilité selon la technique de Sobol | 62 |
| 4.6. Choix de l'approche de SOBOL dans le cadre de la contamination microbienne des eaux récréatives | 63 |
| 4.7. Méthodologie | 64 |
| 4.7.1. Sélection des variables influant la contamination microbienne des eaux..... | 67 |
| 4.7.1.1. Les variables choisies pour l'analyse de sensibilité | 67 |
| 4.7.1.2. Définition des variables : domaine de variation et loi de distribution | 67 |
| 4.7.2. Propagation et analyse des incertitudes | 70 |
| 4.7.3. Analyse de sensibilité..... | 71 |
| 4.8. Résultats et leur interprétation | 71 |
| 4.8.1. Étude de la convergence | 72 |
| 4.8.2. Propagation et analyse des incertitudes | 73 |
| 4.8.2.1. Variables pédologiques | 73 |
| 4.8.2.2. Variables d'occupation du sol : terre agricole et zone urbaine | 79 |
| 4.7.2.3. Proximité des cours d'eau aux zones agricoles | 82 |
| 4.8.3. Analyse de sensibilité..... | 86 |
| 4.9. Conclusion partielle | 91 |
| Chapitre 5. Variabilité temporelle de la contamination microbienne des eaux : une approche géomatique à l'échelle d'un Bassin versant agricole | 93 |
| 5.1. Introduction..... | 93 |
| 5.1.1. Contexte | 93 |
| 5.1.2. Problématique | 94 |
| 5.2. Objectifs..... | 96 |
| 5.3. Description du modèle AVSWAT et son fonctionnement | 97 |

| | |
|--|-----|
| 5.3.1. Structure d'AVSWAT | 97 |
| 5.3.2. Modélisation des processus hydrologiques..... | 98 |
| 5.3.2.1. Le ruissellement de surface..... | 99 |
| 5.3.2.2. Phase aquatique : Transport et mortalité des bactéries | 100 |
| 5.4. Méthodologie | 102 |
| 5.4.1. Choix du modèle AVSWAT | 104 |
| 5.4.2. Implémentation d'AVSWAT sur le bassin versant du lac Massawippi | 104 |
| 5.4.2.1. Délimitation du bassin versant et extraction du réseau hydrographique | 104 |
| 5.4.2.2. Discrétisation spatiale et temporelle | 106 |
| 5.4.2.3. Optimisation des paramètres..... | 106 |
| 5.4.2.4. La sensibilité du jeu optimal à la fonction objectif..... | 107 |
| 5.4.2.5. Évaluation des performances du modèle | 108 |
| 5.4.2.6. Validation du modèle | 110 |
| 5.5. Résultats et interprétation | 110 |
| 5.5.1. Analyse de sensibilité..... | 110 |
| 5.5.2. Analyse des performances suivant les deux stratégies d'optimisation | 117 |
| 5.5.3. Prédications du modèle en fonction des scénarios temporels..... | 123 |
| 5.6. Conclusion partielle | 130 |
| Chapitre 6. Discussion générale..... | 133 |
| 6.1. La modélisation spatiale de la contamination microbienne des eaux récréatives par une approche probabiliste <i>PdIP</i> | 133 |
| 6.2. Analyse spatiale de la sensibilité du modèle <i>PdIP</i> | 136 |
| 6.3. Variabilité temporelle de la contamination microbienne des eaux : une approche géomatique à l'échelle d'un BV agricole..... | 141 |
| Chapitre 7. Conclusion..... | 144 |
| Références bibliographiques | 150 |
| Servais, P. (2005): Microbiological contamination in the Seine river drainage network. <i>Science of the Total Environment</i> , Submitted. | 167 |
| ANNEXES..... | 171 |
| Annexe 1. Les sous-bassins versants de la rivière Yamaska..... | 171 |
| Annexe 2. Les sous-bassins du bassin versant du Lac Massawippi | 172 |
| Annexe 3. Variables morphologiques | 173 |
| Annexe 3.1. Morphologie du bassin versant du lac Massawippi : les formes | 173 |
| Annexe 3.2. Morphologie du bassin versant du lac Massawippi : les pentes..... | 174 |
| Annexe 3.3. Morphologie du bassin versant du lac Massawippi : la direction des pentes..... | 175 |

| | |
|--|-----|
| Annexe 4. Variables pédologiques | 176 |
| Annexe 4.1. Variables pédologiques du bassin versant du lac Massawippi : le drainage..... | 176 |
| Annexe 4.2. Variables pédologiques du bassin versant du lac Massawippi : le drainage..... | 177 |
| Annexe 5. Processus de sélection de la méthode d'interpolation des données de précipitations. | 178 |
| Annexe 6. Répartition des données pluviométriques..... | 180 |
| Annexe 6.1. Répartition des pluies par stations | 180 |
| Annexe 6.2. Comparaisons des résultats d'interpolation..... | 181 |
| Annexe 7. Occupation du sol du bassin versant du lac Massawippi | 182 |
| Annexe 8. Classes de distance | 183 |
| Annexe 9. Les paramètres testés dans AVSWAT | 184 |

Liste des figures

| | |
|---|-----|
| Figure 1. Les bassins versants de l'étude..... | 27 |
| Figure 2. Distribution spatiale des stations d'échantillonnage. | 32 |
| Figure 3. Arbre de régression de la sélection des variables explicatives de contamination microbienne des eaux..... | 44 |
| Figure 4. Probabilité de contamination microbienne sur la Yamaska à partir du modèle D..... | 49 |
| Figure 5. Évaluation explicative et descriptive des modèles | 51 |
| Figure 6. Illustration du principe de LHS: A) le découpage des 5 sous-intervalles selon la fonction de distribution d'une des entrées, en B) représentation des 5 couplets générés (adaptée de Mara, 2008)..... | 61 |
| Figure 7. La méthodologie de l'analyse de sensibilité de <i>PdIP</i> aux variables spatiales..... | 65 |
| Figure 8. Convergence du taux moyen de coliformes en fonction de la taille de l'échantillon : cas des sols argileux et sols sableux..... | 73 |
| Figure 9. Distributions du taux moyen des coliformes dues à la variabilité des sols sableux et en fonction de la pente..... | 75 |
| Figure 10. Distributions du taux moyen de coliformes dues à la variabilité des sols argileux et en fonction de la pente..... | 76 |
| Figure 11. Distributions du taux moyen des coliformes dues à la variation de terre agricole - TA (A) et de zone urbaine - ZU (B)..... | 80 |
| Figure 12. Distributions du taux moyen de coliformes en considérant simultanément les variables pédologiques et d'occupation du sol | 80 |
| Figure 13. Distributions du taux moyen de coliformes dues à la variation de DIST (A) et en fonction de la pente (B)..... | 84 |
| Figure 14. Effet de la variation des variables sur les indices de Sobol..... | 87 |
| Figure 15. Indices de sensibilité du premier ordre et totaux..... | 89 |
| Figure 16. Discrétisation spatiale dans SWAT (adapté de Rollo, 2012) | 98 |
| Figure 17. Méthodologie de la dynamique temporelle de la contamination des eaux..... | 103 |
| Figure 18. Sensibilité du débit à une augmentation des paramètres (+15 %)..... | 111 |
| Figure 19. Sensibilité du débit à la diminution des paramètres (-15%)..... | 112 |
| .. Figure 20. Sensibilité des flux de coliformes à une diminution des paramètres (-15 %) | 113 |
| Figure 21. Sensibilité des flux de coliformes à une augmentation des paramètres (15%) | 114 |
| Figure 22. Résultats de calage mono-objectif (A) et multi-objectif (B) pour les débits..... | 121 |
| Figure 23. Résultats de calage mono-objectif (A) et multi-objectif (B) pour les coliformes | 122 |
| Figure 24. Distribution des flux de coliformes en fonction de hauteur de la pluie..... | 124 |
| Figure 25. Distribution des flux de coliformes en fonction de la durée de la pluie..... | 125 |
| Figure 26. Distributions des flux de coliformes en fonction de l'intensité de la pluie..... | 126 |
| Figure 27. Variation des coefficients de corrélation en fonction de la hauteur de la pluie par sous- bassin versant..... | 127 |
| Figure 28. Variation des coefficients de corrélation en fonction de la durée de la pluie par bassin versant | 128 |
| Figure 29. Variation des coefficients de corrélation en fonction de la durée de la pluie par sous- bassin versant..... | 129 |

Liste des tableaux

| | |
|--|-----|
| Tableau 1. Caractéristiques des bassins versants à l'étude | 30 |
| Tableau 2. Listes des données utilisées pour l'extraction des variables prédictives..... | 33 |
| Tableau 3. Les variables prédictives et les nombres de classes associées | 37 |
| Tableau 4. Corrélation entre les variables prédictives | 42 |
| Tableau 5. Corrélation spatiale entre les variables explicatives et la variable prédictive (bassin versant de la Tomifobia) | 45 |
| Tableau 6. Les valeurs d'ASC des meilleures simulations obtenues par le poids de la preuve.... | 46 |
| Tableau 7. Les valeurs d'ASC des meilleures simulations obtenues pour la transposition | 48 |
| Tableau 8. Valeurs limites et lois de probabilité des variables spatiales sélectionnées | 69 |
| Tableau 9. Valeurs des critères de Khi-deux (test de Kolmogorov-Smirnov) pour identifier la distribution la plus probable de la variable simulée : Taux de coliformes..... | 69 |
| Tableau 10. Moments statistiques du taux de coliformes en considérant les incertitudes du sol sableux. | 77 |
| Tableau 11. Moments statistiques du taux de coliformes en considérant les incertitudes du sol argileux | 77 |
| Tableau 12. Moment statistiques de la distribution de coliformes en considérant la variation de deux variables d'occupation du sol et celle de la pente..... | 81 |
| Tableau 13. Moment statistiques de la distribution de coliformes en considérant les incertitudes des deux variables pédologiques et les deux variables d'occupation du sol | 81 |
| Tableau 14. Moment statistiques de la distribution des coliformes en considérant : la variation de la distance et de la pente | 85 |
| Tableau 15. Résultats de l'optimisation uni-objective avec différentes combinaisons de paramètres | 118 |
| Tableau 16. Résultats de l'optimisation multi-objective avec différentes combinaisons de paramètres | 118 |

Liste des sigles, acronymes et symboles

A : l'espace de départ.

ANOVA : *Analysis of Variance*.

APSC : Agence Publique de Santé du Canada

ASC : l'aire sous la courbe

BDTQ : Base de données topographiques du Québec

C : contraste

Chi² : Chi carré

COGEBY : Conseil de gestion du bassin versant de la Yamaska

COV : Coefficient de variation

E(.) : Espérance mathématique

E. coli : *Eschèrechia coli*

f(.) : Modèle

FAST : *Fourrier Amplitude Sensivity Test*

LHS : *Latin Hypercube Sampling*

IRDA : Institut de Recherche et de Développement en Agroenvironnement

Log : Logarithme naturel

MC : Monte Carlo

MDDELCC : Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques

MF : méthode de filtration sur membrane

MNA : Modèle numérique d'altitude

MRC : Municipalités régionales de comté

NPP : Nombre le plus probable

NSE: Coefficient de Nash-Sutcliffe (*Nash-Sutcliffe estimator*)

O : La cote ou *odds en anglais*

OAT : *One-At-the-Time*

OCDE : Organisation de coopération et de développement économiques

p : Nombre de variables d'entrée

PBIAS : Pourcentage de biais

PdIP : poids de la preuve (*Weight of Evidence*)

SCS-CN : *Soil Conservation Service- Curve Number*

SIG : Systèmes d'informations géographiques

SPOT HRV : Satellite Pour l'Observation de la Terre Haute Résolution Visible

USGS : *United States Geological Survey*

UFC : unités formant des colonies

URH : Unité de réponse hydrologique

V (.) : Variance

W^- : Poids négatif d'une variable

W^+ : Poids positif d'une variable

μ : Moyenne

σ : Écart type

$\bar{-}$: Absence d'une variable

Remerciements

C'est avec la plus grande gratitude que je souhaite remercier tous ceux qui ont contribué à la réalisation de cette thèse de doctorat. Je voudrais tout d'abord remercier mon directeur de recherche, Richard Fournier de m'avoir fait bénéficier de sa grande compétence et de sa rigueur scientifique, me donnant la possibilité de trouver par moi-même mon cheminement personnel. J'ai particulièrement apprécié sa très grande ouverture face à ma condition de mère étudiante. Il m'a accordé sa confiance, une clé incontestable pour se sentir encouragée et mener à bon port ces travaux de recherche. Qu'il trouve ici l'expression de ma profonde reconnaissance. Richard, Merci beaucoup.

Au même titre, je remercie mon co-directeur Pascal Michel pour son encadrement et ses conseils sur la rédaction de ma thèse. Ses commentaires judicieux m'ont permis de raffiner davantage certains points de l'analyse et d'améliorer la méthodologie. Dès le début de ma thèse, il a tenu à m'encadrer de manière très rapprochée et a apporté un soutien financier pour que ces travaux se déroulent dans de bonnes conditions. Merci Pascal.

Je tiens à souligner aussi l'appui de Daniel Lamothe. Grâce à sa large expérience et sa grande connaissance de la méthode du « poids de la preuve » j'ai pu compléter mes travaux de thèse par une approche originale.

Merci à Serge-Olivier et Stéphanie de l'Agence de Santé Publique du Canada; à Environnement Canada, Luc Lemieux, Alexandra Roy et Roberge, aux comités de gestion des bassins versants de la rivière Yamaska et du lac Massawippi qui ont permis d'accéder aux données utilisées dans le cadre de cette étude.

Mes remerciements s'adressent aussi aux membres du jury pour le temps consacré à la lecture de cette thèse, pour les suggestions et les remarques judicieuses.

Un gros merci à mon conjoint, Benoit Landry, qui a assumé tout seul la responsabilité de nos enfants pendant mes absences répétées. Ce projet scientifique a fait grandir notre relation.

Merci à toi mon rayon de soleil, Daven Landry: le premier résultat de cette thèse de doctorat. Enfin merci à toi mon ange Neilson Landry dont les gazouillis ont embelli la fin de cette aventure.

Dédicaces

Je dédie ce travail à :

Ma mère Éyang Béatrice partie très tôt, qu'elle trouve ici le couronnement de son entreprise, elle qui a conduit mes premiers pas vers l'école. Maman, je te dédie ce rêve qui a pris forme.

Mon père, Cyriaque Essono Mebale de m'avoir inculqué les valeurs nobles de la vie, de m'apprendre le sens du travail et de la persévérance en toute choses. Grâce à tes conseils tu m'as aidé à découvrir que le savoir est un trésor inépuisable. Merci également pour ton soutien téléphonique tout au long de mes études. Papa, Merci infiniment.

Le Seigneur est mon berger, je ne manquerai de rien

Chapitre 1. Introduction générale

1.1. Contexte

Selon le rapport « *La performance environnementale de l'agriculture dans les pays de l'OCDE depuis 1990* » (OCDE, 2008) plusieurs problèmes de la qualité de l'eau persistent encore dans les pays membres de cette organisation. Il s'agit entre autres de l'eutrophisation, de la prolifération des algues, ainsi que la pollution chimique et bactérienne des eaux de surface. Pour plusieurs de ces pays, l'agriculture est souvent désignée comme responsable de ces maux. En effet, depuis près de 50 ans, le secteur agricole dans ces pays a connu de profondes transformations du point de vue de sa croissance. Celles-ci se traduisent par des modifications des systèmes d'exploitation, une hausse spectaculaire des quantités produites, des surfaces cultivées, de la charge animale et de l'utilisation d'intrants tels que les engrais organiques et les pesticides.

Au Canada, ce phénomène est bien marqué. Depuis 1956, les superficies des pâturages ensemencés ont augmenté de 18,5 %; la taille moyenne des exploitations agricoles a augmenté de 6,9 %; la superficie des terres louées pour l'exploitation agricole de 9,9 % et le nombre de porcs de 7,8 % (Statistiques Canada, 2012). Ce modèle de production a d'ailleurs imposé une restructuration importante des élevages orientée vers la spécialisation des productions, l'augmentation de la taille des cheptels, ainsi que le développement des productions hors-sol. Au Québec, l'évolution des élevages s'est démarquée par la production laitière, bovine et porcine.

La contrepartie de cette évolution est une exacerbation des problèmes environnementaux issus de la gestion inadéquate des effluents d'élevage notamment le lisier, le fumier et le purin. Ces effluents d'élevage contiennent de l'azote, du phosphore, du potassium ainsi que de nombreux éléments nutritifs secondaires et oligo-éléments qui présentent un intérêt pour la nutrition minérale des plantes et des apports de matières organiques au sol. De ce fait, ils sont utilisés comme fertilisants par les agriculteurs qui pratiquent l'épandage. Or, il s'avère que dans plusieurs zones au Québec, les producteurs font face à un problème majeur : le nombre et l'étendue des terres disponibles pour l'épandage et le recyclage sécuritaire de ces déjections animales; celles-ci ne sont plus suffisantes pour les volumes d'effluents d'élevage générés (Magnan, 2001). Pour preuve, le Centre de développement du porc du Québec (2001), mentionne

que seulement pour les installations sur les bassins versants de la rivière Chaudière, la superficie de sol manquante pour l'application de lisier et de fumier est estimée à 50 000 hectares.

Pour combler ce manque d'espace, les agriculteurs effectuent des applications répétitives de lisiers sur les champs. Or, lorsque la capacité des terres agricoles d'assimiler ces effluents d'élevage est dépassée, les substances épandues se retrouvent éventuellement dans les eaux de surface et les eaux souterraines (Lavoie, 1994). Cette capacité d'assimilation ou de réception est limitée par l'aptitude physique des sols à retenir ou transformer ces résidus d'élevage (Lavoie, 1994). Cette situation de surplus de fumiers ou de lisiers favorisent les mauvaises pratiques de disposition, et contribuent ainsi à augmenter indirectement les apports en éléments nutritifs aux eaux de surface et souterraines.

Au Canada, il appert que de nombreuses fermes sont en situation de surplus de lisier. À titre d'exemple, la quantité de déjections par ferme a été multipliée par quatre entre 1951 et 2001 pour l'ensemble des productions animales, mais cette quantité s'est vue multipliée par 90 dans le cas du porc et par plus de 100 pour la volaille (Ministère de l'environnement, 2003). Selon les chiffres d'Agriculture et Agroalimentaire Canada (2003) sur la production totale de fumier, le bétail des fermes canadiennes a produit près de 152 millions de tonnes de fumier par jour. La même étude mentionne que l'intensité de la production de fumier, c'est-à-dire la quantité produite dans une région donnée, a grimpé d'environ la moitié des sous-bassins de drainage. De ce total, une proportion de 38 % provenait des vaches de boucherie, suivies des vaches laitières (12 %), des veaux (12 %), des génisses (12 %), des bouvillons (10 %), des porcs (9 %), de la volaille (3 %), des chevaux (2 %), des taureaux (2 %) et des moutons (moins de 1 %). Cependant, cette étude mentionne que la production de fumier se concentre dans trois grands groupements où la production de fumier excède les 2 000 kg à l'hectare de superficie totale. Il s'agit du centre et du sud de l'Alberta, du sud du Manitoba et de l'Ontario ainsi que du sud-est du Québec.

Au Québec seulement, une quantité de 14 millions de tonnes de lisier est produite annuellement dont environ 4 millions de tonnes étant du lisier de porc spécifiquement (Baril, 2002; BAPE, 2003). Cette quantité varie considérablement d'une région à l'autre du Québec. Selon les données de l'Institut de la statistique du Québec, (2011) près de 82 % du cheptel porcin québécois est concentré dans quatre régions administratives, à savoir : Montérégie, Chaudière Appalaches, Centre-du-Québec et Estrie. La Montérégie demeure la principale région

productrice avec 33 % du cheptel. Elle est suivie par la Chaudière-Appalaches (31 %), le Centre-du-Québec (13 %) et l'Estrie (5 %). À elles seules, ces régions sont les hôtes de 77 % de la production porcine au Québec (Statistique du Québec, 2011). La culture intensive du maïs, l'élevage porcin industriel et l'épandage excessif des effluents d'élevage s'avèrent être les principales causes de dégradation de la qualité des cours d'eau en Montérégie, notamment la rivière Yamaska (OBV Yamaska, 2013).

Au cours des dernières années, une grande attention a été portée sur les risques physico-chimiques associés aux épandages d'effluents d'élevage : fumiers et lisiers. Toutefois, la contamination microbienne des eaux, servant aux activités récréatives qui y sont associées arrive souvent plus loin dans la liste des priorités des problèmes à résoudre. Or, les déjections animales contiennent des concentrations importantes de micro-organismes dont certains peuvent être pathogènes. Ces derniers peuvent présenter des risques importants pour l'environnement. Ces risques sont plus élevés si le fumier est peu pailleux ou s'il est à proximité d'une voie de circulation. Dans ce contexte, certaines études (Gangbazo *et al.*, 1996 ; Gangbazo *et al.*, 1997) ont démontré que l'application de lisier de porc accroît les risques de ruissellement, et que cette situation s'explique principalement par l'augmentation de l'humidité du sol. Par conséquent, les risques de ruissellement augmentent avec le volume de lisier épandu ainsi qu'avec la rapidité avec laquelle la première pluie survient après l'application (Gingras, 1998 et Gingras *et al.*, 2000).

Au-delà des impacts sur les écosystèmes, la présence dans des cours d'eau des bactéries pathogènes telles que *Escherichia coli* communément appelée *E. coli* présentes dans les effluents d'élevages est susceptible de compromettre les activités récréatives comme la baignade ou les activités nautiques impliquant un contact avec l'eau. Elles peuvent être responsables des infections intestinales, principalement la gastro-entérite (Mariani-Kurkdjian, 2012). Dans la majorité des cas, cette infection est de courte durée, et les symptômes apparaissent brutalement et disparaissent au bout d'un à trois jours (OMS, 2013). Cependant, dans 10 % des cas, cette dernière peut évoluer en un syndrome hémolytique et urémique. Les personnes les plus susceptibles de développer la maladie sont les enfants en bas âge (< 5 ans), les personnes âgées (> 65 ans) et les personnes ayant pris récemment un traitement antibiotique (Griffin et Tauxe, 1991). Néanmoins, toute personne est potentiellement à risque et peut développer une diarrhée lors d'une ingestion d'*E. coli*.

Au Québec, malgré les incertitudes et le peu de cas confirmés de maladies humaines associées à cette contamination, le risque pour la santé publique est bien présent et pourrait même être en augmentation (Gingras *et al.* 2000). D'ailleurs, durant les dix dernières années, le nombre d'infections déclarées liées au stéréotype *E. coli* O157:H7 a subi une hausse de 239 à 383 (Louchini et Douville, 2001). Une étude épidémiologique sur les maladies reliées aux loisirs aquatiques (Sanborn et Takaro, 2013) rapporte qu'au Canada, durant la saison des baignades, la majorité des grands risques de gastroentérites bactériennes, ne viennent probablement pas de sources alimentaires, ni de l'eau potable, mais plutôt de l'exposition à l'eau durant des activités récréatives. Si au Canada, le premier cas d'infection d'origine hydrique par la souche O157:H7, lié aux activités agricoles, a été mis en évidence au début des années 1990 (Louchini et Douville, 2001), ce n'est qu'en 2001, que la première éclosion d'*E. coli* O157:H7 associée à l'eau de baignade a été découverte à Montréal-Centre. Dès lors, quelques études au Québec (Chagnon et Bolduc, 2001; GCÉ, 2002; MAPAQ, 2003, Rousseau et Normand, 2004) ont montré une forte corrélation entre les taux d'infections transmissibles par l'eau, possiblement d'origine animale et l'application de lisier ou de fumier de bovins. Les auteurs mentionnent que cette corrélation, quoique modérée, est plus importante dans le sous-groupe des enfants âgés de moins de 5 ans. Dans le même ordre d'idées une autre étude (Lainesse, 1999) mentionne une éclosion de gastro-entérite causée par *E. coli* O157:H7 lors d'une baignade en rivière, dans un secteur agricole de la région Chaudière-Appalaches. Il a d'ailleurs été démontré que cette région est le siège du plus grand nombre d'épisodes et avec le plus de personnes touchées par épisode. Les principales raisons de ces épisodes de contamination sont liées à l'importance des activités agricoles dans cette région. D'autres études (Gingras *et al.* 2000; Ministère de l'Environnement du Québec, 2001) ont démontré que les taux des infections transmissibles par l'eau et possiblement d'origine animale sont légèrement plus élevés dans les municipalités agricoles en surplus de fumier. Selon le recensement des éclosions d'origine hydrique effectué auprès des directions de santé publique (DSP) et le Ministère de la Santé et des Services Sociaux (MSSS), une éclosion de gastro-entérite d'origine hydrique est survenue en Montérégie au mois de septembre 2000. Concernant cette éclosion, huit (8) cas étaient de nature infectieuse et causés par l'ingestion d'eau dans un milieu de loisir (GCÉ, 2002).

À l'été 1998 et au début des années 2000, plusieurs cas de dermatite du baigneur ont été déclarés dans les Centres de Santé et de Services Sociaux (CSSS) de Beauce et de la région de Thetford

(Bolduc, 1998; Chagnon et Bolduc, 2000; MSSS, 2005; Turcotte, 2006). Il ressort que le lac Fortin à Saint-Victor et le lac Poulin à Saint-Benoît-Labre sont les principaux plans d'eau où des cas de dermatite du baigneur se sont déclarés, pour le CSSS de Beauce. Quant aux CSSS de la région de Thetford, plusieurs épisodes de dermatites sont survenus au Lac Aylmer, qui chevauche Chaudière-Appalaches et l'Estrie, et quelques cas au lac à la Truite à Thetford Mines. Là encore, le surplus de fumier dans ces municipalités en serait la cause (Chagnon et Bolduc, 2001, MSSS). En Estrie, l'un des cas les plus médiatisés est celui d'une fillette de trois ans et demie, qui reste avec des séquelles après avoir contracté la gastro-entérite suite à une baignade à plage du parc de la Plage-Municipale de Deauville (Bombardier, 2011).

Au Québec, bien que les problèmes de santé associés aux contaminants dans les eaux récréatives soient peu documentés, du fait que les cas sont rarement signalés aux responsables de la santé, ces cas de gastro-entérites et de dermatites laissent entrevoir l'existence d'un risque potentiel de santé publique en lien avec la pollution des eaux récréatives par des bactéries. Il est alors important d'améliorer les mécanismes de surveillance des risques liés à l'eau.

La surveillance des agents pathogènes dans l'eau est indispensable pour comprendre leur variation spatio-temporelle, leur concentration et les facteurs qui contrôlent leur dynamique. Jusqu'à présent, la surveillance de la qualité des eaux récréatives est réalisée par des prélèvements d'échantillons d'eau. Ces méthodes de surveillance ne protègent pas suffisamment contre les maladies dues à la baignade et, au Canada la plupart des baignades se font dans des eaux sans surveillance (Sanborn et Takaro, 2013). Il y a donc un besoin de développer des méthodes efficaces de surveillance de la qualité microbienne des eaux des plages publiques.

La compréhension des processus hydrologiques et ceux impliquant le devenir et la mortalité des bactéries est essentielle pour évaluer la dynamique spatio-temporelle de la contamination des eaux. De ce fait, la modélisation agro-hydrologique spatialisée devient l'outil privilégié pour la simulation de la qualité des eaux au niveau des bassins versants. Actuellement, il existe plusieurs modèles agro-hydrologiques couplés aux Systèmes d'Informations Géographiques (SIG), afin d'étudier la dynamique des flux de contaminants et la qualité de l'eau en rivière dans le temps (Prado et Daniel, 2000; Pei et Zhao, 2000) et l'espace (Zeilhofer *et al.*, 2007, Nas et Berktaç, 2010; Sener et Davraz, 2013). Si l'usage des modèles agro-hydrologiques permet d'identifier plusieurs éléments compromettant la qualité des eaux, peu de modèles opérationnels, ne permettent pas encore de définir un schéma fonctionnel suffisant pour

cartographier et prédire les dynamiques temporelles de contamination des eaux par les coliformes dans des zones à vocation agricole.

C'est dans ce contexte que nous soulevons l'importance de développer un modèle spatio-temporelle impliquant les processus hydrologiques, les caractéristiques agro-pédo-climatiques et leurs interactions avec la mortalité et le devenir des bactéries en vue de développer un schéma de prévisions des épisodes de contamination bactériennes des eaux récréatives.

1.2. Objectifs de la thèse

L'objectif de cette thèse est de comprendre, de prédire et de cartographier la dynamique spatio-temporelle de la contamination microbienne des eaux récréatives par une modélisation spatialement explicite. Celle-ci intègre les processus hydrologiques de surface, les caractéristiques agro-pédologiques et pluviométriques. Pour répondre à cet objectif, nous avons procédé en trois étapes, chacune reposant sur une approche de modélisation et correspondant aux trois parties de ce travail de thèse.

Dans la première partie de notre travail, l'objectif a été d'identifier et de hiérarchiser les facteurs spatiaux qui influencent considérablement la contamination des eaux à partir d'un modèle probabiliste bivarié : le « poids de la preuve ». Ce premier aspect a ainsi conduit à caractériser l'occurrence spatiale de la contamination microbienne et produire des cartes de susceptibilité, afin de déterminer les secteurs les plus favorables aux épisodes de contamination microbienne.

Le deuxième objectif de ce travail a consisté à étudier comment la variabilité des facteurs spatiaux du modèle probabiliste bivarié se répercute sur la contamination microbienne, et présenter la méthodologie d'analyse de sensibilité du modèle probabiliste bivarié. Il s'agit dans cette partie, de quantifier l'aide des indices de SOBOL, le domaine de variabilité des sorties de ce modèle probabiliste, en déterminant la part de variance due à l'occupation du sol et aux caractéristiques pédologiques.

Enfin, le troisième objectif a consisté en la modélisation des flux de coliformes dans les eaux de surface dans un contexte événementiel. Dans cette partie, un modèle événementiel AWSWAT a permis de représenter la dynamique de transfert temporel des flux de coliformes à travers un large panel de scénarii sur plusieurs années. Ce modèle a tenu compte de l'ensemble des connaissances disponibles sur les caractéristiques d'une pluie, en particulier la hauteur, la durée et l'intensité de la pluie. À travers ce modèle, nous avons mis en évidence une valeur seuil

des intensités de pluies nécessaire au déclenchement du transport des coliformes vers les eaux.

1.3. Organisation de la thèse

Cette thèse est organisée en quatre chapitres auxquels s'ajoutent une introduction générale et une conclusion finale ainsi que les perspectives de recherche et une liste des références bibliographiques consultées.

Dans le premier chapitre 1 il s'agit de situer la problématique de la contamination microbienne dans un contexte général.

Le chapitre 2 expose la cartographie de la contamination microbienne des eaux à l'aide d'une approche probabiliste issue du poids de la preuve (*PdIP*). Il s'agit dans cette partie de quantifier et cartographier les niveaux de probabilité des eaux à être contaminées par des effluents agricoles, sur la base des données spectrales et des variables géomorphologiques, tout en tenant compte de leurs occurrences spatiale et temporelle.

Le chapitre 3 est consacré à la quantification des incertitudes liées aux simulations de la contamination microbienne des eaux et l'analyse de sensibilité paramétrique à l'aide des indices de Sobol. L'enjeu principal est la compréhension du comportement de *PdIP* aux changements de variables et à travers différents scénarios, pour caractériser sa réponse dans un intervalle de confiance en tenant compte des incertitudes associées aux variables spatiales.

Dans le chapitre 4 nous abordons l'influence des caractéristiques pluvieuses sur la dynamique de la contamination microbienne des eaux à travers le modèle agro-hydrologique ArcView Soil and Water Assessment tools (AVSWAT). Il s'agit ici de prédire de manière temporelle la charge bactérienne véhiculée lors de la pluie dans les eaux récréatives à l'échelle du BV du lac Massawippi, en tenant compte de la hauteur, la durée et l'intensité d'un épisode pluvieux. Ce chapitre expose la démarche de calage et de validation du modèle et propose une discussion des résultats obtenus qui identifient les principaux traits de fonctionnement du BV. La conclusion de ce mémoire présente une synthèse de nos contributions proposées dans les différents chapitres ainsi que les perspectives de recherche qu'il serait intéressant de développer pour compléter les limites et élargir nos travaux.

Chapitre 2. Revue de Littérature

2.2. Revue de littérature

Pour percevoir l'avancée des recherches sur la cartographie de la qualité des eaux, nous avons procédé à une revue de littérature permettant de bien cibler les facteurs influant la contamination microbienne ainsi que les problématiques générales liées à sa cartographie dans le cadre de la contamination des eaux. Nous donnons en introduction une description de la bactérie *E. coli*.

2.2.1. *Escherichia coli* et sa survie dans les eaux

Escherichia Coli dénommé *E. coli* est une bactérie de la microflore commensale intestinale de l'homme et des animaux à sang chaud (mammifères et oiseaux) (Greatorex et Thorne, 1994). Le genre *Escherichia* appartient à la famille des Enterobacteriaceae qui regroupe des espèces fréquemment isolées du tube digestif ou des fèces des mammifères. Leur présence dans l'eau fournit ainsi une indication sur une éventuelle contamination fécale par des bactéries. En dehors d'*E. coli*, le genre *Escherichia* regroupe *E. blattae*, *E. fergusonii*, *E. hermanii* et *E. vulneris*. Chaque espèce d'*Escherichia* possède des caractéristiques biochimiques spécifiques, permettant de les différencier (Grimont, 1987). Par ailleurs, bien que la majorité des souches d'*E. coli* soient commensales, certaines d'entre elles sont associées à des pathologies intestinales (Levine, 1987) ou extra-intestinales (Pohl *et al.*, 1989) très diverses chez l'homme.

Concernant, la survie et la croissance d'*E. coli* dans divers milieux, plusieurs études (Yaun *et al.*, 2003) s'accordent sur le fait que les paramètres environnementaux tels que la température, le PH, et les radiations solaires (UV) semblent avoir une influence négative sur la survie de celle-ci.

L'effet sur la température est beaucoup documenté comparativement aux deux autres facteurs. Cependant, les résultats obtenus sont contradictoires. La principale raison de ces différences est que la majorité de ces études sont réalisées sur des fèces contaminées artificiellement, à des niveaux différents et dans des conditions expérimentales variées. Certains auteurs comme Ogden *et al.* (2002); Gagliardi et Karns, 2002; Islam *et al.* (2005) ont décrit un effet positif des faibles températures sur la survie d'*E. coli* O157:H7. Ils associent cet effet à un ralentissement du métabolisme cellulaire du pathogène et celui de toutes les activités compétitives et/ou antagonistes des bactéries indigènes du sol ou de l'eau. Ces résultats sont les

mêmes que ceux de nombreux auteurs qui ont montré que les basses températures prolongeraient la survie des coliformes fécaux (Smith *et al.*, 2006 ; Hughes, 2003; Noble *et al.*, 2004). Cependant, Jiang *et al.* (2002) ont démontré le contraire dans leur étude. D'après ces derniers, la survie d' *E. coli* pourrait atteindre 231 jours à 21°C. Alors que cette durée baisse de 77 jours à 5°C. Mukherjee *et al.* (2006) soutiennent également que la survie de cette bactérie pathogène est plus faible à 4°C qu'à température ambiante (10 contre 92 jours). En fait, l'examen de ces études laisse croire que d'autres paramètres en plus de la température contrôlent la survie des bactéries. D'ailleurs, dans ce contexte Wang *et al.* (2004) ont étudié les effets simultanés de la température, l'activité de l'eau et le niveau de contamination initial des fèces sur la survie d'*E. coli*. Dans un premier temps les auteurs ont conclu que la durée de vie est plus longue à une température de 5°C comparativement à 22°C ou 35°C pour un même taux de contamination initial soit 10⁵ ufc/g (respectivement, 70 jours, 56, jours et 49 jours). Dans les réservoirs d'eau douce Wang et Doyle, (1998) ont démontré que la survie du pathogène est d'autant plus longue que la température est elle-même plus basse. Ils affirment qu'à 8°C, la population d'*E. coli* O157:H7 diminue de 2 unités décimales après 91 jours d'incubation alors qu'à 15 et 25°C, le pathogène ne serait pas détecté après 14 et 21 jours.

2.2.2. La distribution des bactéries dans l'eau

Le phénomène de distribution, c'est-à-dire le déplacement des polluants dans leur milieu récepteur, est primordial puisqu'il détermine leur accumulation ou leur dilution dans l'eau. Pour comprendre et représenter les facteurs influençant ce phénomène, il nous est apparu nécessaire d'examiner chacun d'entre eux. Nous nous sommes limités ici à la prise en compte des deux facteurs météorologiques prépondérants dans notre problématique, à savoir le vent et les processus hydrodynamiques.

Une fois rejetées dans les milieux aquatiques, les bactéries peuvent se disperser rapidement dans le milieu. Le vent est l'un des paramètres les plus importants pour le transport et la dispersion des polluants dans l'eau des lacs (Servais *et al.*, 2005). Les océans et les rivières ont un système naturel de dilution pour les polluants arrivant, tandis que les lacs n'ont pas de réel point de sortie. Dans ces écosystèmes, la dispersion sera plus influencée par l'action du vent et quelques processus hydrodynamiques (Pommepuy *et al.*, 2005).

Le vent intervient à toutes les échelles tant par sa direction que par sa vitesse. De ce fait,

son influence sur la pollution bactérienne est variable selon la profondeur du lac. D'une part, les courants induits par le vent conduisent à un taux de renouvellement de l'eau et un pouvoir dispersif du milieu. Ces derniers modifient la circulation des masses fluides dans un lac, ce qui affecte directement la dispersion des nutriments, des polluants et par voie de conséquence, la productivité des écosystèmes. Ainsi, dans des lacs peu profonds, la circulation d'eau induit une augmentation du mélange, une augmentation de la suspension des sédiments et des nutriments à partir du fond. Ceci conduit d'une part à une plus grande accessibilité aux nutriments et, d'autre part à l'accroissement de la turbidité de l'eau. Or, pour une turbidité importante, l'effet germicide des rayons solaires est limité et le T90 d'*E. coli* est plus long (Pommepuy *et al.*, 2005). Par ailleurs, des vents faibles sont à l'origine d'une stabilité plus grande de la colonne d'eau, un mélange plus réduit, de moindres concentrations en polluants et nutriments. Les déplacements des masses d'eau dus aux vagues et provoqués par les vents entraînent alors le brassage des eaux, occasionnant ainsi la dispersion horizontale ou verticale des polluants, la remise en suspension des contaminants bactériens piégés dans les sédiments superficiels. Le brassage engendré par ces vents est donc susceptible de rendre les nutriments plus accessibles aux bactéries. Il existe un lien de causalité entre le brassage créé par les vagues et la vitesse du vent, qui sont des mécanismes par lesquels les bactéries sont dispersées dans le milieu lacustre et remis en suspension, avec altération consécutive de la qualité microbiologique de l'eau.

Les facteurs hydrodynamiques qui s'appliquent à l'*E. coli* incluent la dilution dans le milieu aquatique récepteur, la dispersion, la sédimentation et la suspension. Les deux premiers processus dépendent uniquement de l'hydrodynamique du système tandis que les deux derniers sont conditionnés par l'attachement des bactéries fécales à des matières en suspension (Pommepuy *et al.*, 2005). À la diffusion réelle due aux phénomènes hydrodynamiques, s'ajoute une décroissance bactérienne due aux phénomènes d'autoépuration. Les mécanismes explicatifs de cette décroissance bactérienne sont assez mal connus. Mais quelques paramètres semblent avoir plus d'influence sur le pouvoir auto-épurateur des eaux. Il s'agit de la température et de l'intensité lumineuse ainsi que de la turbidité des eaux. D'autres facteurs jouent également un rôle direct ou indirect : le pH et les antagonismes entre organismes vivants. L'impact général du vent et des facteurs hydrodynamiques sont connus qualitativement mais sans avoir de moyens précis de les quantifier.

2.2.3. Transport des bactéries par les matières en suspension dans l'eau

Il est unanimement admis que les particules en suspension jouent un rôle important comme vecteur physique dans le transport des polluants bactériens. Les bactéries sont ainsi véhiculées par ces particules sur de grandes distances en comparaison avec les distances qu'elles pourraient parcourir seules en leur absence (Kretzschmar *et al.*, 1999). En effet, les bactéries sont soumises à des forces biologiques et physico-chimiques qui influencent leur mouvement. Ces forces produisent alors des mécanismes de rétention qui incluent la filtration, l'absorption, l'interception, l'attachement et la sédimentation. De ce fait, elles peuvent s'accrocher aux surfaces des particules et former des biofilms qui leur permettent une meilleure protection aux fluctuations chimiques et une teneur parfois élevée en nutriments et en matière organique. Puisque les bactéries ont tendance à être absorbées par des particules et former des bio-films, celles-ci suivent le déplacement de ces supports et au final, sédimentent aux fonds des cours d'eau dépendamment de leur débit. Les bactéries se retrouvent ainsi sur des fonds vaseux plus favorables à leur survie. Ainsi, lors de brassages dus au vent, aux activités dans l'eau, ces bactéries peuvent être éventuellement remises en suspension et polluer les plages sans qu'il n'y ait de déversement d'eaux usées. D'ailleurs, les concentrations en bactéries fécales observées dans les sédiments sont souvent de 10 à 1000 fois supérieures à celles de la colonne d'eau sus-jacente.

Dans la littérature, plusieurs auteurs s'accordent à dire (Kretzschmar *et al.*, 1999), qu'une fraction allant jusqu'à 94 % peut être fixée sur le matériel particulaire. Cette fraction fixée correspondrait à la partie la plus active du peuplement bactérien. Le transfert des bactéries sur des particules leur occasionne un stress puisqu'elles peuvent passer d'un milieu riche en nutriments, correspondant à leur habitat primaire, à un milieu oligotrophe. En réponse donc à ce stress, les bactéries fécales développent des stratégies de survie leur permettant de persister dans un tel milieu. Après leur arrivée dans le cours d'eau, le nombre de bactéries attachées aux sédiments et leur croissance dépendent de plusieurs facteurs. Un des premiers paramètres dominants est la température (Annexe 1). Un autre facteur de survie est la profondeur de sédimentation. Quelques études ont démontré une abondance graduée qui correspond à un gradient décroissant de concentration des bactéries au fur et à mesure que l'on s'éloigne de la surface. Jacquet *et al.* (2005) indique que l'abondance bactérienne dans les sédiments lacustres varie entre 2,5 et 600×10^{10} ml⁻¹. L'abondance diminue toutefois avec la profondeur. Des

résultats similaires ont été obtenus par Etienne *et al.* (2007) qui ont dénombré les bactéries dans les sédiments d'un marais salé de Caroline du Nord. Les résultats obtenus ont montré que la concentration des bactéries varie de $2,19$ à $2,58 \times 10^9$ cells/cm³ de sédiments à une profondeur de 20 cm, tandis qu'à la surface, cette valeur se situe entre $8,36$ et $10,90 \times 10^9$ cells/cm³ de sédiments.

2.2.4. Corrélation entre *E. coli* et les matières en suspension

La principale étude qui établit la corrélation entre *E. coli* et les matières en suspension (MES) des zones de baignade en eau douce, est celle de Servais *et al.* (2005). Cette étude propose un modèle expliquant la relation entre la concentration des bactéries et les MES. L'approche de mesure utilisée est le suivi de l'activité glucuronidase d'*E. coli* pour expliquer la présence de coliformes dans les MES. Elle ne nécessite pas le décrochage d'*E. coli* du support solide pour les quantifier. En comparaison, les méthodes de mise en culture, sont beaucoup plus laborieuses.

À partir de la méthode de suivi de l'activité glucuronidase d'*E. coli* et un échantillonnage des teneurs très variées des MES, une régression linéaire entre le taux de bactéries attachées et la teneur en MES a été établie. Ce modèle explique à peu près 80 % des variations des concentrations d'*E. coli*. De manière générale, le modèle semble prédire avec une efficacité assez importante les taux d'*E. coli* dans les MES, et ainsi être en mesure de prévoir les journées d'ouverture ou de fermeture de la plage. La mise en place de cette équation revient aussi à dire que, plus la teneur en MES est élevée, plus la part de l'activité due aux *E. coli* fixés aux MES augmente. Par ailleurs, les auteurs ont remarqué que dès que la concentration en MES dépasse 200 mg/l, la quasi totalité de l'activité est due aux *E. coli* fixés aux MES. D'autres résultats de cette étude montrent que la quantité d'*E. coli* libre dans la colonne d'eau disparaît deux fois plus vite que la quantité d'*E. coli* fixée aux MES. Servais *et al.* (2005) ont aussi observé que les *E. coli* sont distribués de façon assez homogène entre les différentes classes de taille de particules. Les résultats obtenus par Servais *et al.* (2005) mettent en évidence l'importance du lien existant entre les concentrations en MES et les coliformes, d'où l'importance de considérer les MES lors de la modélisation de la dynamique d'*E. coli*. Ce lien s'avère donc un élément clé pour évaluer par télédétection le risque sanitaire associé aux bactéries ainsi que l'évaluation de leur concentration en zones de baignade.

L'intérêt d'une telle relation permet tout d'abord de mieux comprendre la dynamique d'*E. coli* en relation avec les MES. Ensuite, si cette relation est bien établie, elle peut permettre de cartographier les niveaux de pollution bactérienne sur des secteurs non mesurés à partir des mesures satellitaires. En effet, les concentrations des bactéries sont estimées sur la base d'une relation statistique entre la teneur en MES et la réflectance diffuse des eaux, à une longueur d'onde donnée. Or, les MES modifient les interactions de la lumière avec l'eau et donc également le signal reçu par le satellite. L'inversion de ce signal pour estimer directement les concentrations d'*E. coli* devient alors difficile. En principe, dans la bande bleue du visible la réflectance de l'eau pure est assez forte. Elle diminue ensuite rapidement pour devenir très faible dès le proche infrarouge. En effet, il y a diffusion de la lumière par les molécules d'eau et les corps dissous (la diffusion de Rayleigh), qui joue un rôle important uniquement dans les courtes longueurs d'ondes (couleur bleue). Par contre, quelques centimètres suffisent pour absorber tout le rayonnement solaire dans le proche infrarouge (0,7 μ m à 1,4 μ m) et le moyen infrarouge (1,5 μ m à 3 μ m). L'eau apparaît donc noire dans ces longueurs d'onde (Guyot 1989). La réflectance sous la surface dépend uniquement des propriétés inhérentes de la colonne de l'eau, et non pas des conditions instantanées d'éclairement. À cela s'ajoutent, la sous-représentativité des phénomènes expliquant la présence des particules dans l'eau et de manière plus récurrente, les erreurs liées à la paramétrisation des coefficients d'absorption et de rétrodiffusion des MES. Une approche, basée sur la modélisation de la qualité de l'eau permettra d'appréhender certaines caractéristiques de la pollution bactérienne susceptibles d'influencer les mesures de réflectances à différentes longueurs d'ondes. Cet outil est susceptible de prédire la concentration des bactéries en étudiant l'impact de scénarios de changement d'utilisation du sol sur la qualité des eaux. Il serait donc nécessaire d'utiliser conjointement le modèle de la qualité de l'eau avec l'approche par télédétection dans le but de fournir un aperçu général de la contamination bactérienne que l'on ne peut pas obtenir à l'aide des approches optiques par télédétection.

2.2.3. Facteurs influant la contamination des eaux de surface

2.2.3.1. Les variables géomorphologiques

La topographie du milieu, conjuguée aux conditions climatiques, notamment l'intensité des pluies, est un facteur déterminant du transport des bactéries en autant que les conditions du sol s'y prêtent (Crane *et al.*, 1983; Gabet *et al.*, 2003). Ce facteur intervient indirectement en agissant

sur le ruissellement de l'eau, la vitesse de son écoulement de surface et de ce fait, sur les processus de transport (Etzelmüller *et al.*, 2000; Gabet *et al.*, 2003). En principe, l'influence potentielle de la topographie au transport des bactéries est fonction des caractéristiques de la pente qui sont : la forme, la longueur ainsi que l'inclinaison. Ces dernières agissent principalement sur les paramètres de détachement des matières en suspension, favorisant ainsi la dynamique des contaminants. D'ailleurs, ces auteurs (Mutchler et Greer, 1980; Poesen, 1984; Poesen, 1987) ont démontré que, plus la pente est forte, plus le ruissellement est important, entraînant une importante quantité de sédiments par augmentation de la vitesse d'écoulement; puisque les particules en suspension jouent un rôle important comme vecteur physique dans le transport des polluants bactériens. Ce dernier, *via* les sédiments, dépend non seulement de l'inclinaison de la pente mais aussi de la rugosité du sol. Cependant, certains auteurs (Machado *et al.*, 2006) estiment que ce transfert de bactéries est influencé par d'autres facteurs tels que l'occupation du sol et la rugosité de la surface. Ces auteurs soulignent d'ailleurs que sur un sol nu, le ruissellement serait directement proportionnel à l'inclinaison de la pente et aux concentrations des polluants. Cependant, aucune relation significative n'a été observée sur des surfaces couvertes. Par ailleurs, on estime que cette relation est fonction de la perméabilité des sols (Berville *et al.*, 2006). En effet, sur des sols imperméables, le ruissellement pluvial n'est pas fonction de la pente, mais de l'intensité de la pluie. Par contre, si les sols sont relativement perméables, la pente aura une influence certaine sur l'infiltration et donc le ruissellement. Dans le même contexte certains auteurs (Etzelmüller *et al.*, 2000; Kirsch *et al.*, 2002) admettent que le transport des sédiments croît avec la longueur de la pente. Cela s'explique par le fait que les pentes les plus longues permettent une plus forte accumulation du ruissellement, ce qui accroît l'énergie globale de celui-ci et sa force de détachement.

2.2.3.2. Les variables pédologiques

Les caractéristiques du sol sont un facteur de grand intérêt dans la dynamique des coliformes vers les eaux de surface (Green *et al.*, 2007). Par sa texture, le sol intervient sur le bilan hydrique en agissant directement sur la vitesse d'infiltration et de ruissellement, et indirectement sur le transfert des polluants. À ce sujet plusieurs études (Fenlon *et al.*, 2000; Kay *et al.*, 2005b) ont démontré que la texture du sol, sous l'action conjuguée de paramètres environnementaux, a une influence directe la survie des polluants. D'ailleurs, ces études ont prouvé que les sols compactés

et peu perméables, notamment argileux, sont propices au ruissellement des coliformes. Les mêmes auteurs soutiennent que la survie d'*E. coli* est inférieure dans un sol sableux par rapport à des sols argileux ou limoneux. Ces études admettent que *E. coli* est capable de survivre pendant quatre mois dans des sols argileux et limoneux contre seulement huit semaines dans un sol sableux. En effet, l'action et le rôle des sols argileux sont d'une importance capitale en termes de protection aux fluctuations chimiques et au transfert des bactéries. Green *et al.* (2007) et Guber *et al.* (2007) ont démontré que la présence de la matière organique dans le sol joue également un rôle considérable dans le transport et la survie d'*E. coli*. Ce deuxième effet compense largement le premier puisque c'est le taux de matière organique qui favorise la rétention des bactéries aux particules solides, formant ainsi un filtre qui empêche leur mouvement en profondeur, les bactéries s'accumulant aux premiers centimètres du sol.

2.2.3.3. Les variables climatiques : les précipitations

Quelques recherches ont été réalisées afin de quantifier l'influence de la pluie sur la contamination microbiologique des eaux de surface (Kay *et al.*, 2005a; Guber *et al.*, 2005; Grimaldi *et al.*, 2006). Ces travaux de recherche ont mis en évidence le rôle important de la pluie dans le transport par ruissellement des contaminants vers les milieux récepteurs. En effet, en tant qu'agent érosif, la pluie par la force de son énergie occasionne le décollement des particules tandis que le ruissellement génère une mise en suspension qui favorise le déplacement des polluants. Cependant, ces auteurs s'accordent sur le fait que la pollution par le ruissellement ne dépend pas essentiellement des caractéristiques de la pluie; mais plutôt de la conjonction de plusieurs phénomènes dont : les délais entre l'épandage de fumier et la texture du sol. Dans ce cadre, d'autres études (Auzet, 1990) ont démontré que l'action dégradante de la pluie est maximale pour des textures dont le diamètre médian est compris entre 63 et 250 μm . Ce qui correspond à la gamme de sol limoneux et limono-sableux.

2.2.3.4. Les variables liées à l'occupation du sol

La connaissance de l'occupation du sol d'un BV est fondamentale pour déterminer la distribution spatio-temporelle des apports de polluants. Ce facteur influence indirectement la répartition du transfert entre infiltration et ruissellement mais aussi les quantités de polluants transférées (Durand *et al.*, 2006; Goss, *al.*, 2008). Dans la littérature, plusieurs catégories d'occupation du

sol ont été identifiées afin d'expliquer cette influence. Il ressort que la relation existante entre l'occupation du sol et le flux de polluants transférés est fonction d'un jeu complexe de facteurs interdépendants (Aitken, 2003; Kay *et al.*, 2005b). Cependant, il est unanimement admis que la végétation joue un rôle prépondérant dans le transport des polluants. L'analyse de Goss *et al.* (2008), montre que sous les mêmes conditions climatiques, le ruissellement et le flux des polluants augmentent avec une diminution du couvert végétal. Ils soulignent par ailleurs que le coefficient de ruissellement favorisant le transport, est presque nul lorsque la couverture végétale dépasse 65 %. La principale raison évoquée est l'obstacle que forme la végétation face aux écoulements ainsi que son caractère stabilisant de particules du sol. Les auteurs concluent enfin qu'à l'occupation du sol s'ajoutent d'autres variables comme la matière organique du sol, la densité apparente du sol, le couvert de litière et le gradient de pente qui peuvent affecter la production de ruissellement.

La distance des zones d'épandage au réseau hydrographique est également un facteur à prendre en considération pour les études concernant la qualité des eaux. En effet, comme l'ont démontré plusieurs études (Kay *et al.*, 2005b; Durand *et al.*, 2006), plus la distance entre une parcelle et le réseau hydrographique est grande, plus les processus de rétention et de mortalité des bactéries sont élevés et plus la pollution est réduite. Cependant, peu de travaux chiffrant la distance au-delà de laquelle une parcelle n'est plus contributrice à la pollution d'un cours d'eau.

2.2.4. Méthode de cartographie de la qualité des eaux de surface

Nombreuses sont les études portant sur la cartographie de la qualité des eaux (Ruhard, 1978; Schnebelen *et al.*, 2001; Seguin et Baudry, 2002). De façon générale, les approches utilisées peuvent se répartir en deux grandes catégories : les approches qualitatives et les approches quantitatives.

2.2.4.1. Les méthodes qualitatives

Traditionnellement, la cartographie de la qualité des eaux face à une pollution se fait à l'aide de l'approche qualitative à travers les méthodes de pondération et d'indexation des critères (Pételet *et al.*, 2006; Dörfliger *et al.*, 2004; Kavouri et Guoin, 2009). Le principe de ce type d'approches consiste à combiner les cartes de chacun des paramètres morphologiques en donnant un index numérique à chaque paramètre. Pour ce faire, l'expertise d'un professionnel est indispensable

pour l'établissement de la liste des facteurs et des modalités des classes. L'intervention de ce dernier contribue ainsi à l'élaboration de procédures et de méthodes permettant de mettre en place des bases de règles pour la modélisation de la dynamique de la pollution. Bien que la cartographie à index avec pondération des critères soit l'approche la plus utilisée pour déterminer la susceptibilité des eaux à la pollution, plusieurs auteurs (Plagnes, 2006; Pranville, 2008) estiment que les cartes mises en place pour une même zone peuvent varier d'un concepteur à un autre à cause de la détermination des poids des paramètres parfois subjective. Ce constat pose ainsi la question de validation de cette approche encore problématique.

À la lumière de la littérature consultée, les points de vue sont très controversés quant à la pertinence des approches qualitatives pour l'évaluation de la susceptibilité des eaux à la pollution. Schnebelen *et al.* (2001); Li *et al.*, (2010) sont très sceptiques quant à l'opportunité d'utiliser la cartographie à index. Ces auteurs soulignent que ces méthodes nécessitent de prendre en compte tous les aspects subjectifs du problème considéré, pour l'attribution des poids aux différents facteurs. À ce problème, s'ajoute un questionnement sur l'incertitude et l'imprécision des données utilisées pour quantifier les différents facteurs et la façon d'en tenir compte (Corgne, 2004). Pour prendre en charge ces limites, des approches alternatives telles que les lois de la probabilité et la règle de décision bayésienne, la théorie de l'évidence sont proposées (Maquaire *et al.*, 2004). Cependant, Corgne (2004) présente des conclusions plutôt enthousiastes. Il soutient que la subjectivité liée à la mise en place de ces méthodes n'influence pas autant la qualité des résultats lorsque les paramètres environnementaux sont bien connus.

2.2.4.2. Les méthodes quantitatives

Dans l'objectif de mettre au point des méthodes plus objectives, les approches quantitatives ou probabilistes ont été développées pour cartographier la susceptibilité des eaux aux pollutions diffuses ou ponctuelles. Le principe de ces méthodes est basé sur l'information obtenue des interrelations entre les facteurs conditionnant un phénomène donné et sa distribution. Ces méthodes peuvent être regroupées en deux principaux groupes: l'approche bivariée et l'approche multivariée. La première approche consiste à assigner des poids aux variables prises en compte et à chacune des classes de chaque variable, pour l'élaboration d'une carte (Omlin et Reichert, 1999; Sin *et al.*, 2009). La pondération de chaque facteur contributif est ainsi obtenue à l'aide de différents algorithmes précis de manière objective, sans l'intervention d'experts.

Dans la littérature, l'une des techniques probabilistes les plus répandues est celle des méthodes bayésiennes (Kuczera, 1997; Kuczera et Parent, 1998; Corgne, 2004; Kanso, 2004). Le principe de cette approche considère qu'il existe deux sources d'informations à propos d'un paramètre inconnu : une information *a priori*, basée sur des données historiques ou sur la connaissance du modélisateur avant de n'assembler aucune observation, et une information *a posteriori* obtenue par des observations ou par expérience. Par suite, l'information sur les paramètres est actualisée à partir d'une connaissance *a priori* des paramètres, représentée par une probabilité Pp , vers une probabilité *a posteriori* $Ppst$ en utilisant les informations dans les données observées (Kanso, 2004).

Dans la littérature, plusieurs auteurs (Beven et Binley, 1992; Kuczera et Parent, 1998; Campbell *et al.*, 1999; Kanso, 2004) ont essayé l'utilisation des modèles bayésiens pour la modélisation de transfert de produits phytosanitaires, des flux de rejets urbains à l'échelle du BV. De ces études, il ressort que les modèles bayésiens permettent de mieux prendre en compte les incertitudes sur les paramètres d'entrée. Par ailleurs, ces facteurs n'ont aucune influence l'un sur l'autre dans le calcul des pondérations (Kuczera et Parent, 1998). En somme, les approches probabilistes permettent d'identifier et de hiérarchiser les variables géographiques qui ont un rôle important dans la genèse d'un phénomène donné. Ces modèles vont au-delà de simples modèles descriptifs en permettant la prédiction des occurrences spatiales (Campbell *et al.*, 1999; Kanso, 2004; Smith et Marshall, 2008).

Au niveau des méthodes multivariées, les analyses logistiques sont souvent les plus utilisées (Reghunath *et al.*, 2002; Shrestha et Kazama, 2007). Il s'agit de méthodes correspondant à un modèle non-linéaire, permettant de calculer la probabilité d'occurrence spatiale d'un phénomène en fonction des variables de prédisposition et de déclenchement, en se basant sur une fonction sigmoïdale (Steven *et al.*, 2001). Cette approche a permis à Turgeon (2011) d'analyser la relation entre un groupe de déterminants agroenvironnementaux et la contamination fécale des eaux récréatives. Des auteurs (Boyacioglu, 2008; Yang *et al.*, 2010, Fan *et al.*, 2010) soutiennent d'ailleurs que de tels modèles permettent de prévoir les changements temporels dans une zone donnée, ce qui permet de déterminer le lieu du changement. L'utilisation fréquente dans plusieurs études des régressions logistiques s'explique par leur capacité d'analyse des causes explicites dans l'espace et de prévision des changements potentiels dans une zone donnée (Zhang *et al.*, 2009; Zhang *et al.*, 2011).

Dans la même catégorie d'approches, on distingue l'analyse des régressions multiples. Cette méthode consiste à établir un rapport fonctionnel entre des variables (Neter et Wasserman, 1974; Hébert, 2001) ou évaluer le rapport entre une ou plusieurs variables indépendantes et une variable continue dépendante. Cette approche a été utilisée par Patoine (2011) pour évaluer les impacts de la densité animale sur la contamination microbienne des eaux. Cet auteur a démontré que, du point de vue de leur applicabilité, ces modèles ont le potentiel d'être beaucoup plus flexibles. Ils permettent d'intégrer dans une même relation plusieurs variables indépendantes comme les variables liées à l'occupation du sol, de données pluviométriques. Après plusieurs études au niveau de la prédiction de la contamination microbienne, Gangbazo (2000) et Hébert (2001) ont remarqué l'importance des régressions statistiques à expliquer et prédire les attributs géomorphologiques et météorologiques associés à la dynamique de la contamination microbienne. Cependant, les auteurs mentionnent que la principale limite de ces modèles est qu'ils sont très liés aux spécificités géographiques des sites sur lesquels ils s'appliquent et demeurent donc difficilement transposables à d'autres espaces.

Finalement, à la lumière de toute l'information présentée et par analogie avec les méthodes utilisées en géologie et en épidémiologie, nous avons choisi une nouvelle méthode de la cartographie spatio-temporelle de la contamination microbienne des eaux. Il s'agit de celle basée sur la théorie de l'évidence. Ce modèle probabiliste bivarié est basé sur une approche statistique dite bayésienne. La théorie de l'évidence a largement été étudiée et appliquée dans les domaines de recherches scientifiques telles que l'écologie (Semenzin, *et al.*, 2007) et la géologie (Van, *et al.*, 1997; Prabin *et al.*, 2012). Cependant, son application semble rare dans l'évaluation de la qualité des eaux où une grande attention est portée sur l'efficacité des modèles agro-hydrologiques.

Fréquemment utilisée en épidémiologie (Gemperli *et al.* (2005); Juan *et al.* (2005); Swaen et Van (2009)), la théorie de l'évidence permet de mettre en exergue les régions avec un risque anormalement élevé afin de mieux comprendre l'étiologie des maladies. Ces auteurs soutiennent que la théorie de l'évidence fournit des cartes de risque plus fiables que celles obtenues par les méthodes empiriques conceptuelles, avec moins d'erreurs de classification et plus de régions clairement délimitées en zones de risque. Dans une étude en géologie, Thierry *et al.* (2004) soulignent que la théorie de l'évidence est considérée comme l'approche probabiliste la plus robuste pour étudier la prédiction du devenir d'un territoire, pour l'identification des

facteurs contributifs ainsi que leur hiérarchisation quantitative. C'est l'une des principales raisons du choix de la théorie de l'évidence pour notre étude. La deuxième raison est la capacité de la théorie de l'évidence à évaluer des incertitudes liées à l'estimation des paramètres du modèle, ce qui permet d'aller plus loin et de tester la validité du modèle en explorant tout l'espace des valeurs des paramètres (Bonham-Carter, 1994; Thierry *et al.*, 2004).

Chapitre 3. Modélisation probabiliste de la contamination microbienne des eaux récréatives

3.1. Introduction

La présence de bactéries fécales telles que *E. coli* dans les eaux récréatives pose un problème de santé publique. Certains types d'*E. coli* sont pathogènes et représentent ainsi un danger pour la santé humaine. La surveillance de ces agents pathogènes dans l'eau est alors indispensable afin de comprendre leur variation spatio-temporelle, leur concentration et les facteurs qui contrôlent leur dynamique. Jusqu'à présent, la surveillance de la qualité des eaux récréatives est réalisée par des prélèvements d'échantillons d'eau. Afin de quantifier la contamination microbienne des eaux récréatives, le Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques (MDDELCC) a mis en place des programmes de surveillance. Ceux-ci recommandent l'interdiction de la baignade et restreignent certaines activités nautiques lorsque la concentration en coliformes dépasse 200 ufc/100 ml (MDDELCC, 2009). Pour établir ces recommandations, il est nécessaire de mettre en évidence la présence des bactéries d'origine fécale, notamment *E. coli*, dans les échantillons d'eau.

L'approche utilisée implique habituellement des méthodes de dénombrement fondées sur la mise en culture des bactéries. Dans la pratique, deux méthodes sont régulièrement utilisées. Il s'agit de la détermination du nombre le plus probable (NPP), et de la méthode de filtration sur membrane (MF). La première méthode est une estimation statistique de la densité des micro-organismes. Il s'agit d'inoculer des dilutions décimales dans une série de tubes contenant un milieu de culture liquide spécifique. Ensuite, la loi de Poisson permet de calculer le NPP sur la base de la proportion de tubes positifs dans chaque dilution. La deuxième méthode quant à elle consiste à filtrer un volume d'eau connu sur une membrane poreuse. Cette dernière est ensuite mise dans des conditions favorisant le développement des coliformes. Après 24 heures, les bactéries forment des colonies identifiables à l'œil nu. Les résultats obtenus par cette méthode s'expriment en unités formant des colonies (ufc) par volume d'eau filtrée. Récemment, le dénombrement de la totalité des bactéries a été automatisé; elle se fait par des compteurs automatiques de particules. L'inconvénient majeur de cette méthode est qu'elle ne fait pas de distinction entre les bactéries viables et mortes. Elle n'est donc fiable que dans les conditions où la plupart des bactéries sont vivantes. À cela s'ajoute le long temps de réponse qui varie de 24 h

à 48 h. Aussi, les procédures liées à l'acquisition d'échantillons et leur mesure imposent souvent un délai minimum de quatre à cinq jours pour donner lieu à un avis de fermeture de plage (MDDELCC, 2009). Au-delà des délais requis, s'ajoute le problème du petit nombre d'échantillons prélevés sur de petites superficies, ce qui amène une certaine imprécision dans l'évaluation réelle du niveau de bactéries.

Une alternative à ces méthodes est l'utilisation des modèles hydro-agricoles (Guber *et al.*, 2007; Rogel, 2007; Qiu *et al.*, 2009). Ces derniers simulent la dynamique de la contamination microbienne des eaux à partir des données hydrométéorologiques et la variabilité spatiale du milieu. À ce jour, on peut dire que les objectifs de ces modèles sont en partie atteints puisque les principaux facteurs affectant le transfert des coliformes sont identifiés. Néanmoins, la plupart de ces modèles restent encore peu opérationnels (Goovaerts, 2002) et présentent un certain nombre de problèmes méthodologiques, parmi lesquels :

- la représentation des processus hydrologiques et les conséquences sur les modalités de segmentation de la surface en unités hydrologiques;
- les difficultés de paramétrisation dues aux grands nombres de paramètres à renseigner, impliquant une masse de données à acquérir et surtout à adapter aux non-linéarités des équations mises en jeu à une échelle donnée (Corgne, 2004);
- les problèmes inhérents à la spatialisation des données.

Face à ces limites et malgré les efforts substantiels consacrés à l'étude de la pollution microbienne des eaux de surface, il reste encore difficile d'établir avec certitude les conditions favorables à une contamination microbienne. Une meilleure compréhension de ce phénomène implique nécessairement que les différents paramètres favorisant le transfert des coliformes soient caractérisés de manière explicite, de telle sorte que l'on puisse localiser et spatialiser les zones susceptibles de favoriser le déclenchement d'un tel phénomène.

La mise en œuvre de cette approche permettra de quantifier, de hiérarchiser et de cartographier les facteurs potentiels qui interviennent dans le déclenchement de la pollution microbienne.

3.2. Objectifs

Au regard des options de modélisation, nous proposons comme objectif principal de cette première partie de notre thèse, le développement d'un modèle prédictif, selon une approche probabiliste, par *le poids de la preuve* ou *PdIP* (Bonham-Carter, 1994, Lamothe, 2009), pour

l'estimation cartographique de la contamination microbienne des eaux récréatives à l'échelle spatiale du BV. En se basant sur le concept de la théorie de l'évidence, la contamination microbienne des eaux est prédite à partir des relations conditionnelles (probabilités de Bayes) entre la teneur en coliformes fécaux et les paramètres agro-pédo-climatiques caractérisant la superficie des bassins versants analysés. Une contrainte importante dans le développement du modèle prédictif est d'assurer sa transposabilité sur différents bassins versants des basses-terres du St-Laurent comme domaine d'application. Il s'agit ici de tester le *PdIP* sur d'autres bassins versants aux conditions topographiques, hydrographiques et climatiques variées afin de définir la généralisation de l'approche choisie.

Pour atteindre cet objectif général, les objectifs spécifiques consistent à :

- identifier les facteurs de prédisposition et de déclenchement et valider statistiquement leur choix pour la prédiction spatiale de la probabilité de la contamination microbienne;
- caractériser l'occurrence spatiale et fournir une cartographie des probabilités de contamination microbienne des eaux, répondant aux besoins des acteurs de la gestion des eaux;
- évaluer la robustesse de la technique de modélisation *PdIP* à prédire la contamination microbienne et sa capacité à choisir de façon hiérarchique les variables explicatives destinées au modèle.

3.3. Principe du modèle *PdIP*

Le modèle *PdIP* est une méthode quantitative bivariée qui permet d'estimer l'importance relative des éléments de preuve (*Evidence*) en utilisant la forme log-linéaire du modèle de probabilité bayésien. Le modèle *PdIP* s'appuie sur deux informations qui se proposent de déterminer :

- La probabilité *a priori* d'un événement;
- La probabilité *a posteriori* qui est modifiée par l'intégration d'une nouvelle évidence.

De façon simplifiée, *PdIP* permet de combiner une nouvelle évidence avec une connaissance définie *a priori* sur une hypothèse H donnée, pour estimer la probabilité que celle-ci soit vérifiée ou pas. Notons que Vd est la variable dépendante que l'on souhaite estimer qui est appelé ici hypothèse (dans notre cas c'est la probabilité que les coliformes soient présents dans un lac).

La probabilité *a posteriori* qui actualise la probabilité *a priori*, exprime la probabilité

d'occurrence d'une variable dépendante V_d sous la présence conditionnelle de la variables explicative V_e . Succinctement, il s'agit de calculer les probabilités de présence ou d'absence de V_d dans les zones occupées par une variable explicative V_e . En utilisant les règles de la théorie bayésienne, la probabilité *a posteriori* de trouver V_d dans les zones occupées par la variable V_e s'exprime comme suit:

$$Pp\{V_d|V_e\} = \frac{P(V_e \cap V_d)}{P_{V_e}} = \frac{P\{V_e|V_d\}}{P_{V_e}} \times P_{V_d}. \quad \text{Éq. 1}$$

De même, la probabilité *a posteriori* de trouver V_d en l par la variable V_e est :

$$Pp\{V_d|\overline{V_e}\} = \frac{P(\overline{V_e} \cap V_d)}{V_e} = \frac{P\{\overline{V_e}|V_d\}}{V_e} \times P_{V_d}. \quad \text{Éq. 2}$$

$Pp\{V_d|V_e\}$ et $Pp\{V_d|\overline{V_e}\}$ représentent les probabilités conditionnelles que H soit vraie en considérant respectivement la présence et / ou l'absence de la variable explicative V_e .

Le ratio du logarithme de ces probabilités conditionnelles définit les poids positif et négatif de la variable explicative. Ces poids sont donc définis comme suit :

$$W^+ = \log_e \frac{Pp\{V_e|V_d\}}{Pp\{\overline{V_e}|V_d\}} \quad \text{Éq. 3}$$

et

$$W^- = \log_e \frac{Pp\{\overline{V_e}|V_d\}}{Pp\{V_e|V_d\}}. \quad \text{Éq. 4}$$

W^+ et W^- représentent l'importance relative de la présence et de l'absence de la variable prédictive dans la dynamique de V_d . Si W^+ est > 0 , alors la présence de V_e est favorable à l'occurrence du phénomène, son absence étant défavorable. Si W^- est > 0 , la présence de la classe est défavorable à l'occurrence du phénomène, par contre son absence est favorable à l'occurrence de celui-ci. La différence entre les poids positif et négatif calculée pour V_e est le contraste C . Ce paramètre mesure la corrélation entre V_e et V_d . Lorsque sa valeur tend vers zéro, la présence de la V_e n'influence pas la distribution de V_d ; aucune relation n'existe donc entre V_e et V_d . Cette relation est positivement ou négativement significative lorsque C tend vers 1 ou -1 (Weed, 2005).

Si plus de deux variables explicatives sont utilisées, leurs poids sont ajoutés dans le modèle et la

probabilité *a posteriori* est ensuite calculée à partir des poids globaux des facteurs prédictifs et la probabilité *a priori* :

$$\log_e\{O\} \{V_d | V_{e1}^k \cap V_{e2}^k \cap V_{e3}^k \dots \dots \dots \cap V_{en}^k\} = \sum_{i=1}^n W_{V_{ei}}^k + \log_e\{O_{V_d}\} \quad \text{Éq. 5}$$

$W_{V_{ei}}^k$ est le poids global des facteurs. Ce dernier est obtenu en additionnant l'ensemble des poids des facteurs retenus selon leur hiérarchie. k peut prendre le signe $+$ ou $-$ dépendamment de la présence ou de l'absence de V_e .

3.4. Justification du choix du *PdIP*

Le *PdIP* est un modèle statistique/probabiliste bivariée, fondé sur le théorème général de Bayes (Malet *et al.*, 2006). Fréquemment utilisée en épidémiologie (Gemperli *et al.* (2004); Juan *et al.* (2006); Swaen & Van (2009); Burr *et al.* (2011); Dominik *et al.* (2012)), cette méthode permet de mettre en exergue les régions avec un risque anormalement élevé afin de mieux comprendre l'étiologie des maladies. Ces auteurs soutiennent que la théorie de l'évidence fournit des cartes de risque plus fiables que celles obtenues par les méthodes empiriques conceptuelles, avec moins d'erreurs de classification et plus de régions clairement délimitées en zones de risque. Dans une étude en géologie, Thierry *et al.* (2007) soulignent que le *PdIP* est considéré comme l'approche probabiliste la plus robuste pour étudier la prédiction du devenir d'un territoire, pour l'identification des facteurs contributifs ainsi que leur hiérarchisation quantitative.

Par ailleurs, il est reconnu que le *PdIP* a la capacité d'évaluer des incertitudes liées à l'estimation des paramètres du modèle, ce qui permet d'aller plus loin et de tester la validité du modèle en explorant tout l'espace des valeurs des paramètres (Bonham-Carter, 1994; Thierry *et al.*, 2004). Au plan des concepts, l'approche bivariée du *PdIP* offre un point de vue objectif sur les règles et les critères de décision pour les pondérations des variables. De même, il permet d'étudier la variabilité spatiale des données et présente plus de souplesse dans le choix des sites d'observations ponctuels qui sont aléatoires et peu nombreux. D'ailleurs, plusieurs auteurs s'accordent que cette approche est capable de fournir des prédictions appropriées en présence d'incertitude, de jugements subjectifs d'experts et d'incomplétude de l'information (Neil *et al.*, 2000).

Contrairement aux approches multivariées, les facteurs n'ont aucune influence l'un sur l'autre dans le calcul des pondérations, ce qui permet une analyse de sensibilité sur l'influence des

facteurs sur les cartes finales.

L'avantage principal du *PdlP* est sa simplicité, les valeurs prédictives ne pouvant être que de deux types : événement possible ou événement impossible. Il en découle alors une facilité de lecture et d'interprétation des résultats. Leur souplesse de mise en œuvre par rapport aux méthodes qualitatives est également appréciée car même avec peu de données, il est possible d'introduire l'opinion experte dans le processus de calcul.

Au delà des avantages conceptuels, le *PdlP* permet de faire une comparaison entre la répartition spatiale des événements et les variables considérées comme favorables à sa génération. Cette approche permet de déterminer alors le degré de causalité entre chaque variable et chaque type d'événement, contrairement aux méthodes multi variées dont le principal inconvénient est parfois la création de corrélations parfois grossières et illogiques entre des variables réellement indépendantes et la probabilité d'occurrence d'un événement.

En somme, le *PdlP* présente de nombreux avantages, entre autre : la méthodologie objective, le processus d'évaluation structuré et permettant une certaine automatisation, gestion des données pouvant être standardisée. Ce modèle statistique est introduit dans le SIG ArcView 3.2 par une extension appelée ArcSDM.

3.5. Zone d'étude

Le développement du modèle prédictif *PdlP* s'est fait sur deux bassins versants, ceux de la rivière Yamaska et du lac Massawippi, sous-bassin versant (SBV) hydrographique de la rivière Saint-François (figure 1). Ces bassins présentent des caractéristiques variées en termes morphologique, pédologique, topographique et agricole. Ces bassins versants s'inscrivent dans un contexte similaire : (i) de représentativité des problématiques de la contamination microbienne (ii) de présence d'étendus d'eau voués aux activités récréotouristiques, et (iii) de bassin typique perturbé par l'activité agricole et les pesticides.

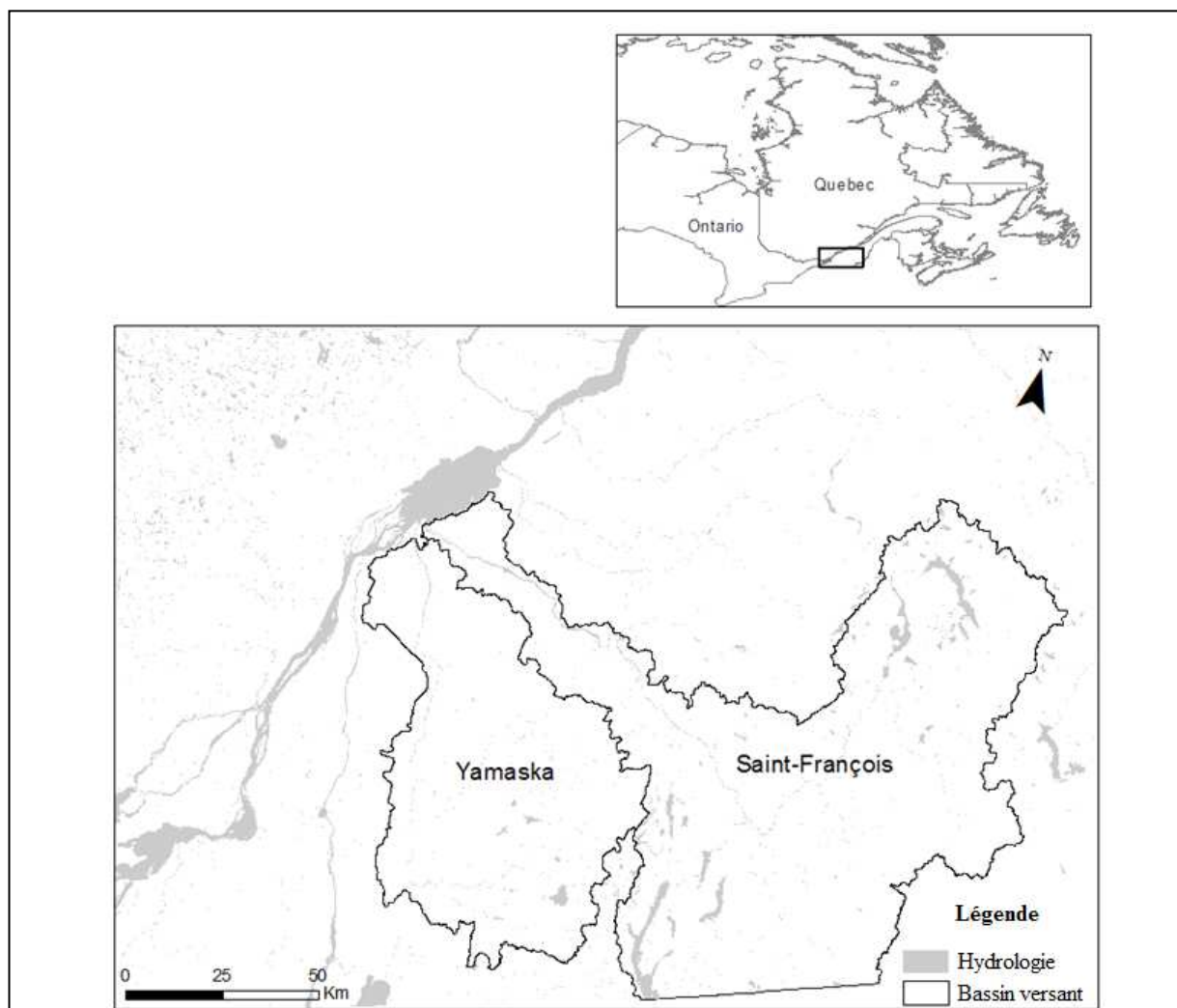


Figure 1. Les bassins versants de l'étude

3.5.1. Aperçu des bassins hydrographiques du lac Massawippi et de la Yamaska

Le bassin hydrographique de la rivière Yamaska draine une vaste région dont la superficie est de 4 784 km². Il est situé sur la rive sud du fleuve Saint-Laurent, entre les rivières Richelieu et Saint-François. Ce BV est drainé par trois principaux tributaires, dont les rivières Noire, Yamaska-Nord et Yamaska Sud-Est. Il se subdivise en sept sous-bassins versants de deuxième niveau dont : Yamaska Sud-Est, Yamaska-Nord, Yamaska, Noire, Salvail, Pot-au-Beurre et David (Annexe 1). Le BV de la Yamaska chevauche deux régions physiographiques naturelles : les basses-terres du Saint-Laurent situées en aval de la rivière et les Appalaches, en amont. Ces deux

reliefs contrastés confèrent au BV une occupation particulière de son territoire. Ainsi, en raison des fortes pentes et des sols à texture grossière, la partie du BV située dans les Appalaches est plus vouée aux activités de loisir en milieu naturel. Cependant, les basses-terres fertiles se caractérisent par l'agriculture, les activités urbaines et industrielles (APPEL, 2012). Chacune de ces régions occupe près de 50 % de la superficie du BV. Au niveau de l'occupation du sol, près de 63 % du bassin de la rivière Yamaska est occupé par des terres agricoles dédiées majoritairement à la production végétale et à la culture fourragère (APPEL, 2012). La forêt occupe 31 % de la superficie totale alors que les territoires urbains, les étendues d'eau et les autres affectations occupent 6 % de la surface restante. Pour ce qui est de la production animale, elle est prédominée par une grande production porcine, laitière et avicole. Ces productions représentent respectivement 58 %, 21 % et 13 % du cheptel du BV et correspondent à 92 % du nombre total d'unités animales (UA) du bassin. Au total, on dénombre plus de 314 500 UA sur ce BV, ce qui représente près de 17,5 % des UA du Québec.

Le BV du lac Massawippi est un SBV de 2^e niveau du BV de la rivière St-François principalement situé en Estrie (figure 1). Avec une superficie de 610 km², il est limité à l'ouest par le BV du lac Memphrémagog et à l'est par celui de la rivière Massawippi (RAPPEL, 2006). Au sud, une partie du bassin se trouve aux États-Unis. Le BV du lac Massawippi est drainé par dix-sept tributaires dont les principaux sont : les ruisseaux Niger, le Brook, le Pont couvert, le McConnell, le Webster, le Abbott et finalement les rivières la rivière Coaticook et Tomifobia (Annexe 2). Ce BV est composé de neuf sous-bassins versants de 3^e niveau dont le plus important en superficie est celui de la rivière Tomifobia avec une superficie de 437 km². Le plus petit SBV est celui du ruisseau « Pères-Servites » avec 4 km² de superficie (COGESAF, 2006). Au niveau topographique, le BV du lac Massawippi se situe à l'intérieur de l'unité physiographique des Appalaches, caractérisée par un relief accidenté et vallonné (Gélinas et Direction générale des eaux, 1977). Au niveau de l'occupation du sol, la forêt occupe 54 % du BV du lac Massawippi. L'agriculture représente 32 % de l'occupation du territoire, avec près de 80 % des sous-bassins considérés « agricoles ». On y trouve majoritairement (61 %) des prairies, des pâturages et des champs. 28 % de cette superficie est consacrée aux céréales et aux protéagineux. L'acériculture occupe 6 % tandis que les arbres de Noël représentent 5 % de la superficie cultivée. Au niveau des productions animales, l'élevage du bovin laitier est le type d'élevage le plus fréquemment rencontré dans ce BV. Il compte 42 % du cheptel, tandis que la

production représente 31 % des unités animales. Le sous-bassin de la rivière Tomifobia compte la plus grande proportion d'unités animales par rapport à la superficie en culture avec 1,78 UA par hectare (RAPPEL, 2006).

Pour la généralisation du modèle et sa transférabilité sur d'autres bassins versants, nous proposons de travailler sur un ensemble de petits sous-bassins hydrographiques, appartenant aux bassins versants susmentionnés.

3.5.2. Choix des sous-bassins versants de l'étude

Le choix des sous-bassins versants a été effectué en fonction du paramètre principal étudié qui est le flux de coliformes dans les eaux. Dans un premier temps, les bassins sont sélectionnés en fonction de l'intérêt qu'ils représentent face à cette problématique de la contamination microbienne des eaux, bien évidemment de la disponibilité des données requises. Il s'agit plus précisément du nombre minimal de station d'échantillonnage par SBV. Se référant aux études antérieures de Turgeon (2011), un minimum de trois stations d'échantillonnage par SBV a été fixé. Le deuxième critère de sélection est fonction de la superficie. De ce fait, les sous-bassins versants ayant une superficie comprise entre 20 km² et 400 km² ont été sélectionnés. Pour une certaine hétérogénéité physique des sous-bassins versants, une autre sélection a été faite en fonction de l'hétérogénéité au niveau de la pédologie et de la morphologie. Plusieurs études (Fenlon *et al.*, 2000 ; Kay *et al.*, 2005b) ont mentionné que le rôle de la pédologie dans le transport des bactéries est primordial. De ce fait, la couverture pédologique et le relief des sous-bassins versants nous sont apparus comme les caractères les plus sélectifs pour apprécier l'hétérogénéité physique des bassins versants considérés. Les paramètres pris en compte ici sont la texture, le drainage des sols et les gradients de pente. Pour ce faire, les bassins sélectionnés ont été ceux qui présentent la plus grande diversité des types de sol et la plus grande disparité de valeurs de pente. Après avoir travaillé sur le critère de diversité des types de sol par SBV, nous avons pris en compte l'hétérogénéité spatiale afin d'effectuer un choix final. Les sous-bassins versants devaient donc être agricoles et peu urbanisés. Nous avons ainsi considéré les superficies agricoles comme critère de sélection et les bassins versants dont la superficie agricole est supérieure à 25 % ont été considérés comme BV à vocation agricole. Ce dernier critère a permis d'établir un choix final de sept sous-bassins. Trois de ces derniers appartenant au BV du Lac Massawippi ont été utilisés pour la calibration. Les quatre autres (deux dans les Terres-basses et

deux dans les Appalaches) appartenant au BV de la Yamaska, ont été utilisés pour la transposition du modèle. Un descriptif des sous-bassins versants est présenté au tableau 1.

Tableau 1. Caractéristiques des bassins versants à l'étude

| Bassins versants | | Modélé | Drainage | Granulométrie | Superficie totale (Km ²) | Superficie en culture (%) |
|-----------------------|-----------------|---------|-------------|--------------------|--------------------------------------|---------------------------|
| Bassins de validation | Yamaska-Nord | Plat | Bien drainé | Loameuse-grossière | 303 | 28,2 |
| | Salvail | Plat | Mal drainé | sableuse | 206 | 53,7 |
| | Yamaska Sud-Est | Plat | Bien drainé | Loameuse grossière | 326 | 59,8 |
| | Pot au beurre | Plat | Mal drainé | Argile fine | 208 | 45,8 |
| Bassin de calibration | Tomifobia | Plat | Bien drainé | Loameuse fine | 437 | 35 |
| | Brook | Ondulé | Bien drainé | Loameuse fine | 35 | 43 |
| | McConnell | Incliné | Bien drainé | Loameuse grossière | 22 | 37 |

3.6. Matériels

3.6.1. Base de données de la variable prédictive : la teneur en coliformes fécaux

La teneur en coliformes fécaux des différents sous-bassins versants de 2^e niveau constitue l'élément de base pour la prédiction de son occurrence spatiale ou temporelle. Pour cette étude, nous avons utilisé les données de concentration d'*E. coli* issues de plusieurs campagnes d'échantillonnage de l'été 2011. Ces données proviennent principalement de la banque de données sur la qualité des milieux aquatiques du MDDELCC, de plusieurs municipalités régionales de comté (MRC) (MRC Memphrémagog, MRC Coaticook, MRC de la Haute Yamaska), des organismes de bassins versants (RAPPEL, COGEBY) et de l'Agence de santé publique du Canada (APSC). Une campagne d'échantillonnage a été effectuée sur une période de quatre mois dans le but d'avoir une variété de mesures sur l'état de la pollution par *E. coli* des eaux du lac Massawippi. Des échantillons d'eau ont été prélevés sur une plage du mois de mai au mois d'août trois fois par semaine (lundi, mercredi et vendredi). À des endroits préalablement définis pour différentes profondeurs de la plage (15 cm, 60 cm et 70 cm), les échantillons d'eau sont prélevés dans des contenants de 250 ml et gardés à 4°C dans une glacière jusqu'à leur analyse en laboratoire (Laboratoire de lutte contre les zoonoses d'origine alimentaire de l'Agence de santé publique du Canada à Saint-Hyacinthe), dans un délai de 24 h. Ces mesures sont effectuées pour un total de 12 plages en Estrie et 15 plages en Montérégie afin de créer une base

de données complète sur deux bassins versants : ceux des rivières Saint-François et Yamaska.

La méthode utilisée pour le dénombrement d'*E.coli*, est celle de Quantitray 2000 avec le milieu de culture de Colilert. Il s'agit d'une méthode de dénombrement semi-automatisée basée sur le principe du nombre le plus probable (NPP). Dans un premier temps, les échantillons sont sortis à l'avance des réfrigérateurs, pour qu'ils se réchauffent et qu'ils soient facilement utilisables. À ces échantillons d'eau de 100 ml, une ampoule du milieu de culture Colilert est ajoutée. Les échantillons sont alors agités de façon à aider le réactif à se dissoudre complètement. Après ce mélange, le contenu est distribué dans une galerie Quantitray 2000 puis cette dernière est scellée à l'aide d'un appareil. Après incubation à 35°C pendant environ 24 h, le nombre de puits présentant une coloration jaune (coliformes totaux) est compté en premier. Ensuite, pour les puits de couleur jaune, se fait le décompte du nombre de puits fluorescents (*E. coli*). Enfin, il suffit de se référer à une table NPP fournie par le laboratoire Idexx (C'est le leader mondial dans la fourniture des services de contrôle de la qualité microbiologique de l'eau). Cette table donne le nombre de coliformes totaux et *E.coli*/100mL d'eau en fonction du nombre de puits positifs.

Au niveau du BV de la Yamaska, nous disposons de 111 stations d'échantillonnage de la qualité de l'eau. Parmi ces dernières, (37) proviennent du MDDELCC; (9) de APSC, (25) de la MRC de la haute Yamaska et (40) de la COGEBY. Par ailleurs, en ce qui concerne le BV du lac Massawippi, nous disposons des données des stations d'échantillonnage recueillies sur différents ruisseaux et rivières tributaires du BV du lac Massawippi. La majorité des stations d'échantillonnage proviennent des données disponibles de la base de données de la qualité de l'eau des MRC Coaticook et de Memphrémagog. Une seule station provient de l'APSC.

Sur la base des concentrations moyennes des coliformes fécaux calculées, les sites d'échantillonnage ont été classés en deux groupes : (1) sites contaminés, (2) sites non contaminés, avec 200 ufc/100 ml comme valeur seuil. Celle-ci a été sélectionnée sur la base des valeurs préétablies par le MDDELCC.

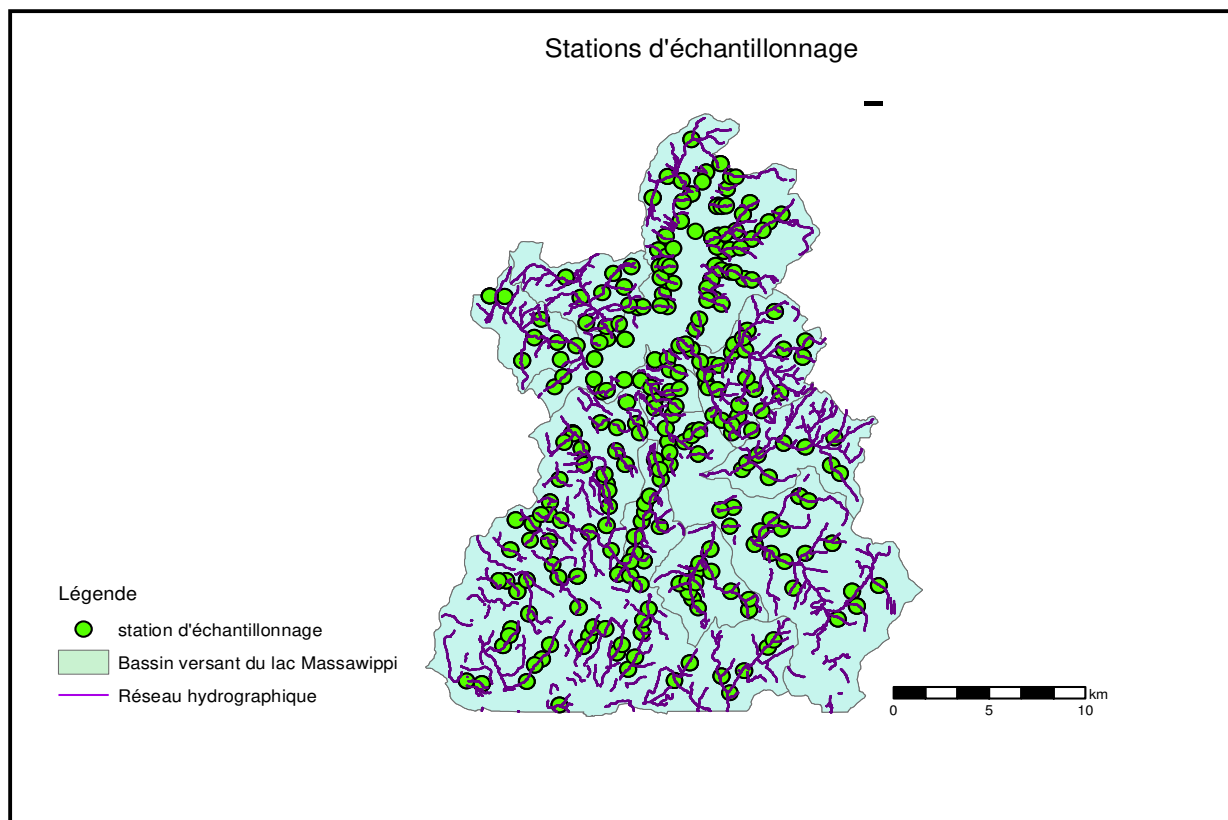


Figure 2. Distribution spatiale des stations d'échantillonnage.

3.6.2. Base de données nécessaires à l'extraction des variables explicatives

La deuxième base de données dont nous disposons regroupe les informations relatives aux variables prédictives. Dans la littérature, plusieurs variables sont connues pour avoir une influence soit sur la viabilité des coliformes, soit dans le transport de ceux-ci. Par rapport à la disponibilité des données et des connaissances que nous avons des sites d'étude, cinq principales variables ont été retenues. Il s'agit de la pédologie, du climat, de l'occupation du sol, de la distance aux cours d'eau et de la géomorphologie. Nous disposons donc :

- d'un modèle numérique d'altitude (MNA) à l'échelle de 1/250 000 des deux bassins versants de l'étude. Le MNA a été produit à partir de la base des données topographiques du Québec (BDTQ) et a permis de dériver les paramètres descripteurs de la topographie et leurs caractéristiques hydrologiques afin de cartographier les zones de ruissellement maximal;
- d'une carte des sols à l'échelle de 1/20 000 fournie par l'Institut de Recherche et de Développement en Agroenvironnement (IRDA). Cette carte couvre complètement les

zones d'étude et compte 600 polygones dont chacun est relié à une base de donnée contenant la superficie, la texture du sol, son épaisseur, des pourcentages de sable, le limon, les classes de matière organique ainsi que le drainage;

- des données hydrométéorologiques tirées de la base de données d'Environnement Canada, au pas de temps journalier; cette base de données contient les précipitations journalières mesurées dans les 34 stations recouvrant le BV du lac Massawippi et de la rivière Yamaska;
- d'une image du capteur satellitaire SPOT-5 acquise le 17 juillet 2011 pour la caractérisation de l'occupation du sol.

Le tableau 2 résume l'ensemble des données nécessaires pour l'extraction des variables explicatives.

| Données et paramètres descriptifs | | Origine des données de base |
|--|--|---|
| Limites des BVs | | Ministère de l'agriculture, des pêcheries et alimentations/Sherbrooke. Conseil de gestion du BV de la Yamaska (COGEBY) |
| Limites des SBV | | Base de données topographiques du Québec (BDTQ) (1/20 000) |
| Densité du réseau hydrographique | | Base de données topographiques du Québec (BDTQ) (1/20 000) |
| Données climatiques | - Précipitations (mm) | Environnement Canada, « Normales climatiques au Canada 1971-2000 ». |
| Données Topographiques | - Altitude (m) - Pente (%); Formes, Orientation | • Base de données topographiques du Québec (BDTQ) (1/25 000) Données numériques d'élévation du Canada |
| Données Pédologiques | - Type de sol - Granulométrie et drainage | • Bases nationales des données sur les sols (1/20000) • Carte des sols (1/20 000) IRDA |
| Données satellitaires (image Spot HRV) | | • Résolution spatiale : 250 m (rouge et infrarouge) et 500 m (bleu-vert) • Date d'acquisition : 17 juillet 2011 |
| Données bactériologiques | | • Campagnes de terrain (2010 - 2011) • MDDELCC (2010 - 2011) • MRC Memphrémagog (2010 - 2011) • MRC de Coaticook (2010-2011) • COGEBY (2010 - 2011) |

Tableau 2. Listes des données utilisées pour l'extraction des variables prédictives

3.7. Méthodologie

La méthodologie proposée repose sur deux phases distinctes, mais complémentaires. La première phase a consisté à structurer la base de données afin de regrouper des informations relatives au flux de coliformes fécaux (variable à modéliser) et aux cinq variables explicatives retenues. Une fois structurées, les corrélations entre les variables explicatives ont été évaluées à partir du test d'indépendance de *Chi carré* (Chi^2). La seconde phase a porté sur la hiérarchisation et la caractérisation de l'influence des variables explicatives sur l'occurrence d'apparition de la contamination microbienne. Le but de cette partie était d'éliminer les variables inutiles et de garder uniquement celles qui apportent une réponse significative à la prédiction de la contamination microbienne. Enfin, avec les variables explicatives sélectionnées, le *PdlP* a été mis en application. Pour ce faire, les poids positifs et négatifs (W^+ et W^-) ont été calculés pour chacune des classes de variables explicatives. Ensuite, le contraste C a été calculé afin d'avoir une idée quantitative des corrélations entre chaque variable explicative et la variable prédictive. Inspiré de Corby *et al.* (2005), nous avons retenu les classes de variables explicatives ayant un contraste C supérieur à 0,5 ; celles ayant un contraste supérieur à 0,6 ont été considérées comme les principales classes favorables à l'occurrence d'un épisode de contamination microbienne des eaux. À l'issue de cette étape, les poids les plus significatifs de chacune des variables prédictives ont ensuite été combinés et la probabilité *a posteriori* de la contamination microbienne a été calculée. Il s'agissait d'effectuer plusieurs itérations en introduisant les variables explicatives dans le modèle probabiliste afin d'obtenir un résultat optimal. Une analyse comparative des résultats a permis de choisir le meilleur modèle suivant deux critères : (1) la capacité explicative du modèle, évaluée à l'aide de la valeur de l'aire sous la courbe (ASC) qui traduit le pouvoir prédictif de chaque modèle; et (2) la capacité descriptive du modèle qui traduit la capacité de ce dernier à identifier les stations présentant les problématiques de contamination microbienne. Afin d'exploiter la reproductibilité de l'approche *PdlP* le meilleur modèle a été transposé pour validation sur les trois bassins versants de la rivière Yamaska.

3.7.1. Choix des variables explicatives

L'influence prépondérante de plusieurs facteurs conditionne la dynamique des bactéries, favorisant ainsi la contamination microbienne des eaux de surface. Dans la littérature, plusieurs études (Crowther *et al.*, 2001; Green *et al.*, 2007; Parajuli *et al.* 2007) ont démontré qu'à

l'échelle parcellaire, la quantité de fumier ou de lisier épandue dans les champs agricoles, les cycles de successions culturales, le type de sol ainsi que son utilisation seraient responsables de la contamination microbienne. Par ailleurs, au niveau du BV, ces études mentionnent qu'en plus de la quantité de fumier, les facteurs les plus déterminants et dont l'influence est la plus immédiatement perçue, sont celles liés aux caractéristiques physiques telles que la topographie et la pédologie. À ces facteurs s'ajoutent le fonctionnement hydrologique, la proximité des cultures au réseau hydrographique ainsi que la pluviométrie.

L'objectif de cette partie de l'étude consiste à extraire de cette masse de données, les variables pertinentes, disponibles et susceptibles d'expliquer, mais aussi d'évaluer au travers d'un modèle probabiliste, la dynamique de la contamination microbienne des eaux récréatives. Pour chaque variable explicative construite, un «nom de code» lui est attribué. C'est ce dernier qui sera utilisé dans le reste du document. Au total cinq variables explicatives de bases ont été utilisées pour cette étude. De nature quantitative et qualitative, il s'agit des données géomorphologiques, pédologiques, climatiques, hydrologiques et d'occupation du sol.

- Les variables géomorphologiques

Pour explorer l'influence des variables géomorphologiques sur la contamination microbienne des eaux, nous disposons d'un modèle numérique d'altitude (MNA) à l'échelle de 1/250 000 des deux zones d'étude. Ce dernier, issu de la base des données topographiques du Québec (BDTQ) a permis de dériver les paramètres descripteurs de la topographie et leurs caractéristiques hydrologiques afin de cartographier les zones de ruissellement maximal. À partir du MNA, nous avons ainsi procédé au calcul de variables morphométriques et leurs dérivées soient : le gradient de pente (pente), la forme (forme) et la direction (direction). Pour chacun de ces trois facteurs, un certain nombre de classes représentant les évidences a été défini. Au total, 18 évidences ou classes ont été dérivées de la variable «topographie». Ces dernières sont présentées en annexe 3.

- Les variables pédologiques

Pour caractériser l'influence des caractéristiques pédologiques sur la contamination des eaux du BV du lac Massawippi, une principale source d'informations cartographiques a été utilisée. Il s'agit d'une carte des séries de sol à l'échelle de 1/20 000 fournie par l'Institut de Recherche et de Développement en Agroenvironnement (IRDA). Cette carte couvre complètement la zone

d'étude et compte 600 polygones dont chacune est reliée à une base de donnée contenant la superficie, la texture du sol, son épaisseur, des pourcentages de sable, le limon, les classes de matière organique ainsi que le drainage. L'ensemble de ces informations générales a permis de classer les sols dépendamment de la texture (texture) et du drainage (drainage). La classification ainsi que la codification des évidences de chacun de ces facteurs est présentée en annexe 4.

- Les variables climatiques

Pour prendre en compte les variabilités spatiale et temporelle du climat, les données hydrométéorologiques tirées de la base de données d'Environnement Canada, au pas de temps journalier ont été utilisées. Cette base de données contient les précipitations journalières mesurées dans les 34 stations recouvrant le BV du lac Massawippi. Afin de représenter la variation spatiale des pluies (pluie) sur la zone d'étude, plusieurs méthodes d'interpolation ont été testées : krigeage ordinaire, co-krigeage ordinaire, distance inverse pondérée, spline, surface de tendance (trend, polynomiales locale et globale. La méthode retenue est la distance inverse pondérée qui a permis une automatisation des interpolations des données journalières de 2010 et 2011. La distance inverse pondérée est une méthode déterministe exacte et locale. L'avantage est que cette méthode permet de prendre en compte plus de données du champ d'observation. Un poids plus important est affecté aux sites les plus proches, un poids moindre aux sites plus éloignés (Baillargeon, 2005). De plus, cette approche permet d'intégrer l'interpolation des données provenant d'autres sources que les stations météorologiques. Finalement, quatre classes de pluviométrie ont été définies. Les annexes 5 et 6 présentent la répartition des précipitations, leur minimum, maximum, moyenne et écart type.

- Les variables liées à l'occupation du sol

La caractérisation de l'occupation du sol a été obtenue à partir de la classification dirigée de l'image SPOT-5. Pour ce faire, des traitements préliminaires ont été effectués sur cette image. Ces traitements impliquaient les processus de fusion, de corrections géométrique et atmosphérique pour éliminer ou, tout au moins, corriger les distorsions et les dégradations de l'image. Suite à cette étape, la classification dirigée par la méthode de maximum de vraisemblance, en s'appuyant sur les données directes de terrain a permis d'obtenir une carte d'occupation du sol. Sept classes ont été identifiées: forêt, sol-nu, terre agricole, eau, zone-

urbaine, zone-humide et autoroute. La carte de l'occupation du sol du BV du lac Massawippi est représentée en annexe 7.

- La proximité des zones agricoles aux cours d'eau

La pertinence de la proximité des zones agricoles au cours d'eau (distance) a été testée à l'aide de la distance euclidienne entre chaque type d'occupation du sol et les cours d'eau. Cette distance définit la ligne droite la plus courte entre un point quelconque et un point du réseau hydrographique. Ainsi, la proximité entre une catégorie d'occupation du sol et un cours d'eau a été évaluée en extrayant les surfaces d'eau à partir de la carte de couverture du sol et en calculant la distance horizontale entre chaque catégorie d'occupation du sol et le plan d'eau le plus proche. Trois zones tampons (buffers) autour du réseau hydrographique ont été créées afin d'obtenir des zones différenciées. Cela a permis d'identifier une distance seuil pour laquelle les corrélations entre les flux de coliformes et l'occupation du sol est optimale. Cette distance seuil représentera un critère de zonation des zones les plus contributrices aux flux de coliformes. La carte des distances présentant trois classes est illustrée en annexe 8.

En somme, cette partie de notre travail a permis de passer en revue l'ensemble des VP qui ont été utilisées dans cette étude. Au total 45 évidences (ou classes) ont été utilisées pour l'estimation de la cartographie de la contamination microbienne. Le nombre de facteurs de chaque VP ainsi que les évidences ou classes sont présentés dans le tableau 3.

Tableau 3. Les variables prédictives et les nombres de classes associées

| Variables explicatives | Facteurs | Nombres de classes ou d'évidence |
|------------------------|----------------------|--|
| Géomorphologie | Gradient des pentes | - Nul; - Faible; - Douce; - Modérée; - Forte; - Excessive |
| | Forme des pentes | - Dorsale; - Incliné; - Ondulé; - Placage; - Plat; - Vallonné |
| | Direction des pentes | - Nord; |

| | | |
|--------------------------|-------------------|---|
| | | <ul style="list-style-type: none"> - Nord-Ouest; - Sud-Ouest; - Sud; - Est; - Nord-est |
| Pédologie | Texture | <ul style="list-style-type: none"> - Loam; - Loam-argileux; - Loam-sableux; - Loam-Sableux fin; - Organique; - Sable fin; - Textures variables |
| | Drainage | <ul style="list-style-type: none"> - Bien drainé; - Imparfaitement drainé; - Mal drainé; - Rapidement drainé; - Très mal drainé; - Très rapidement drainé |
| Occupation du sol | Occupation du sol | <ul style="list-style-type: none"> - Terre agricole; - Forêt; - Zone urbaine; - Zone Humide; - Eau; - Sols nuls; - Eau; - Autoroute |
| Climat | Pluviométrie | <ul style="list-style-type: none"> - Faible; - Moyenne; - Forte; - Très-Forte |
| Distance aux cours d'eau | Longueur | <ul style="list-style-type: none"> - Faible; - Moyenne; - Forte |
| Nombre total d'évidence | | 45 |

3.7.2. L'évaluation de l'indépendance conditionnelle entre les variables explicatives

L'approche *PdIP* exige une indépendance conditionnelle entre les variables explicatives. La vérification de cette indépendance est une étape importante car elle permet d'apprécier le degré de corrélation et la force du lien entre elles. Cela permet ainsi d'éviter la redondance de l'information et de limiter le nombre de données aux seules variables dont l'influence est significative à l'occurrence de l'évènement. La violation de cette condition peut dégrader les

performances prédictives si elle n'est pas respectée, en entraînant une surestimation ou une sous-estimation des résultats finaux (la probabilité *a posteriori*). En comparant différentes méthodes évaluant la corrélation entre les variables explicatives et variable à modéliser, Ghattas. (2008) montrent que le χ^2 fournit des coefficients plus stables entre les variables, comparativement aux autres méthodes. Le test de χ^2 a été effectué entre les différentes variables pour tester l'hypothèse d'indépendance conditionnelle. Une première étape a consisté à convertir les cartes multi-classes de chacune des variables en carte binaire. Il s'agissait en effet de créer des cartes où la valeur 1 était attribuée à une classe en particulier de la variable prédictive et la valeur 0 à toute autre classe de la même variable. Cette carte binaire était ensuite combinée avec la couche d'informations montrant la répartition spatiale des points d'échantillonnage au sol. Puisque ces derniers avaient déjà été classés en deux catégories (contamination et non contamination) la nouvelle couche créée correspondait ainsi à la distribution spatiale d'une des deux valeurs en présence d'une certaine évidence. Cette démarche a donc été appliquée pour chacune des 45 preuves retenues dans la base de données globale (tableau 3). Après l'obtention des cartes binaires, le test d'indépendance conditionnelle a été effectué en croisant deux à deux les classes des variables explicatives sous la forme d'un tableau de contingence.

3.7.3. Application de l'approche *PdIP*

Une fois l'indépendance conditionnelle vérifiée et les variables explicatives choisies, *PdIP* a été mise en œuvre à travers différentes étapes dont : (1) le calcul des poids positifs (W^+) et négatifs (W^-) ainsi que le contraste C ($W^+ - W^-$) des variables choisies; (2) la simulation : Cette étape a consisté à combiner les variables explicatives sélectionnées une à une par ordre hiérarchique jusqu'à obtenir le résultat optimal. Pour ce faire, des tests avec plusieurs combinaisons de variables explicatives ont été réalisés; et enfin la dernière étape consiste au choix du meilleur modèle par comparaison.

Le meilleur modèle a été choisi en fonction de deux critères basés sur les capacités explicative et descriptive. Le premier critère, évalué sur l'analyse de l'ASC traduit la capacité du modèle à discriminer les échantillons contaminés des non-contaminés. Graphiquement, il s'agit d'une courbe qui donne le taux de bonne classification dans un groupe en fonction du taux de mauvaise classification pour ce même groupe. Lorsque l'ASC est supérieure à 0,8, le pouvoir discriminant est excellent ; il est bon lorsque l'ASC est comprise entre 0,7 et 0,8. Une ASC à 0,5

correspond à une prédiction au hasard. Le deuxième critère de comparaison repose sur la capacité descriptive du modèle. Cette capacité a été évaluée en comptabilisant le nombre d'observations concordantes entre la carte simulée et les observations sur le terrain. Au final, le meilleur modèle est utilisé pour une validation locale puis transposé sur les bassins de la rivière Yamaska pour une validation externe.

3.7.4. Calage

Le choix du protocole de calage a été orienté sur la base de la littérature (Rogel, 2007; Rousseau *et al.*, 2008). L'un des objectifs de ce travail était de déterminer le jeu de paramètres optimal qui influence la contamination microbienne sur un BV. Par conséquent, le modèle proposé doit prendre en compte l'homogénéité spatiale des caractéristiques de chacun des bassins versants. Cette homogénéité spatiale est essentielle, puisqu'elle permet d'évaluer le rôle des organisations hydro-géomorphologiques sur la réponse de chaque BV lors d'un épisode de contamination microbienne. Dans ce contexte, plusieurs études soulignent le faible apport des calages par combinaison de bassins versants différents (Vaze *et al.*, 2010; Wagener *et al.*, 2003). Ceci est lié au fait qu'en combinant les formes spatiales de différents bassins versants, leurs systèmes de pente ainsi que l'organisation de leur réseau hydrographique, les paramètres ayant un impact significatif peuvent perdre leur pertinence et vice versa, lors du transfert sur d'autres régions. De plus, les erreurs relatives telles que celles des précipitations peuvent s'additionner (Kanso, 2004). Or, de telles erreurs peuvent biaiser la qualité des résultats obtenus et fausser les conclusions quant à la robustesse et à l'adaptabilité du modèle (Wilby, 2005; Kanso, 2004). En principe, dans le cas théorique, les paramètres obtenus par une méthode de combinaisons de différents bassins versants permettent souvent un ajustement des sorties du modèle aux données de calage comparativement aux méthodes de calage locales (Wriedt *et al.*, 2006). Cependant, cet avantage perd son efficacité lorsque les données réelles sont utilisées (Vase *et al.*, 2011). Avec une méthode de combinaison, contrairement à une méthode locale, la stabilité des paramètres est moins garantie lorsqu'on change de période (Rosero *et al.*, 2010). Cette stabilité est primordiale pour la transférabilité du modèle, puisque l'impact des paramètres peut être diminué ou augmenté. D'ailleurs, quelques auteurs (Wagener *et al.*, 2003; Wriedt *et al.*, 2006) détaillent différentes situations où la transférabilité des paramètres par combinaison de bassins versants hétérogènes peut engendrer une perte de robustesse. Vase *et al.* (2011) constatent que lors du

calage, les paramètres optimaux sur certains ensembles de bassins versants hétérogènes ne conviennent pas à d'autres. Dans le même contexte, Rosero *et al.* (2010) soulignent que lors du calage par combinaison des bassins hétérogènes, la transférabilité des paramètres optimaux est fortement influencée par les conditions climatiques. Cette dépendance des paramètres optimaux au climat a été également observée par Vaze *et al.* (2010); Merz *et al.* (2011); Andréassian *et al.* (2012). Ces derniers concluent que l'identification d'un jeu de paramètres jugé optimal durant le calage d'un modèle ne garantit pas toujours une reproduction fidèle des mêmes processus dans toutes les situations et en tout temps.

En somme, la vision de calage par combinaison de bassins versants différents ne paraît pas adéquate dans le cas de notre étude, car elle ne permet pas de quantifier l'impact réel des paramètres environnementaux et leur dynamique dans le déclenchement d'un processus (la contamination microbienne dans notre cas). À partir de ces considérations, deux bases de bassins versants ont été constituées. Un premier échantillon se compose de trois sous-bassins du lac Massawippi en Estrie, où la forte variabilité agricole et la taille ont permis d'effectuer les tests de validation locale. Un second échantillon se compose de sous-bassins de la rivière Yamaska situés en Montérégie où sera testée la transférabilité spatiale du modèle *PdIP*, dans d'autres régions dont la morphologie est différente de celle du BV du lac Massawippi.

3.7.5. Validation

La validation se fait sur deux niveaux. D'abord, une validation dite interne ou locale s'effectue sur les bassins versants de la Tomifobia (en utilisant les points qui n'ont pas servi de calage), de Brook et de McConnell. Ensuite une validation dite externe consiste à vérifier la capacité de reproductibilité et de transférabilité du modèle à d'autres zones géographiques avec le même type de variables.

3.8. Résultats et leur interprétation

3.8.1. Analyse de la dépendance conditionnelle

Les résultats du test d'indépendance sont présentés dans le tableau 4.

Tableau 4. Corrélations entre les variables prédictives

| Variabes | <i>Gradient</i> | <i>Nord-Ouest</i> | <i>Nord-Est</i> | <i>Dorsale</i> | <i>Plat</i> | <i>Vallonné</i> | <i>Drainage</i> | <i>Distance</i> | <i>Pluviométrie</i> |
|---------------------|-----------------|-------------------|-----------------|----------------|-------------|-----------------|-----------------|-----------------|---------------------|
| <i>Gradient</i> | 1 | | | | | | | | |
| <i>Nord-Ouest</i> | 0,018 | 1 | | | | | | | |
| <i>Nord-Est</i> | 0,212 | -0,77 | 1 | | | | | | |
| <i>Dorsale</i> | 0,011 | 0,021 | 0,15 | 1 | | | | | |
| <i>Plat</i> | 0,32 | 0,16 | 0,29 | 0,24 | 1 | | | | |
| <i>Vallonné</i> | 0,002 | -0,11 | 0,25 | 0,16 | 0,31 | 1 | | | |
| <i>Drainage</i> | 0,26 | 0,72 | 0,41 | 0,33 | 0,68 | 0,57 | 1 | | |
| <i>Distance</i> | -0,13 | 0,42 | 0,21 | -0,27 | 0,13 | 0,15 | 0,21 | 1 | |
| <i>Pluviométrie</i> | 0,37 | 0,12 | 0,36 | -0,19 | 0,33 | 0,08 | 0,01 | 0,33 | 1 |

Ces résultats révèlent certains liens significatifs entre les variables explicatives. Il ressort qu'il y a une très forte relation entre les deux variables de l'orientation de la pente (*Nord-Ouest* et *Nord-Est*). Par ailleurs, le test de χ^2 , montre une violation d'indépendance conditionnelle entre le drainage, l'orientation des pentes et les formes du terrain (*plat et vallonné*). Pour éliminer ce problème de dépendance conditionnelle, les variables explicatives corrélées n'ont pas été utilisées simultanément dans le processus de modélisation.

Une analyse fine sur la base de l'arbre de segmentation (figure 3) montre que, seules, trois variables peuvent contribuer activement à la prédiction de la contamination microbienne. Ces variables actives sont : la pluviométrie (*pluie*), la distance (*distance*) et la texture du sol (*texture*). Rappelons que, lors de la construction de l'arbre, les variables morphométriques et d'occupation du sol n'ont pas été utilisées. En principe, vu que les corrélations au niveau de l'arbre de segmentation sont basées sur des variables standardisées (variance totale correspond aux nombres de variables), la part explicative sur la variance totale de ces deux variables auraient été beaucoup trop importante. Cependant, ces dernières ont été prises en compte dans la suite des travaux.

Au niveau de la hiérarchie des variables explicatives obtenues par l'arbre de

segmentation, la VP *distance* est le facteur a semblé être le plus susceptible d'influer sur la contamination microbienne. On remarque que la première division associée à cette variable a divisé les échantillons suivants de données en trois sous-ensembles. Une analyse de sensibilité sur cette première division par la variable *distance* montre que la contamination des eaux peut connaître une certaine amplification lorsque la variable *distance* est faible ou moyenne (0-5 km et 5-10 km).

La variable *pluie*, quant à elle, est aussi apparue comme l'une des variables les plus actives dans le processus de partitionnement de l'ensemble des données puisqu'elle a servi pour la division au deuxième niveau. À ce niveau, cette variable apparaît encore plus en profondeur de l'arbre au dernier niveau. En principe, selon les résultats, la *pluie* permet de raffiner l'interprétation des effets de la distance sur la contamination microbienne des eaux au moyen de son interaction avec la texture du sol (*texture*). Nous pouvons d'ailleurs observer que l'effet de la variable *pluie*, n'est réellement significatif sur la contamination microbienne que si la variable *distance* est faible ou moyenne. Dans ce cas, la VP *pluie* intervient encore plus en profondeur de l'arbre comme facteur explicatif de la contamination microbienne, notamment au dernier niveau. Ces divisions portent toutes sur des échantillons pour lesquels la variable *texture* a trait au *sable* ou à l'*argile*. Dans le partitionnement de l'ensemble des données, l'implication active de la variable *texture* montre que ce dernier a également des effets significatifs dans le processus de prévision de la contamination microbienne. Cette variable est la plus discriminante dans le groupe d'échantillons localisés sur des valeurs de *distance* inférieures ou égales à 10 km.

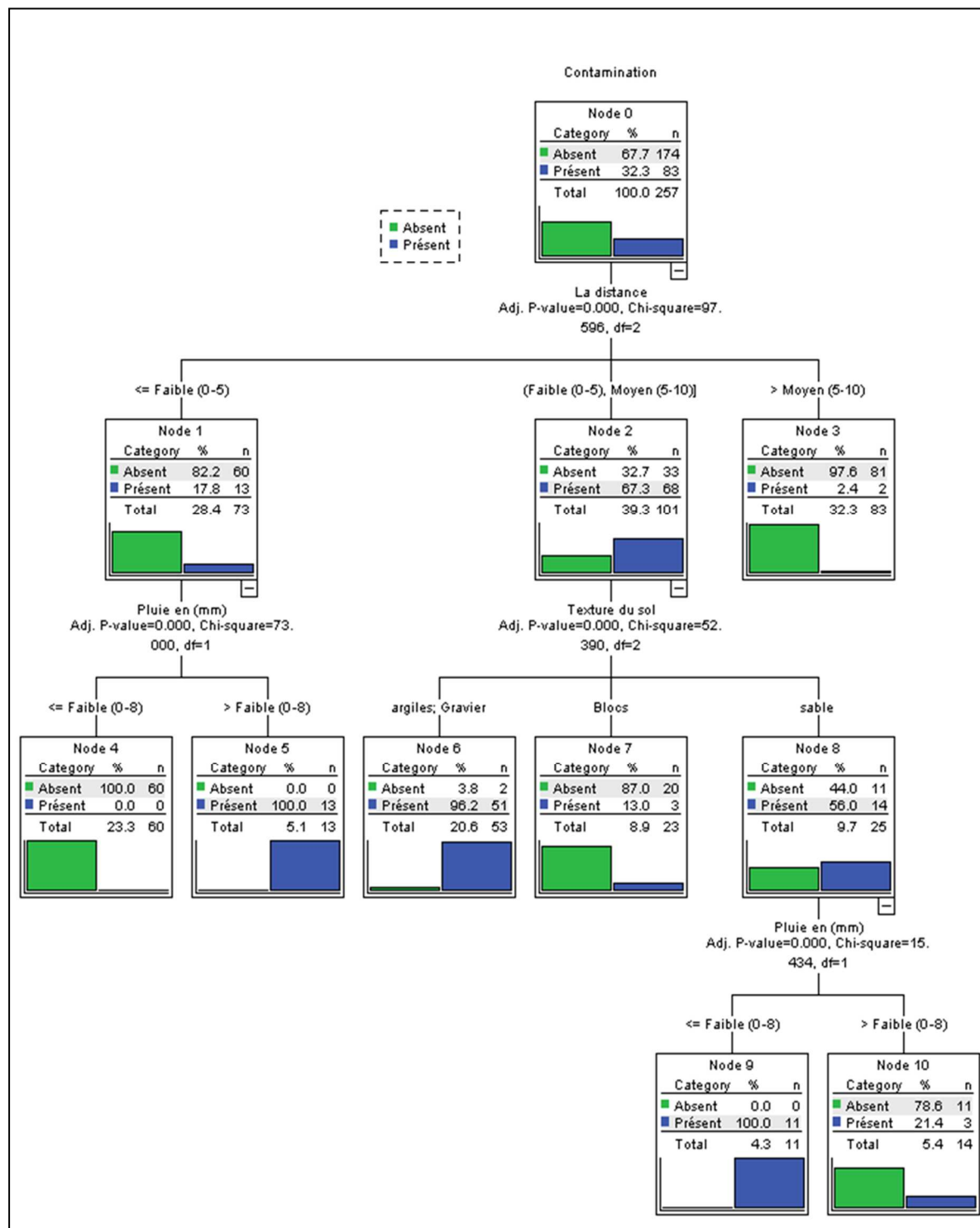


Figure 3. Arbre de régression de la sélection des variables explicatives de contamination microbienne des eaux.

3.8.2. Attribution des poids

Les résultats des poids positifs et négatifs ainsi que le contraste figurent dans le tableau 5.

Tableau 5. Corrélation spatiale entre les variables explicatives et la variable prédictive (bassin versant de la Tomifobia)

| Variables prédictives | Classe des preuves | W+ | W- | Contraste |
|--------------------------------|------------------------|-------|-------|--------------|
| Occupation du sol | forêt | -0,37 | 0,35 | -0,72 |
| | terre agricole | 0,77 | -0,07 | 0,84 |
| | zone urbaine | 0,34 | 0,01 | 0,33 |
| | sol nu | 0,36 | -0,03 | 0,39 |
| | autoroute | 0,01 | -0,21 | 0,22 |
| | zone humide | 0,4 | -0,06 | 0,46 |
| Gradient de pente | nulle | 0,31 | -0,02 | 0,33 |
| | faible | 0,1 | 0 | 0,1 |
| | douce | 0,01 | 0,01 | 0 |
| | Modérée | 0,4 | -0,13 | 0,53 |
| | forte | 0,67 | -0,09 | 0,76 |
| | excessive | 0,62 | 0,23 | 0,39 |
| Texture du sol | loam | 0,11 | -0,14 | 0,25 |
| | loam argileux | 0,69 | 0,16 | 0,53 |
| | loam sableux | 0,59 | 0,03 | 0,56 |
| | loam- sable fin | 0,62 | 0,05 | 0,57 |
| | organique | -0,37 | 0,13 | -0,5 |
| | textures variables | 0,01 | 0,03 | -0,02 |
| hauteurs de précipitation (cm) | faible (10,1-11,5) | 0,29 | -0,04 | 0,33 |
| | moyenne (11,5-12,2) | 0,41 | -0,04 | 0,45 |
| | forte (12,2-12,8) | 0,86 | 0,22 | 0,64 |
| | très forte (12,8-13,3) | 0,42 | -0,17 | 0,59 |
| distances (km) | faible (0-5) | 0,61 | 0,11 | 0,5 |
| | moyenne (5-10) | 0,59 | -0,28 | 0,87 |
| | forte (10 et plus) | 0,02 | 0,13 | -0,13 |

Au regard de ces résultats, pour le BV de la Tomifobia, la distance semble être un facteur déterminant dans l'estimation de la contamination microbienne. Une analyse plus détaillée des preuves de ce facteur montre que le contraste le plus élevé est celui de la classe moyenne 5-10 km avec une valeur de 0,87. Cette valeur décroît assez vite lorsque la distance augmente atteignant une valeur négative, soit -0,13. La tendance est la même pour le gradient de pente et les hauteurs de précipitation dont les contrastes optimaux atteignent les valeurs de 0,76 et 0,64 respectivement. Concernant l'occupation du sol, le paramètre *terre agricole* donne le meilleur contraste, avec une valeur de 0,84. En ce qui concerne la pédologie, les preuves interviennent avec des ordres de grandeur presque similaires. Les valeurs de ces dernières sont comprises entre -0,5 et 0,57. Cette observation montre que la pédologie est un facteur non négligeable dans

l'estimation des modèles expliquant la variabilité de la contamination microbienne. Au niveau de ce paramètre, les preuves *loam argileux* et *loam sableux* ont les valeurs de contraste les plus élevés atteignant respectivement de 0,56 et 0,57. Au vu de ces résultats, six preuves principales ont été sélectionnées pour la construction du modèle : *terre agricole*; *pente forte*; *loam-argileux*; *loam-sableux*; *forte précipitation* et *faible distance*.

À partir des paramètres présélectionnés, plusieurs tests de simulation ont été effectués sur les bassins versants de Tomifobia, McConnell et Brook. Dans un premier temps, seules, les classes associées à l'occupation du sol sont utilisées. Le second test en plus des classes de l'occupation du sol, contient les variables dérivées de la pluviométrie. À ce niveau, plusieurs simulations ont été effectuées en faisant varier la répartition de la pluie et en tenant compte ou non alternativement de la répartition des types de sol. Le troisième test incluait les deux variables précédentes en plus des variables de la pente. Enfin, le dernier test regroupait toutes les variables. En vue de simuler la contamination microbienne sur différents types de bassin versant, nous avons cherché à optimiser la procédure de calage afin de déterminer les paramètres optimaux pour chaque BV. Ces tests avaient pour objectif de garder les meilleures combinaisons permettant de mieux cartographier des zones vulnérables à la contamination microbienne des eaux. Ainsi, pour chacune des simulations considérées, nous avons évalué le degré de succès de la prédiction à partir des valeurs d'ASC. Dans le tableau 6, seules, les meilleures simulations sont présentées.

Tableau 6. Les valeurs d'ASC des meilleures simulations obtenues par le poids de la preuve.

| Modèles | Combinaison des preuves | Tomifobia | McConnell | Brook |
|----------|---|-------------|-------------|-------------|
| A | <i>AGRI</i> | 0,27 | 0,59 | 0,65 |
| B | <i>AGRI + PL_forte + DIST_Faible</i> | 0,71 | 0,22 | 0,38 |
| C | <i>AGRI + PL_forte + loam_sableux + Sol_nu</i> | 0,45 | 0,32 | 0,11 |
| D | <i>AGRI + PL_forte + GRAD_forte +DIST_Faible</i> | 0,73 | 0,13 | 0,21 |
| E | <i>AGRI + loam argileux</i> | 0,31 | 0,12 | 0,03 |
| F | <i>AGRI + DIST_Faible</i> | 0,21 | 0,52 | 0,58 |
| G | <i>PL_forte + DIST_Faible+ loam argileux</i> | 0,31 | 0,23 | 0,32 |
| H | <i>AGRI + loam argileux + DIST_Faible + PL_forte + GRAD_forte</i> | 0,36 | 0,15 | 0,28 |
| I | <i>AGRI + Sol_nu</i> | 0,25 | 0,36 | 0,39 |
| J | <i>AGRI + PL_forte</i> | 0,44 | 0,40 | 0,49 |

AGRI = terre agricole

PL_forte = pente forte

DIST_Faible = faible distance

PL_forte =forte précipitation

À l'instar des résultats du tableau 6, les performances moyennes obtenues des modèles A, B, D et F sont meilleures que celles obtenus avec les modèles C, E, G, H, I et J. Nous observons que la

valeur moyenne d'ASC varie de 0,27 à 0,65 pour le modèle A; de 0,22 à 0,71 pour le modèle B, de 0,13 à 0,73 pour le modèle D et de 0,21 à 0,58 pour le modèle F. Une analyse plus détaillée montre que les valeurs d'ASC sont uniquement élevées sur un seul BV, celui de la Tomifobia. Sur ce dernier, les valeurs moyennes d'ASC atteintes sont respectivement de 0,71 et 0,73 pour les modèles B et D. Cependant, nous remarquons une grande difficulté de ces deux modèles (B et D) à être transposés sur les sous-bassins de McConnell et Brook. Avec ces modèles, les valeurs d'ASC se stabilisent à 0,21 et 0,13 respectivement pour les sous-bassins de Brook et McConnell. Si l'utilisation des modèles B et D dégrade le critère d'ASC au niveau des sous-bassins versants de Brook et McConnell, il n'en ait pas de même pour le modèle A dont les performances augmentent pour ces deux bassins versants. Avec le modèle A, les valeurs d'ASC atteignent 0,59 et 0,65 respectivement pour les sous-bassins versants de Brook et McConnell. Dans l'ensemble, les validations locales permettent de mettre en évidence un certain nombre d'éléments :

- Le BV de McConnell présente les moins bons résultats avec des valeurs d'ASC faibles quelque soit le modèle employé.
- Le BV de la Tomifobia obtient les plus fortes valeurs d'ASC avec le modèle B.
- Le modèle A semble plus performant que les modèles B et D sur les sous-bassins versants de Brook et McConnell.

D'après ces résultats, nous pouvons déduire que si la variable «*Pente*» apporte une explication significative à la contamination microbienne au niveau du sous-bassin de la rivière Tomifobia, elle ne l'est pas de façon automatique pour les sous-bassins de Brook et McConnell (modèles B et D). D'ailleurs, dans ce contexte plusieurs auteurs (Hébert, 2001; Servais *et al.*, 2005) ont démontré que pour les bassins dont la taille est supérieure à 200 km² les variables spatiales telles que la pédologie et la topographie sont relativement homogènes; par conséquent, la pente et / ou l'occupation du sol permettent d'améliorer l'explication de la contamination des eaux. Alors que, pour les bassins dont la taille n'atteint pas 100 km², l'occupation du sol est assez suffisante pour expliquer la variation de la contamination microbienne.

Au vue de ces résultats, seuls, les modèles B et D ont été pris en considération et transférés sur les bassins versants de la Yamaska dont les superficies sont supérieures à 200 km². Le tableau 7 présente les résultats de la transposition des modèles B et D sur les sous-bassins versant de la Yamaska.

Tableau 7. Les valeurs d'ASC des meilleures simulations obtenues pour la transposition

| Modèles | Combinaison des preuves | Yamaska-Nord | Yamaska Sud-Est | Salvail | Pot au beurre |
|---------|---|--------------|-----------------|---------|---------------|
| B | <i>AGRI + PL_forte + DIST_Faible</i> | 0,76 | 0,75 | 0,62 | 0,54 |
| D | <i>AGRI + PL_forte + DIST_Faible + GRAD_forte</i> | 0,79 | 0,77 | 0,21 | 0,36 |

Nous pouvons constater que la transposition du modèle D reste meilleure par rapport au modèle B pour les bassins versants de Yamaska-Nord et de Yamaska Sud-Est. Les valeurs d'ASC sur ces bassins versants atteignent respectivement 0,79 et 0,77. Cependant, le modèle D perd sa performance lorsqu'elle est transposée sur les sous-bassins versants de Salvail et Pot-au-beurre. Cette perte est caractérisée par une baisse notable du taux de bonne prédiction dont les valeurs passent de 0,21 à 0,36 pour les sous-bassins versants de Salvail et de Pot-au-beurre.

En comparant les résultats par type de BV, nous constatons que ce sont les bassins versants des Appalaches (Yamaska-Nord et Yamaska Sud-Est) qui ont les valeurs d'ASC les plus élevées : 0,79 et 0,77 respectivement pour « Yamaska-Nord » et « Yamaska Sud-Est ». Cependant, le groupe des sous-bassins versants de basses terres (Salvail et Pot-au-beurre) ne suit pas cette règle, puisque les valeurs d'ASC obtenues sont inférieures à 70 %. Du point de vue descriptif, de telles valeurs d'ASC n'indiquent aucune concordance entre la carte simulée et les points d'observations sur le terrain. Cela rend la distribution instable et difficile à interpréter. Pour avoir une idée plus nette de la façon dont les deux modèles estiment la contamination microbienne, nous avons visuellement comparé les cartes de probabilités simulées et les points d'observation sur le terrain. Les résultats obtenus sont présentés aux figures 3 et 4.

Du point de vue descriptif, la carte de distribution de la contamination microbienne générée par le modèle B (figure 4) sous-estime certaines zones à fortes probabilités de contamination microbienne. C'est le cas du SBV Yamaska.

Bien que les valeurs d'ASC du modèle B soient inférieures à celle du modèle D, le modèle B présente de meilleures performances descriptives (figure 4). Ce modèle prédit plus de 70 % de la dynamique de la contamination microbienne aux niveaux des bassins des Appalaches et plus de 50 % pour les bassins versants des basses terres. Cette analyse comparative montre que le modèle B est celui qui peut être transposable sur les bassins versants de la région des Appalaches et celle des terres basses. Il explique non seulement une grande part de la variabilité de la contamination microbienne, mais aussi une concordance réaliste entre les cartes simulées et

les observations sur le terrain. Les mauvais résultats en validation proviennent majoritairement d'erreurs de paramètres calés trop faibles et moins représentatifs sur certains bassins versants. Malgré tout, le calage et la validation des modèles B et D ont permis de sélectionner un jeu de paramètres permettant de simuler au mieux la distribution de la contamination microbienne de chaque BV.

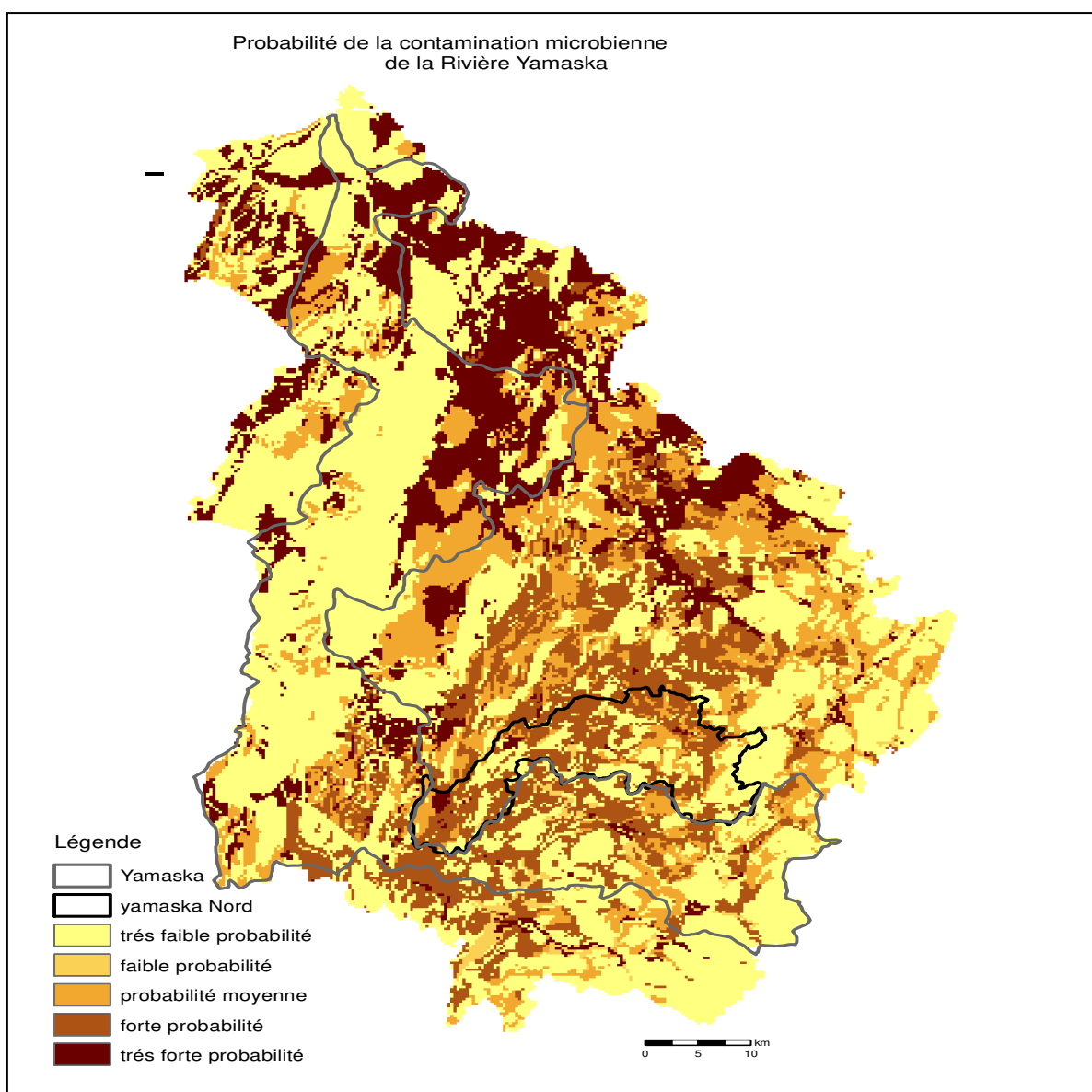
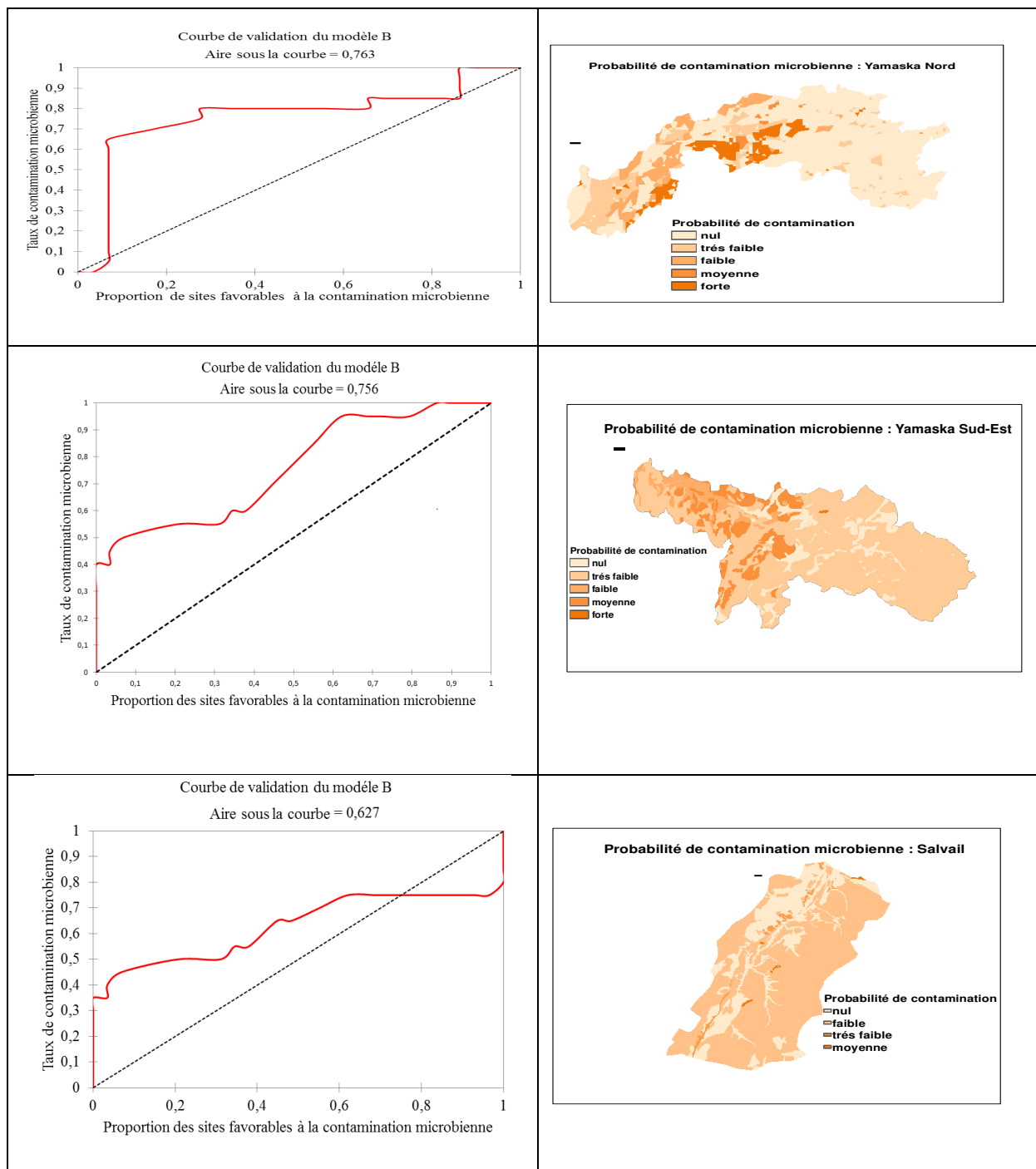


Figure 4. Probabilité de contamination microbienne sur la Yamaska à partir du modèle D



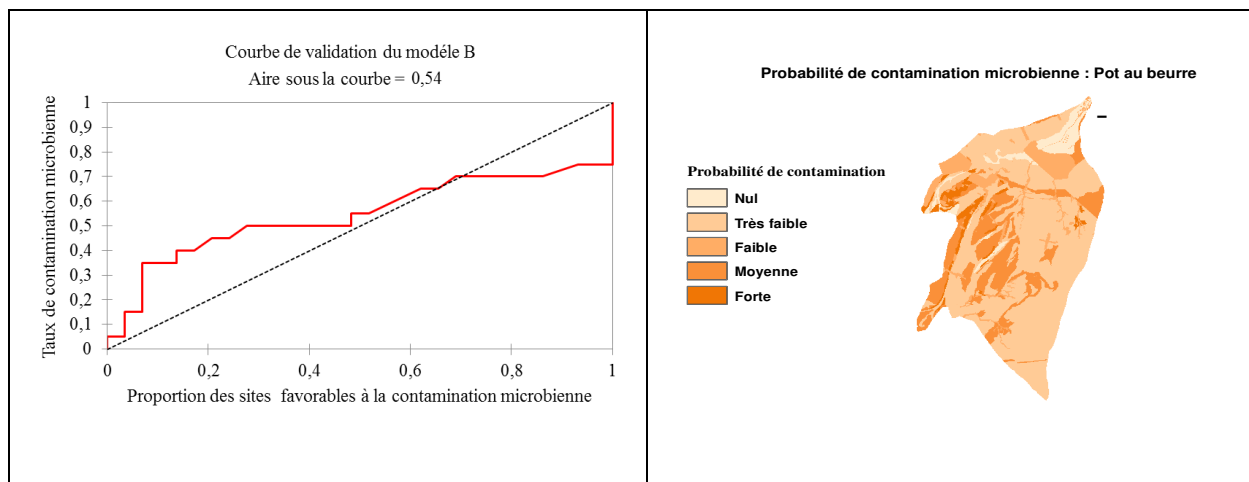


Figure 5. Évaluation explicative et descriptive des modèles

3.9. Conclusion partielle

Dans cette étude, nous avons évalué la performance de prédiction du modèle *PdIP* sur la contamination microbienne des eaux. Pour cela, nous avons réalisé une validation externe dans un contexte de généralisation globale. Dans la littérature, cette dernière a rarement été effectuée, car la plupart du temps, les modèles de contamination microbienne des eaux sont construits localement, et évalués dans la région où ils ont été calibrés. Un point fort de ce travail a été la mise en place d'un protocole d'évaluation mettant en exergue les données de calibration du modèle à une échelle inférieure à 1 000 km² et indépendants des données d'évaluation dans des régions de caractères géomorphologique et pédologique différents. Il a donc été possible de tester à la fois la capacité du modèle à prédire non seulement la contamination microbienne des eaux à l'échelle d'un BV, mais aussi la capacité d'être transposé sur d'autres bassins versants. Le présent modèle, avec une force de prédiction d'environ 71 % est assez fort si on tient compte des études antérieures (Pitts, 2002; Hubert, 2002; Pitoine *et al.*, 2012) qui mentionnent respectivement des forces de prédiction de la contamination microbienne d'environ 50 % selon le type de marée et 70 % à partir de SLAMM (Source Loading and Management Model).

Parmi les résultats obtenus, il ressort que les paramètres qui influencent le plus, la variabilité de contamination microbienne sont la distance des cultures au réseau hydrologique, l'occupation du sol, la pluie, ainsi que les caractéristiques pédologiques dans une moindre mesure. Dans ce contexte, la précision du modèle et la hiérarchisation des facteurs de la contamination microbienne des eaux sont similaires à ceux des modèles de régression (Patoine, 2009) de REP-PCR (Kon *et al.*, 2009) et des modèles d'analyse multivariée. Ces études ont montré que l'agriculture et la densité animale contribuent à plus de 59 % d'*E. coli* sur un BV de 100 km². Les mêmes conclusions ont été avancées par Turgeon (2011) qui a démontré une association significative entre le pourcentage des terres agricoles et la présence d'*E. coli*.

L'évaluation, la validation ainsi qu'une tentative de transposabilité des modèles proposés ont constitué une étape importante de ce travail. Dans la littérature, de nombreuses études testent les ajustements des modèles agro-hydrologiques aux variations de la contamination microbienne, mais très peu se sont attachées à vérifier la validité des modèles dans un but de transférabilité spatiale. En principe, certains auteurs (Hansen *et al.*, 2001) pensent que les interactions entre les facteurs environnementaux et la pollution sont complexes et compliquées pour être transférées

systematiquement sur toutes les échelles spatiales. D'autres, plus optimistes, affirment que la connaissance des facteurs affectant la contamination microbienne a déjà des principes bien définis qui sont suffisants pour généraliser la modélisation du flux de coliformes. À notre échelle de travail et sur la base des sept sites d'études, nous montrons qu'un certain degré de transposabilité géographique peut être fait sur les relations entre les paramètres environnementaux et la contamination des eaux. Pour des bassins versants dont la superficie est supérieure à 100 km² et inférieure à 1 000 km² le modèle *PdIP* est correctement généralisable. À cette échelle, il est possible de cartographier de manière réaliste la dynamique de la contamination microbienne. La tentative de transférabilité donne une certaine robustesse au modèle *PdIP* et de son utilité pour le contrôle de la qualité microbiologique des eaux récréatives.

Chapitre 4. Analyse spatiale de la sensibilité du modèle *PdlP*

La première partie de notre travail a exploré l'utilisation du modèle *PdlP* comme un outil prédictif de la contamination microbienne. Ce modèle a procuré des éléments de réponse sur l'identification et la localisation des zones susceptibles de dégrader la qualité bactériologique des eaux. Cette conclusion ne peut être renforcée que si l'incertitude due aux données spatiales et temporelles de ce modèle est prise en compte. Le but de cette section est donc d'analyser la sensibilité de *PdlP* à travers l'incertitude des variables spatiales utilisées comme intrant au modèle. Il s'agit ici de quantifier le domaine de variabilité des sorties du modèle *PdlP*, en déterminant quelle part de variance est due à l'occupation du sol et aux caractéristiques pédologiques. Le travail d'analyse de sensibilité globale sur le modèle *PdlP* permet d'identifier les variables explicatives dont l'effet est peu important en fixant leur valeur, et de réduire l'incertitude dans la prédiction de la contamination microbienne. La littérature existante sur les modèles de la qualité de l'eau procure des analyses de sensibilité, mais elles se résument principalement à étudier les variations d'une sortie d'intérêt en faisant varier à tour de rôle les valeurs des variables explicatives autour de leur valeur nominale. Ce type d'analyse ne permet pas de comprendre et d'appréhender la sensibilité des variables sur la qualité des eaux de surface. C'est pourquoi, nous proposons l'utilisation d'une nouvelle approche fondée sur la décomposition de la variance pour l'analyse de sensibilité globale du modèle *PdlP*. Il s'agit des indices de sensibilité de Sobol.

4.1. Introduction

La modélisation de la contamination microbienne à partir du modèle *PdlP* est issue des considérations probabilistes et repose entièrement sur un large éventail de données d'entrée spatiales, notamment la topographie du terrain, le réseau hydrographique, la carte d'occupation du sol ainsi que les caractéristiques pédologiques. Chacune de ces données contient un niveau d'incertitudes liées aux erreurs d'acquisition, de traitement et d'échelle (Corgne, 2004; Maggioni *et al.*, 2012). À ces incertitudes s'ajoutent celles du modèle liées à la simplification des structures complexes, à la description mathématique ou à la discrétisation spatiale du modèle (Uhlenbrook *et al.*, 1999; Kanso, 2004). L'ensemble de ces incertitudes peut réduire les capacités prédictives de la modélisation. Devant l'incertitude des variables d'entrée, il est impératif d'effectuer une analyse de sensibilité du modèle pour s'assurer de la fiabilité, de la robustesse et surtout de la

représentativité des prédictions. Le contexte dans lequel nous nous intéressons à cette incertitude, est celui de l'analyse de sensibilité permettant de mieux comprendre le comportement du modèle, face aux variations de valeurs des variables d'entrée. Cette analyse permet donc d'orienter la sélection de variables explicatives et la validation du modèle en définissant l'intervalle de variations des variables sur lesquelles il faut agir pour que la sortie du modèle soit dans une gamme bien précise (Hamby, 1994; Murphy, 2004; Hasmadi *et al.*, 2008).

Il existe différentes méthodes permettant de quantifier des incertitudes sur les variables d'entrée spatiales. Ces méthodes, largement décrites dans Kleijnen. (1995) et Saltelli *et al.* (2000) peuvent se répartir en trois grandes catégories : (i) les méthodes de criblage ou « screening »; (ii) les méthodes d'analyse de sensibilité locale; et (iii) les méthodes d'analyse de sensibilité globale.

Les méthodes de criblage consistent à hiérarchiser les incertitudes des variables d'entrée en fonction de leur influence sur la variabilité de la sortie du modèle (Campolongo *et al.*, 2007; Weichel *et al.*, 2007; Carlos García-Díaz *et al.*, 2012). Ce sont des méthodes considérées locales car elles permettent d'estimer localement les dérivées partielles des variables vis-à-vis des paramètres incertains, et de les normaliser par l'espérance ou l'écart-type des différentes variables (Homma et Saltelli, 1996). Cette approche est appropriée si, seule, une région spécifique de l'espace doit être étudiée. Par cette approche, la sensibilité est mesurée en analysant les variations de la sortie lorsqu'une seule variable d'entrée varie, les autres étant fixées à leurs moyennes. C'est pourquoi elle sert, en première approche, à détecter les variables les plus influentes d'un modèle, pour pouvoir ensuite les étudier par des méthodes plus approfondies. Dans la plupart des cas, ces méthodes sont utilisées lorsque le nombre de variables est très grand (>100). Toutefois, il s'agit des méthodes très restrictives. Elles ne représentent pas toute l'interaction entre les variables du modèle. Par ailleurs, elles fournissent uniquement une information qualitative contrairement aux méthodes d'analyse de sensibilité locale et globale qui permettent de quantifier et classer les variables par ordre d'importance (Hamby, 1994; O'Hagan, 2012).

Pour sa part, l'analyse de sensibilité locale détermine la réponse du modèle à une variation des variables d'entrée, au voisinage d'une configuration paramétrique donnée (Hamby, 1994; Mishra *et al.*, 2003). La mise en place de cette méthode repose principalement sur le calcul des dérivées partielles des fonctions de sortie par rapport aux variables d'entrée. Les calculs

numériques impliquent de faire varier les variables d'entrée du modèle dans un intervalle très restreint autour d'une valeur nominale (Mara *et al.*, 2012; Marrel *et al.*, 2012). Dans ce contexte, l'approche la plus utilisée est celle dite *Un-à-la-fois*, (en anglais *One-At-a-Time* ou OAT), où une seule variable d'entrée varie à la fois, alors que toutes les autres sont maintenues constantes (Turányi et Rabitz, 2004). L'inconvénient majeur est que cette catégorie d'analyse de sensibilité ne permet pas de prendre en compte la variabilité totale des variables, ni leurs dépendances (Saltelli *et al.*, 2010). Par conséquent, les résultats obtenus par une analyse de sensibilité locale ne reflètent généralement pas la complexité du modèle considéré puisqu'ils sont obtenus sans tenir compte des distributions des variables d'entrée, de leurs intervalles de valeur et de leurs interactions.

Complémentaire aux méthodes locales, l'analyse de sensibilité globale est fondée sur une méthode probabiliste considérant chacune des variables comme étant aléatoire. Elle se caractérise par l'exploration de l'étendue admissible des variables d'entrée et permet de prendre en compte la distribution de probabilité de chaque entrée en faisant varier simultanément toutes les variables (Xu et Gertner, 2011; Kucherenko *et al.*, 2012). Dans l'analyse de sensibilité globale, on distingue principalement trois techniques : Monte Carlo (MC), Fourier Amplitude Sensivity Test (FAST) et les indices de Sobol. Ces méthodes se caractérisent par l'exploration de l'espace des variables d'entrée et sont basées sur la variabilité de la sortie du modèle dans son domaine de variation. Parmi ces techniques, les indices de Sobol sont les plus utilisés pour des analyses de modèles hydrologiques. La raison principale de cette popularité repose sur l'attribution d'une loi de probabilité aux variables d'entrée afin de modéliser l'incertitude sur chaque paramètre d'entrée. Basés sur la décomposition de la variance des valeurs de sortie, les indices de Sobol permettent de quantifier la sensibilité du modèle non seulement aux interactions paramétriques des variables temporelles ou spatiales, mais ils prennent en compte l'espace global des incertitudes sur les variables (Marrel *et al.*, 2012, Storlie *et al.*, 2009; Tissot, 2012). Ces indices permettent d'estimer l'influence des paramètres d'entrée sur la sortie du modèle en utilisant la décomposition de la variance fonctionnelle (appelée aussi analyse de variance fonctionnelle) en somme de fonctions élémentaires.

Dans le but d'explorer son potentiel, les indices de Sobol ont été utilisés par Cibin *et al.* (2010) et Nossent *et al.* (2011). Le but de ces études étaient d'analyser la sensibilité globale du modèle SWAT aux propriétés du sol de deux BV en considérant différents changements

climatiques. Il ressort de ces études que, vu la difficulté de paramétrisation que nécessitent les modèles hydrologiques, les indices de Sobol sont assez robustes pour repérer les principales sources d'incertitude du modèle. Ils sont en mesure d'identifier et de hiérarchiser les variables d'entrée les plus influentes. Ces indices sont également en mesure de diminuer l'incertitude des entrées. Un autre avantage qu'on attribue aux indices de Sobol est qu'ils sont peu coûteux en calcul et ne nécessitent aucune hypothèse sur le modèle considéré. Au vu de ces avantages, les indices de Sobol constituent une voie prometteuse pour l'analyse de sensibilité globale des modèles utilisant les variables spatiales comme intrants.

4.2. Objectifs

L'objectif de ce chapitre est de quantifier les incertitudes liées aux simulations de la contamination microbienne des eaux et d'effectuer une analyse de sensibilité du modèle aux variables explicatives à l'aide de la méthode de Sobol. L'enjeu principal est la compréhension du comportement de *PdlP* aux changements de variables et à travers différents scénarios, pour caractériser sa réponse dans un intervalle de confiance en tenant compte des incertitudes associées aux variables spatiales. Compte tenu de l'aspect aléatoire du risque de contamination microbienne et de l'aspect aléatoire du comportement des variables spatiales, il apparaît nécessaire d'introduire une dimension stochastique dans le modèle *PdlP*. Pour ce faire, il faut :

- quantifier l'impact des incertitudes associées aux variables d'entrée spatiales sur la prédiction de la contamination microbienne des eaux et leur propagation dans la structure du modèle *PdlP*;
- déterminer le degré de confiance à accorder au modèle *PdlP* dans la simulation de la contamination microbienne des eaux, face aux incertitudes de nature spatiale.

4.3. Cadre mathématique : principe d'analyse de sensibilité de Sobol

La méthode de Sobol est une technique d'analyse de sensibilité globale qui consiste à déterminer l'impact de la variabilité des variables d'entrée du modèle sur les valeurs de sortie (Jacques, 2011). Initialement, l'application de cette méthode se concentrait sur des cas où les variables d'entrée sont scalaires, tout comme la sortie. Basée sur le théorème de Hoeffding, cette méthode stipule que la variance des valeurs de sortie peut être décomposée de façon unique en somme de termes de dimensions croissantes soumis à une contrainte d'orthogonalité (Saltelli *et al.*, 2010).

De façon concrète, il s'agit d'estimer l'influence des variables d'entrée sur la sortie du modèle en utilisant la décomposition de la variance fonctionnelle en somme de fonctions élémentaires (Saltelli et Annoni, 2010; Saltelli *et al.*, 2010; Sudret, 2008).

Supposons que le modèle étudié est sous la forme $Y = f(X_1, \dots, X_p)$ avec (X_1, \dots, X_p) , l'ensemble des variables d'entrées incertaines du modèle; Y est la sortie et P le nombre de variables indépendantes. Pour quantifier l'importance d'une variable d'entrée sur la variance de la sortie Y , en fixant la variable X_i à une valeur x_i^* il faut calculer la variance conditionnelle de Y sachant que $X_i = x_i^*$ notée $V(Y|X_i = x_i^*)$. Or, selon le théorème de la variance totale (Saporta, 2006), la variance de Y est liée à la variance et l'espérance conditionnelles de X selon l'équation suivante :

$$V(Y) = V(E[Y|X_i]) + E[V(Y|X_i)]. \quad \text{Éq. 10}$$

Le premier terme représente la variance de l'espérance de Y conditionnellement à X_i . Alors, pour pouvoir quantifier l'effet de X_i sur Y , on évalue la part de la variance de Y induite par X_i . Cette quantité est définie par Sobol (1993) comme étant l'indice de sensibilité S_i de premier ordre:

$$S_i = \frac{V(E[Y|X_i])}{V(Y)} \quad \text{Éq. 11}$$

Cet indice S_i est appelé indice de sensibilité de premier ordre par Sobol (1993) et « *correlation ratio* » ou encore « *importance measure* » par McKay (McKay, 1979). Il quantifie la sensibilité de la sortie Y à la variable d'entrée X_i , ou encore la part de variance de Y due à la variable X_i . Cet indice a été introduit par décomposition de la fonction f du modèle en somme de fonctions de dimensions croissantes :

$$f(x_1, \dots, x_p) = f_0 + \sum_{i=1}^p f_i(x_i) + \sum_{1 \leq i < j \leq p} f_{i,j}(x_i, x_j) + \dots + f_{1,2,3,\dots,p}(x_1, \dots, x_p) \quad \text{Éq. 12}$$

où

$$f_0 = E[Y];$$

$$f_i(X_i) = E[Y|X_i] - E[Y]; \quad f_{i,j}(X_i, X_j) = E[Y|X_i, X_j] - E[Y|X_i] - [Y|X_j] + E[Y],$$

suit de la composante d'ordre supérieure ($f_{1,2,3,\dots,p}(x_1, \dots, x_p)$). Suivant le même principe de l'équation 12 la variance se décompose alors en :

$$V = \sum_{i=1}^p V_i + \sum_{1 \leq i < j \leq p} V_{ij} + \dots + V_{1\dots p}, \quad \text{Éq.13}$$

où

$$V_i = V(E[Y|X_i]),$$

$$V_{ij} = V(E[Y|X_i, X_j]) - V_i - V_j,$$

$$V_{ijk} = V(E[Y|X_i, X_j, X_k]) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k$$

$$V_{ijk\dots p} = V - \sum_{i=1}^p V_i - \sum_{1 \leq i < j \leq p} V_{ij} - \dots - \sum_{1 \leq i_1 < \dots < i_{p-1} \leq p} V_{i_1 \dots i_{p-1}}$$

Le dernier terme de la décomposition $V_{1\dots p}$, n'est que la différence entre la variance de Y et toutes les parts de variance d'ordre inférieur. Si on divise chaque membre de l'équation 13 par $V(Y)$ on obtient :

$$\sum_i S_i + \sum_{i < j} S_{ij} + \sum_{i < j < k} S_{ijk} + \dots + S_{ijk\dots p} = 1 \quad \text{Éq.14}$$

où

$$S_i = \frac{V_i}{V} = \frac{V(E[Y|X_i])}{V} \quad \text{Éq.15}$$

$$S_{ij} = \frac{V_{ij}}{V} \quad \text{Éq.16}$$

$$S_{ijk} = \frac{V_{ijk}}{V} \quad \text{Éq.17}$$

Les équations 15, 16 et 17 représentent respectivement:

- l'indice du premier ordre qui quantifie la sensibilité de la sortie Y par rapport aux variables d'entrée X_i ;
- les indices de sensibilité d'ordre deux permettant de quantifier la sensibilité de la sortie Y par rapport à l'interaction croisée de X_i et X_j sans prendre en compte l'effet des variables seules; et
- les indices de sensibilité d'ordre trois qui permettent de quantifier la sensibilité de la sortie Y par rapport à l'interaction croisée de X_i et X_j et X_{ij} sans prendre en compte l'effet des variables seules. Le principe se poursuit ainsi de suite jusqu'à l'ordre p .

La somme de tous les indices est égale à 1. Puisque tous les indices sont positifs, plus la valeur

de l'indice de Sobol est grande (proche de 1), plus la variable aura d'importance. Le nombre d'indices de sensibilité dépendra donc de l'ordre utilisé (1 à p) puisqu'il est égal à $2(p - 1)$. Si le nombre de variables d'entrée p est trop important, on obtient un très grand nombre d'indices de sensibilité; par conséquent, leur estimation peut se compliquer (Jacques, 2005). Pour ce faire, Homma et Saltelli (1996) ont alors introduit des indices de sensibilité totaux qui correspondent à la somme de la sensibilité de la seule variable et à la sensibilité aux interactions de cette variable avec toutes les autres. Elle est définie par l'équation suivante :

$$S_{t1} = S_{ti} = S_i + S_{ij} + S_{ik} + \dots + S_{ijk} + \dots \quad \text{Éq. 18}$$

Ces indices de sensibilité ont l'avantage de n'imposer aucune hypothèse sur la forme du modèle, mais nécessitent une hypothèse d'indépendance des variables d'entrée.

4.4. L'échantillonnage

Pour évaluer les indices de Sobol, on s'appuie souvent sur la création d'échantillonnage dans l'espace des paramètres d'entrées (Iooss, 2011). Cette étape est fondamentale pour une bonne analyse de sensibilité, car elle permet de déterminer la confiance accordée aux estimations de l'incertitude par de nombreux essais expérimentaux. Plusieurs techniques d'échantillonnage ont été proposées dans la littérature (McKay *et al.*, 1979; Stein, 1987; Fishman et Rubbin, 1992(a); Frey et Burmaster, 1999; Saltelli *et al.*, 2010). Parmi les plus populaires, on distingue : l'échantillonnage aléatoire de Monte Carlo (MC), l'échantillonnage stratifié et l'échantillonnage selon les hypercubes latins.

La stratégie d'échantillonnage aléatoire de MC est la plus utilisée dans des études d'analyse de sensibilité (Saltelli *et al.*, 2010). La majorité des auteurs s'accordent à l'idée que l'avantage de cette approche réside dans la simplicité de sa mise en œuvre. Toutefois, le temps de calcul nécessaire en constitue l'inconvénient principal (Hamby, 1994; Christiaens et Feyen, 2002). Par ailleurs, la méthode nécessite beaucoup de données pour l'attribution d'une fonction de densité de probabilité de chaque variable choisie (Guyonnet *et al.*, 1999, 2003). Enfin, à partir d'un échantillonnage de ce type, il y a un risque que les réalisations de l'échantillon utilisé soient focalisées sur une certaine région au dépend d'autres faiblement sondées (Geours, 2011). Pour cela, une alternative efficace consiste à utiliser des techniques de génération de nombres pseudo-aléatoires qui permettent d'obtenir une meilleure précision pour un même nombre de points.

Une de ces méthodes développée par McKay *et al.* (1979) est l'échantillonnage par hypercubes latins, connue sous l'acronyme LHS pour Latin Hypercube Sampling. Ce dernier consiste à répartir uniformément les points de l'échantillon sur toute l'étendue du domaine de chaque variable d'entrée. L'échantillonnage de l'Hypercube Latin sélectionne ainsi n valeurs distinctes pour chacune des k variables $X_1, X_2; \dots X_k$. Dans un premier temps, le domaine de chaque variable est divisé en n intervalles disjoints. Dans un second temps, on choisit aléatoirement dans chaque sous-intervalle une valeur. Les n valeurs choisies pour X_1 sont couplées aléatoirement avec les n valeurs de X_2 . Ces n paires sont combinées avec les n valeurs de X_3 pour former n triplets, etc. jusqu'à ce que n k -uplets soient formés. Cette construction de l'échantillonnage garantit que l'on n'a qu'une valeur dans chaque sous-intervalle de chaque direction de l'espace. L'avantage de ce type d'échantillonnage par rapport à la méthode classique de MC est qu'elle réduit le nombre d'échantillons pour une précision similaire. Elle garantit également une couverture uniforme du domaine de chaque entrée, en évitant de sous-échantillonner certaines zones. La figure 6, illustre le principe de l'échantillonnage LHS.

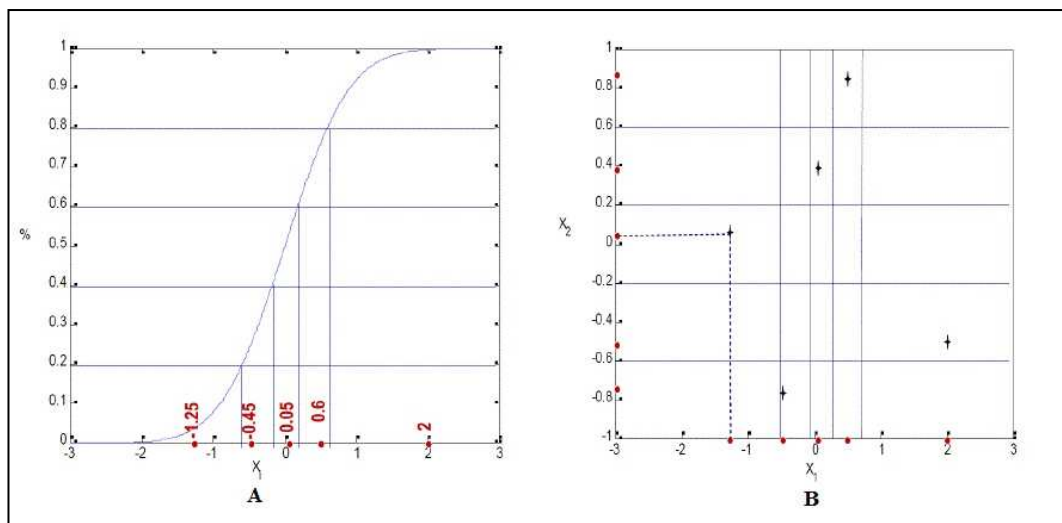


Figure 6. Illustration du principe de LHS: A) le découpage des 5 sous-intervalles selon la fonction de distribution d'une des entrées, en B) représentation des 5 couples générés (adaptée de Mara, 2008)

Dans la littérature plusieurs chercheurs se sont intéressés à comparer l'échantillonnage de MC au LHS (Saltelli *et al.*, 2010; Pételet *et al.*, 2006.). Il ressort que le LHS, qui présente l'avantage de la simplicité, est le mieux adapté lorsqu'on ne dispose d'aucune information sur la relation entrée/sortie. Comparativement à l'échantillonnage de MC cette technique donne

asymptotiquement un estimateur présentant un écart-type plus petit. Au vu de ces avantages, cette méthode est donc celle qui a été retenue dans notre étude afin de réaliser l'échantillonnage des variables considérées incertaines dans le modèle *PdIP*.

4.5. L'estimation des indices de sensibilité selon la technique de Sobol

L'estimation des indices de sensibilité nécessite l'estimation d'espérance de variance conditionnelle. Selon la technique d'estimation par la méthode de Sobol, il faut estimer la quantité V_i selon l'équation suivante :

$$V_i = V(E[Y|X_i]) = \underbrace{[E[Y|X_i]^2]}_{U_i} - [E[Y|X_i]]^2 = U_i - E[Y]^2, \quad \text{Éq. 18}$$

Sobol propose d'estimer U_i (l'espérance du carré de l'espérance de Y en tenant compte de la condition de X_i) et en faisant varier toutes les variables sauf la variable X_i . Pour cela, on génère deux matrices A et B de taille (N, p). N est la taille de l'échantillonnage et k le nombre de variables.

$$A = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_i^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_i^{(2)} & \dots & x_k^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{(N-1)} & x_2^{(N-1)} & \dots & x_i^{(N-1)} & \dots & x_k^{(N-1)} \\ x_1^{(N)} & x_2^{(N)} & \dots & x_i^{(N)} & \dots & x_k^{(N)} \end{bmatrix}$$

$$B = \begin{bmatrix} y_1^{(1)} & y_2^{(1)} & \dots & y_i^{(1)} & \dots & y_k^{(1)} \\ y_1^{(2)} & y_2^{(2)} & \dots & y_i^{(2)} & \dots & y_k^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_1^{(N-1)} & y_2^{(N-1)} & \dots & y_i^{(N-1)} & \dots & y_k^{(N-1)} \\ y_1^{(N)} & y_2^{(N)} & \dots & y_i^{(N)} & \dots & y_k^{(N)} \end{bmatrix}$$

$$A_B = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & y_i^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & y_i^{(2)} & \dots & x_k^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{(N-1)} & x_2^{(N-1)} & \dots & y_i^{(N-1)} & \dots & x_k^{(N-1)} \\ x_1^{(N)} & x_2^{(N)} & \dots & y_i^{(N)} & \dots & x_k^{(N)} \end{bmatrix}$$

La matrice A_B est formée par toutes les colonnes de A exceptée la $i^{\text{ème}}$ colonne qui est prise de B. C'est l'inverse pour la matrice B_A où toutes les colonnes de B exceptée la $i^{\text{ème}}$ colonne qui est prise de A. Partant de ces matrices, les indices de sensibilité de premier ordre sont alors estimés par :

$$V_i = V(E[Y|X_i]) = \frac{1}{N} \sum_{j=1}^N f(A)_j f(B_A^i)_j - f_0^2 \quad \text{Éq. 19}$$

Les indices totaux sont égaux à :

$$ST_i = 1 - \frac{\frac{1}{N} \sum_{j=1}^N f(A)_j f(B_A^i)_j - f_0^2}{\frac{1}{N} \sum_{j=1}^N f(A)_j^2 - f_0^2} \quad \text{Éq. 20}$$

4.6. Choix de l'approche de SOBOL dans le cadre de la contamination microbienne des eaux récréatives

L'originalité de notre travail de thèse a été d'adapter l'analyse des indices de SOBOL à un modèle dont les entrées sont des cartes et de montrer ses avantages potentiels dans la réduction d'incertitudes et l'amélioration des résultats, en combinant notamment méthodes statistiques (technique Bayésienne et analyse de sensibilité de premier ordre et indices totaux) et outils informatiques (couplage entre R et logiciel statistique SENSE (Sensitive analysis). En dépit de la pertinence de la calibration bayésienne, les indices de SOBOL ont été rarement appliquée au domaine de recherche sur les relations entre les variables environnementaux et la qualité de l'eau, en particulier en utilisant les modèles spatiales qui simulent la contamination des coliformes à l'échelle d'un bassin versant a vocation agricole. Cela nous a conduits à produire un nouveau paramétrage pour des sites du bassin versant de la Yamaska soumis à des conditions agricole extrêmes et topologiques variées. Les résultats ont confirmé l'efficacité de cette technique pour la réduction d'incertitude et l'amélioration de la fiabilité du modèle spatiale du *PdIP*.

L'un des points forts de la calibration bayésienne est qu'elle fournit une distribution a posteriori pour chaque paramètre et non pas seulement une valeur unique. Cette distribution nous permet de mieux étudier et de quantifier l'incertitude liée aux variables du modèle. Il s'agit d'une méthode simple et performante en raison de son adaptation aux modèles complexes basés sur les processus et de toutes tailles (plus particulièrement les modèles de ≥ 10 paramètres). Par ailleurs, les indices de SOBOL ne nécessitent pas de connaissances préalables sur la forme de la distribution *a posteriori* à générer. Tout cela prouve la pertinence des indices de SOBOL qui sont

utilisées dans nombreuses disciplines grâce à leur capacité de résoudre plus facilement et plus naturellement les problèmes. En plus, ces indices permettent de combiner le paramétrage du modèle avec l'analyse d'incertitude. Ce qui permet d'intégrer et de synthétiser plus facilement diverses sources d'informations.

Les indices de SOBOL n'ont pas vraiment besoin d'un logiciel ou outil spécifique pour être mises en place. Ils peuvent être développée sous plusieurs environnements et outils génériques de calcul scientifique (ex. R, MATLAB, SENSE) ou codé en langages informatiques ayant des bibliothèques statistiques (ex. C/C++, Java...). Ces atouts majeurs les distinguent des autres approches permettant principalement d'utiliser la nouvelle distribution *a posteriori* des paramètres pour obtenir des nouvelles valeurs plus précises et moins incertaines (Van Oijen *et al.*, 2005).

4.7. Méthodologie

L'analyse de sensibilité globale du modèle *PdIP* aux variables spatiales a été effectuée à partir de la méthode de Sobol. L'approche proposée a pour but de prendre en compte à la fois le caractère aléatoire du taux moyen de coliformes dans les eaux et l'incertitude des variables spatiales définissant leur caractère aléatoire. Cette démarche, scindée en trois étapes, est schématisée en figure 7.

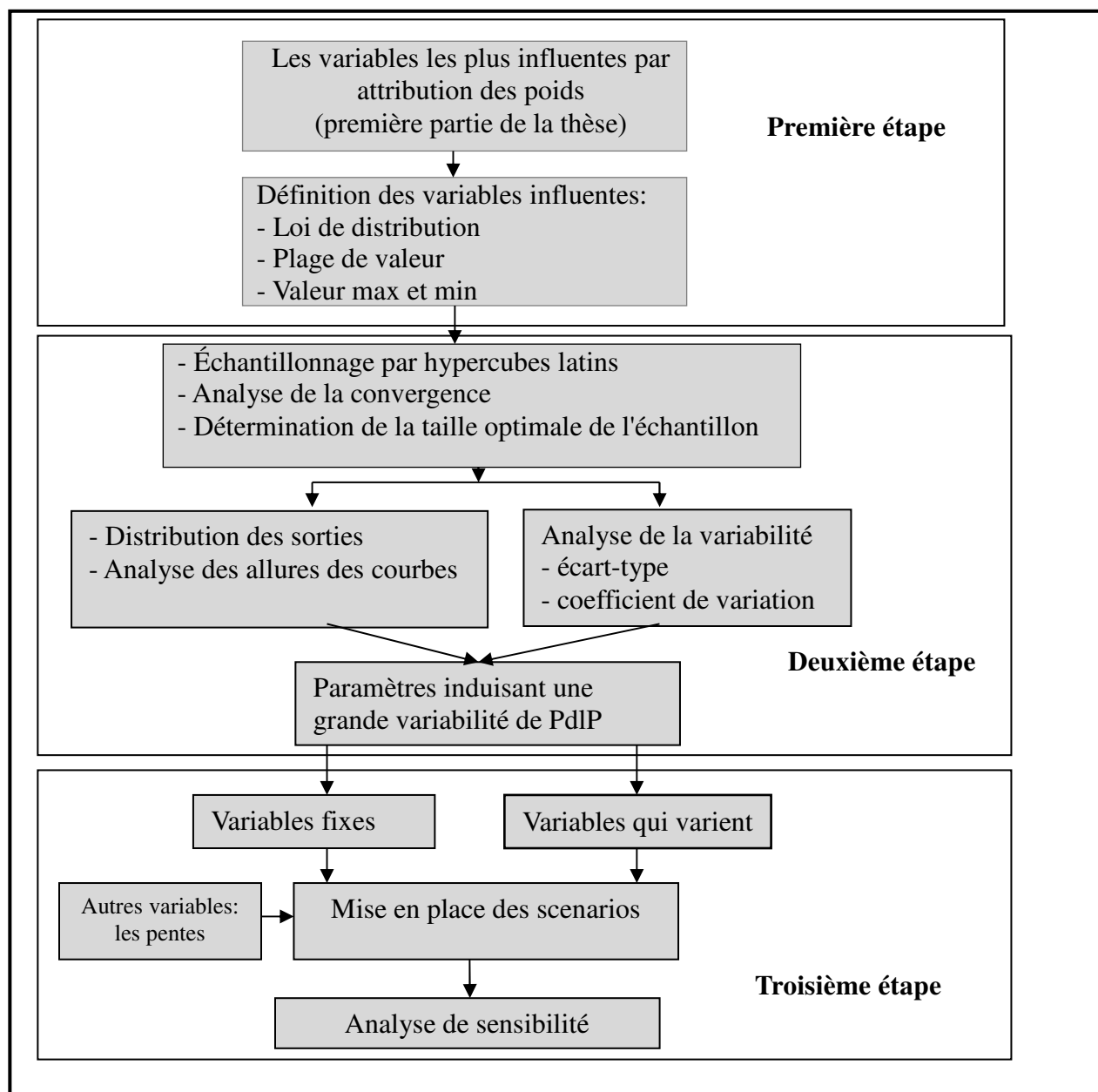


Figure 7. La méthodologie de l'analyse de sensibilité de *PdIP* aux variables spatiales

La première étape de la méthodologie a consisté tout d'abord à sélectionner les principales variables spatiales ayant un impact significatif sur la contamination microbienne des eaux. Dans la majorité des cas, cette sélection s'est effectuée à l'aide du jugement d'experts, d'études analogues, ou de résultats d'études précédentes. Dans cette partie de notre étude, la sélection des variables a été effectuée selon les résultats de la première partie de cette thèse de

doctorat dont l'objectif a été de quantifier l'influence de chaque variable sur la contamination microbienne en leur assignant un poids. Suite à la sélection des variables spatiales les plus influentes, nous avons défini non seulement les différentes valeurs que celles-ci peuvent prendre mais aussi leurs lois de probabilités.

La deuxième étape est la propagation de l'incertitude. Cette étape a permis de déterminer la variabilité de la réponse de *PdIP* induite par les incertitudes des variables d'entrée spatiales, et donc d'exprimer un intervalle de confiance sur des résultats de simulation. Pour cela, le domaine de chacune des variables spatiales définies à la première étape, a été échantillonné à l'aide des plans LHS. Plusieurs tirages aléatoires de tailles différentes ont été réalisés afin d'apprécier le gain de précision réalisé avec l'augmentation de la taille. Ainsi, l'on fait varier la variable aléatoire dont l'influence est étudiée en suivant une distribution préalablement définie tandis que le reste des variables reste fixe. Les résultats ainsi obtenus des tirages ont permis ensuite de construire un histogramme de distribution du taux de coliformes en fonction de la variabilité de chaque variable. Cet histogramme constitue un outil visuel pour appréhender non seulement la distribution du taux de contamination pour une simulation, mais aussi leurs intervalles de variation ainsi que les fréquences relatives des différentes variables. Pour chaque variable, nous avons obtenu une distribution des valeurs représentant son effet pour différents scénarios. Ces distributions sont caractérisées par la moyenne (μ), l'écart-type (σ), les coefficients de variation (COV), les quartiles 2,5 et 97,5 ($Q_{2,5}$ et $Q_{97,5}$).

La troisième et la dernière étape est l'analyse de sensibilité qui consiste à étudier l'impact de la variabilité des entrées spatiales choisies sur la variabilité de *PdIP*. Il a été question d'estimer la variance (la part de la dispersion) du taux moyen des coliformes en fonction de chaque variable aléatoire à partir de la méthode de Sobol. Ainsi, pour chaque variable spatiale sélectionnée, deux grandeurs ont été évaluées : l'indice de premier ordre et les indices totaux. Pour ce faire, suivant un nombre de scénario prédéfinis, toutes les variables d'entrée sont perturbées simultanément à travers une variation arbitraire. Le principe est alors d'explorer l'espace des variables d'entrée le plus efficacement possible afin d'obtenir une meilleure estimation de l'effet de la variabilité des variables d'entrée sur la variable de sortie.

4.7.1. Sélection des variables influant la contamination microbienne des eaux

4.7.1.1. Les variables choisies pour l'analyse de sensibilité

Pour effectuer l'analyse de sensibilité d'un modèle, il est important de sélectionner les variables d'entrée sur lesquelles il faut quantifier les incertitudes. Dans cette étude, le choix des variables considérées incertaines a été fait en fonction de leur poids, calculé dans la première partie de la thèse. La recherche s'est focalisée essentiellement sur les variables spatiales, plus précisément celles associées aux caractéristiques pédologiques et d'occupation du sol, ainsi que la distance des zones agricoles aux réseaux hydrologiques. Pour limiter le nombre de paramètres dans l'analyse de sensibilité, nous avons sélectionné cinq variables représentatives. Par ordre de priorité décroissant, il s'agit de la distance, des superficies des terres agricoles et zone urbaines, des sols loameux argileux et des sols loameux sableux. Dans la suite du texte, ces variables seront respectivement nommées «**DIST**»; «**TA**»; «**ZU**»; «**SLA**»; «**SLS**». Partant du fait que les bactéries fécales sont capables de persister dans le sol jusqu'à 15 cm de profondeur pendant environ deux mois et ce, jusqu'à la disparition complète du fumier ou du lisier, seul, l'horizon A dont l'épaisseur varie entre 25 et 30 cm est considéré pour les deux types de sols.

4.7.1.2. Définition des variables : domaine de variation et loi de distribution

Après la sélection des variables les plus influentes, la deuxième phase de l'analyse a consisté en la définition de leur domaine de variation. Pour les variables pédologiques **SLA** et **SLS** le choix d'un intervalle précis reste difficile à définir en raison du manque d'informations à ce sujet. En effet, dans la littérature, les valeurs de ces variables sont relativement disparates et il est difficile de faire un choix définitif sur la plage de variation de la texture des sols. Pour cela, nous avons choisi un intervalle correspondant aux différentes proportions de leurs teneurs respectives en argile et en sable sur le BV du lac Massawippi. Concernant la variable **DIST**; le même principe a été adopté, c'est-à-dire que les limites inférieure et supérieure sont données respectivement par la valeur minimale et la valeur maximale de **DIST**; du pourcentage dans chaque couche spatiale de la zone étudiée. Enfin, pour les caractéristiques d'occupation du sol nous avons choisi une plage de variation correspondant aux proportions des types d'occupation du sol. Puisque le BV du lac Massawippi est considéré agricole, pour les variables **TA** et **ZU**; les bornes inférieure à 40 et supérieure à 88, ont été choisies pour prendre en compte les incertitudes sur la mesure de ces deux variables.

Suite à la définition du domaine de variation, nous avons procédé au choix des lois de distribution des cinq variables sélectionnées. Au niveau des caractéristiques pédologiques, la loi de distribution reste très controversée : loi exponentielle négative (Baecher *et al.*, 1977), loi log-normale (Rouleau et Gale, 1985), lois normale et Gamma (Huang et Angelier, 1989), loi de Weibull (Bardsley *et al.*, 1990), loi de puissance (Gillespie *et al.*, 1993; Bodin et Razack, 1999). L'attribution des lois de distribution des caractéristiques du sol semble alors difficile à définir en pratique et repose essentiellement sur des avis d'experts. Ce faisant, pour les variables **SLA** et **SLS**, la loi de distribution uniforme a été choisie. Ce choix repose sur le fait qu'une loi uniforme est adaptée pour des variables mal définies, donc pour des variables dont nous disposons uniquement des bornes inférieure et supérieure. Au niveau des variables d'occupation du sol, en raison du manque d'informations sur la loi de distribution de certains variables, une fonction de distribution normale a été utilisée sur toute la gamme de valeurs possibles associées aux deux variables : **TA** et **ZU**. La même loi a été attribuée à **DIST**. Pour l'analyse de sensibilité, nous nous sommes appuyés sur deux hypothèses pour la distribution *a priori* des valeurs des paramètres (uniforme et normale). Dans un premier temps, la distribution normale a été choisie pour toutes les variables sauf (SLA dont la loi de distribution est uniforme). Les distributions normales tronquées ont été considérées pour respecter la plage des valeurs admissibles pour chacun des paramètres. Pour vérifier l'influence de la distribution des variables sur la variabilité de la sortie (taux de coliformes), nous avons simulé les mêmes cas tests en utilisant les distributions log-normale; Gamma et uniforme, de telle sorte que *la Distance et Terre Agricole* appartiennent à l'intervalle spécifié pour la distribution normale avec une probabilité 0.99. À cet effet, le test de Kolmogorov-Smirnov a été réalisé afin de déterminer la distribution statistique qui est la plus probable pour la variable simulée : les taux de coliformes. Ceci nous a conduits à déterminer les intervalles de confiances avec les méthodes de percentiles. Cette approche ne présuppose aucune restriction sur la distribution et, permet donc de mieux tenir compte des phénomènes d'asymétries et de valeurs extrêmes pouvant exister. Les lois de distribution ainsi que les valeurs limites assignées à chaque variable sont présentées dans le tableau 8.

Tableau 8. Valeurs limites et lois de probabilité des variables spatiales sélectionnées

| Variabes | Limite inférieure | Limite supérieure | Loi de distribution |
|----------|-------------------|-------------------|---------------------|
| DIST | 0 | 20 | normale |
| TA | 40 | 88 | normale |
| ZU | 40 | 88 | normale |
| SLA | 1 | 45 | uniforme |

Pour vérifier l'influence de la distribution des variables sur la variabilité de la sortie (taux de coliformes), nous avons simulé les mêmes cas tests en utilisant les distributions log-normale; Gamma et uniforme, de telle sorte que *DIST* et *TA* appartiennent à l'intervalle spécifié pour la distribution normale avec une probabilité 0.99. À cet effet, le test de Kolmogorov-Smirnov a été réalisé afin de déterminer la distribution statistique qui est la plus probable pour la variable simulée : les taux de coliformes. Ceci nous a conduits à déterminer les intervalles de confiances avec les méthodes de percentiles. Cette approche ne présuppose aucune restriction sur la distribution et, permet donc de mieux tenir compte des phénomènes d'asymétries et de valeurs extrêmes pouvant exister.

Les résultats du tableau 9 montrent que les distributions du taux de coliformes sont majoritairement log-normalement distribuées. De ces résultats, on pourra déduire que les résultats de simulation ne varient pas avec la distribution de la variable utilisée.

Tableau 9. Valeurs des critères de Khi-deux (test de Kolmogorov-Smirnov) pour identifier la distribution la plus probable de la variable simulée : Taux de coliformes.

| Scénario/Distribution | Normale | Gamma | Log-normal | La plus probable |
|------------------------------|---------|-------|------------|------------------|
| Distance : Loi uniforme | 10,43 | 23,18 | 9,32 | Log-normal |
| Distance : loi Gamma | 14,69 | 29,43 | 15,23 | Normal |
| Distance loi Log-Normal | 12,01 | 19,46 | 10,37 | Log-normal |
| Terre Agricole: loi uniforme | 15,39 | 22,14 | 13,54 | Log-normal |
| Terre Agricole : loi Normal | 17,29 | 18,46 | 14,27 | Log-normal |

Niveau de signification $P < 0,001$

4.7.2. Propagation et analyse des incertitudes

Après la sélection et la caractérisation des domaines de définition des variables, l'étape suivante a consisté à quantifier la variabilité des sorties de *PdIP* induite par l'incertitude des cinq variables d'entrée. Il s'agit de calculer un intervalle de confiance pour la sortie du modèle permettant de délimiter sa variation dans un domaine de définition donné, tout en tenant compte des incertitudes des entrées incertaines. Cette partie est basée sur l'utilisation des nombres aléatoires pour la réalisation des simulations. Celle-ci est importante dans l'analyse de sensibilité puisqu'une mauvaise taille d'échantillon peut entraîner une mauvaise quantification des incertitudes du modèle et, de ce fait, induire des erreurs dans l'estimation des indices de sensibilité.

En ce qui concerne l'évaluation des incertitudes en fonction du nombre de simulations, les principaux travaux sont ceux de Crosetto *et al.* (2000) et de Gouldby et Kingston, (2007). Il ressort de ces études que la meilleure taille d'échantillon dépend non seulement d'un compromis entre la précision et un temps de calcul acceptable, mais aussi du nombre de variables étudiées et du degré de non-linéarité du modèle analysé. Par conséquent, plus le modèle est complexe, plus il faut un échantillonnage de grande taille pour apprécier cette complexité. Pour la propagation des incertitudes dans notre cas, la méthode LHS a été choisie (pour des raisons évoquées à la section 3.4) afin de réaliser plusieurs expériences numériques avec le code de calcul *PdIP*. Pour ce faire, nous avons initialement considéré deux cas. Tout d'abord, le nombre de points de simulations a été considéré fixe à 10 (double du nombre des variables) tandis que le nombre de variables incertaines variait de 2 à 5. Dans un second temps, nous avons procédé de manière évolutive en faisant varier la taille de l'échantillon de 100, 200, 300, 500 jusqu'à 1000, tandis que le nombre de variables était stable. Enfin, pour chaque simulation, nous avons calculé les moyennes, les écart-types pour définir le pourcentage d'incertitude sur la valeur estimée, les coefficients de variation pour accéder au résultat de la propagation d'incertitudes sur la sortie.

Enfin, pour décider du nombre adéquat de simulations sur lequel les indices de sensibilité seront déduits, nous avons étudié la convergence de la moyenne en comparant les taux moyens de coliformes pour l'ensemble des configurations. Un test de Kolmogorov-Smirnov a été utilisé pour identifier la distribution la plus probable. Cette deuxième étape a permis non seulement d'optimiser l'emplacement des nouvelles expérimentations, mais aussi d'augmenter la quantité d'informations apportée par ces simulations.

4.7.3. Analyse de sensibilité

La troisième et dernière partie de la méthodologie est l'analyse de sensibilité paramétrique qui a permis de déterminer la variable ou la combinaison de variables spatiales responsable de l'incertitude de la contamination microbienne. Pour cela, les indices de sensibilité du premier ordre et totaux ont été estimés au moyen de la méthode de Sobol à partir des équations 19 et 20. Au départ, nous avons généré des échantillons pour chaque variable choisie, dans notre cas cinq matrices de 1 000 simulations. Ensuite, un second échantillon a été construit par permutation pour permettre l'analyse de sensibilité. Le modèle a été ensuite évalué avec ces deux échantillons, à partir desquels les indices de Sobol ont été déduits.

Afin d'évaluer la sensibilité du *PdlP* aux différentes variables spatiales prises en compte, nous avons considéré différentes configurations pour le calcul des indices de sensibilité de Sobol du premier ordre et totaux. Préalablement aux analyses des incertitudes, une étude de convergence a été menée en comparant les résultats de différentes simulations. Cette étape a permis de vérifier parmi les tailles de simulations utilisées pour l'échantillonnage, laquelle est suffisante pour déterminer avec précision les calculs statistiques relatifs aux indices de Sobol. En d'autres termes, le critère de convergence a aidé à certifier que de nouvelles simulations n'apportent plus d'autres informations, puisqu'elles ne modifient plus les résultats déjà obtenus.

Une fois la convergence vérifiée, nous avons obtenu pour chaque variable spatiale les indices de premier ordre et les indices totaux. Ces derniers ont permis de déterminer quelle part d'incertitude sur la réponse est due à l'incertitude de chaque variable d'entrée ou de la combinaison des variables.

4.8. Résultats et leur interprétation

Dans ce qui suit, nous présentons les résultats en fonction des étapes méthodologiques. La première partie présente l'étude de la convergence pour trouver le nombre de simulations minimales qui a permis d'obtenir les indices de Sobol. Suivra l'analyse des incertitudes des variables spatiales sur la réponse de la contamination microbienne des eaux de surface. L'analyse a permis de déterminer la contribution de chaque variable intrant à l'incertitude de la prédiction de la contamination microbienne. Cette analyse a été menée selon deux axes : (1) la quantification des incertitudes sur les variables pédologiques, d'occupation du sol et de la distance et (2) la propagation de ces incertitudes dans le modèle probabiliste du *PdlP*, c'est-à-dire

le calcul de l'erreur induite dans la sortie par les incertitudes des entrées spatiales. Ainsi, comparativement à la quantité des données acquises, une comparaison détaillée des incertitudes a été réalisée pour chacune des variables. Dans l'ordre, nous avons d'abord considéré des cas tests dans lesquels, seules, les incertitudes des caractéristiques pédologiques sont étudiées, ensuite les incertitudes des deux variables d'occupation du sol, enfin celles de la distance. Dans la dernière étape, nous avons analysé les résultats de sensibilité de chacune de ces variables.

4.8.1. Étude de la convergence

La première étape de l'analyse a consisté à s'assurer de la convergence des résultats. Il s'agit de vérifier que le nombre d'itérations effectuées est suffisant pour garantir l'obtention d'un état stationnaire, en s'assurant que la variation relative d'une simulation à une autre est suffisamment faible. Celle-ci a été inspectée graphiquement en regardant l'évolution des moments statistiques durant la simulation pour vérifier que le modèle a atteint la zone stationnaire. De ce fait, nous avons tracé pour chacune des caractéristiques pédologiques **SLA** et **SLS** la convergence de la moyenne spatio-temporelle de la contamination microbienne. Les résultats obtenus sont illustrés sur la figure 8.

Au regard de la figure 8, on peut constater au niveau des sols argileux que la convergence s'est effectuée à partir de 300 simulations après quelques petites oscillations entre 1 et 250 simulations. Pour les sols sableux, après plusieurs fluctuations, la moyenne s'infléchit encore dès 400 simulations, mais se stabilise à partir de 550 simulations. Pour les deux cas, nous avons constaté que les simulations convergent assez rapidement. Mais pour garantir une bonne stabilisation du critère de convergence, $N = 1000$ a été retenu pour l'ensemble des estimations d'incertitudes de la contamination microbienne liées aux caractéristiques pédologiques. Pour la représentativité des échantillons relatifs aux autres variables d'entrées spatiales une vérification a toujours été effectuée sur un $N = 1000$, afin de s'assurer que celui-ci reste suffisant pour donner une estimation correcte des indices de Sobol calculés pour l'ensemble des variables. Cette valeur de N semble suffisante, pour minimiser les coûts en temps des simulations, compte tenu du nombre important de configurations qui y seront effectuées dans les prochaines sections pour la quantification des incertitudes et l'analyse de sensibilité.

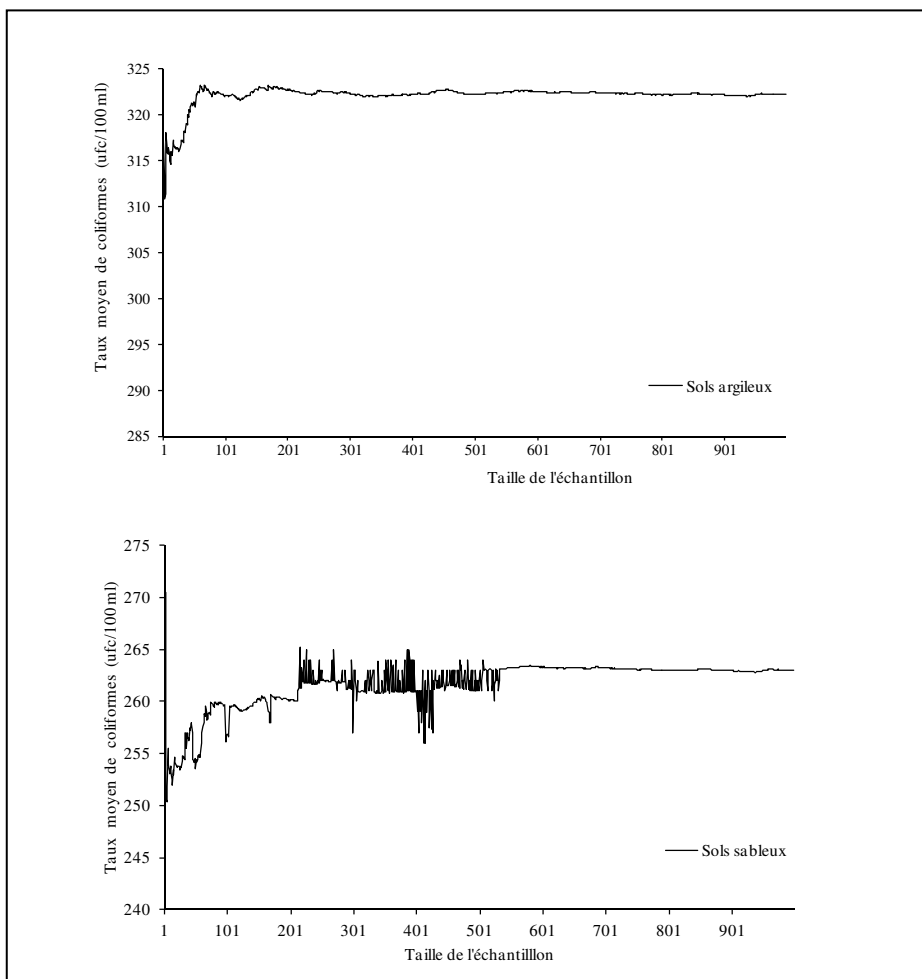


Figure 8. Convergence du taux moyen de coliformes en fonction de la taille de l'échantillon : cas des sols argileux et sols sableux.

4.8.2. Propagation et analyse des incertitudes

4.8.2.1. Variables pédologiques

Cette partie a trait à la propagation des incertitudes dues aux deux variables pédologiques **SLA** et **SLS** sur la variabilité du taux de coliformes. Les deux variables pédologiques ont été préalablement échantillonnées par le plan LHS avec une taille d'échantillon $N = 1\ 000$. Pour chaque variable pédologique six configurations spatiales sont étudiées, à savoir :

- deux configurations (2) à partir des variations de 10 et 15 % des valeurs de référence des teneurs des sols. Ces configurations sont notées : (config1) **10SLS** et (config2) **15SLS**; (config1) **10SLA** et (config2) **15SLA** respectivement pour désigner une variation de 10 et 15 % pour les sols sableux **SLS** et argileux **SLA**. Ces deux configurations dites de

référence étaient constituées par celles qui correspondent aux choix standards des variables pédologiques. Nous avons choisi alors dans cette première analyse de ne pas étudier les effets croisés et de nous limiter donc à ne faire varier qu'une seule des variables pédologiques à la fois par rapport à sa valeur initiale.

- Quatre configurations (4) prenant simultanément en compte les variations (10 et 15 %) des valeurs de référence des teneurs des sols et de la pente (10 et 15 %). Notons que lorsque la pente a été ajoutée dans la simulation, on a ajouté l'indice P. Les configurations pour les sols loameux sableux sont notées : (config1) **10SLS-10P**; (config2) **15SLS-10P**; (config3) **15SLS-10P**; et (config4) **15SLS-15P**. Les sols argileux prendront aussi la même notification. Par exemple (config1) **10SLA-10P** signifie qu'au niveau du sol argileux, la première configuration consiste à faire varier de 10 % la teneur en sable et de 10 % l'inclinaison de la pente.

Les figures 8 et 9 présentent respectivement les histogrammes de distributions du taux moyen de coliformes de **SLS** et **SLA** en fonction de chaque configuration. Les tableaux 9 et 10 représentent dans le même ordre les moments statistiques des deux variables spatiales. Il s'agit de la moyenne (μ), de l'écart-type (σ), des coefficients de variation (COV), des quartiles 2,5 et 97,5 ($Q_{2,5}$ et $Q_{97,5}$) correspondant aux différentes distributions.

L'analyse des histogrammes de distributions relatifs à la variation des **SLS** montre que toutes les distributions sont asymétriques, ni normales (gaussiennes), ni log-normales (figure 9). Quant aux **SLA**, les distributions des moyennes de coliformes pour l'ensemble des configurations se rapprochent toutes de la forme d'une cloche avec des modes bien marqués et légèrement asymétriques (figure 10). Pour les six tests de Kolmogorov-Smirnov relatifs à chacune des configurations, les *p-value* calculées sont inférieures à 0,0001 au niveau de signification $\alpha = 0,05$; l'hypothèse nulle de normalité est ainsi rejetée.

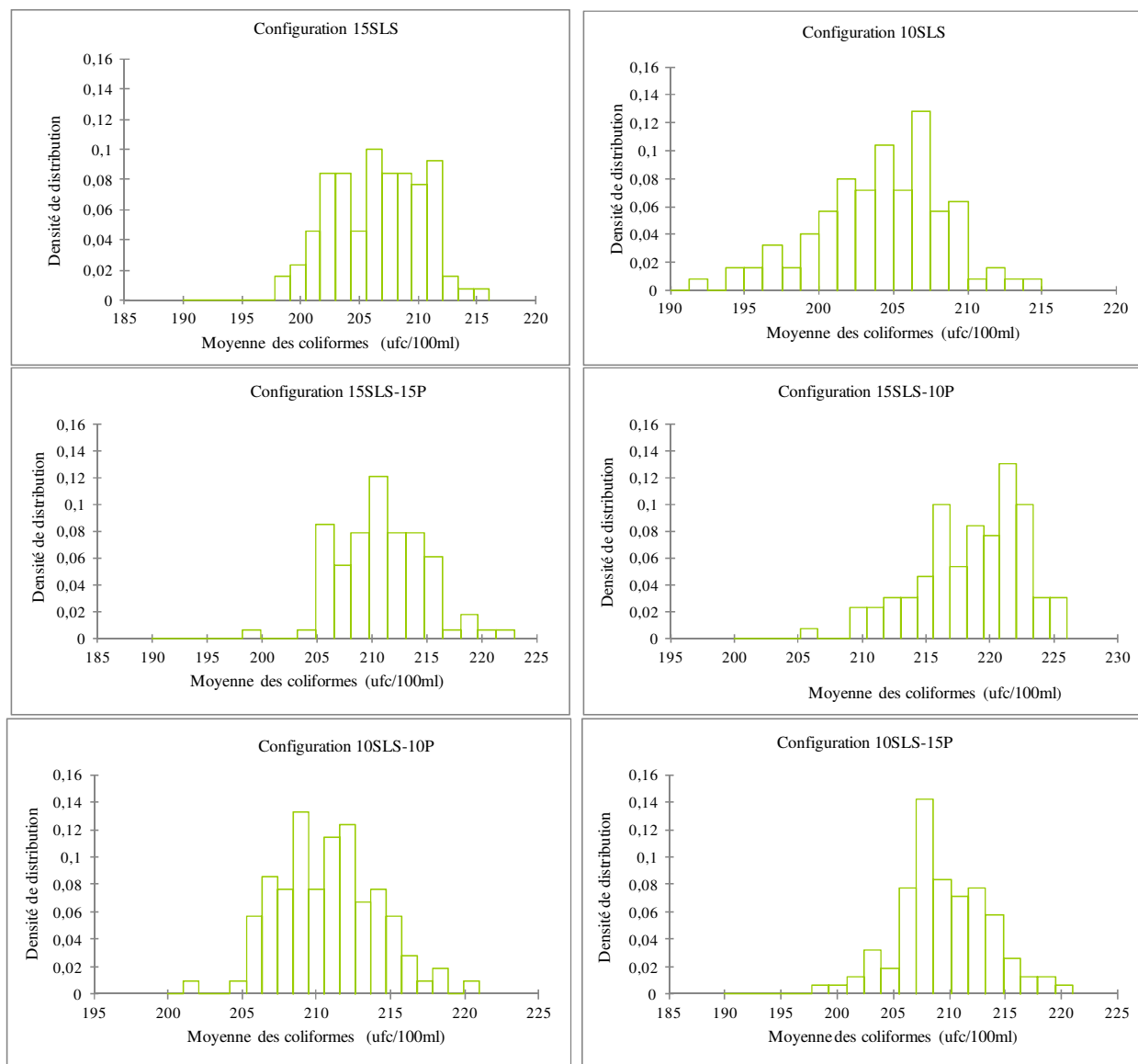


Figure 9. Distributions du taux moyen des coliformes dues à la variabilité des sols sableux et en fonction de la pente.

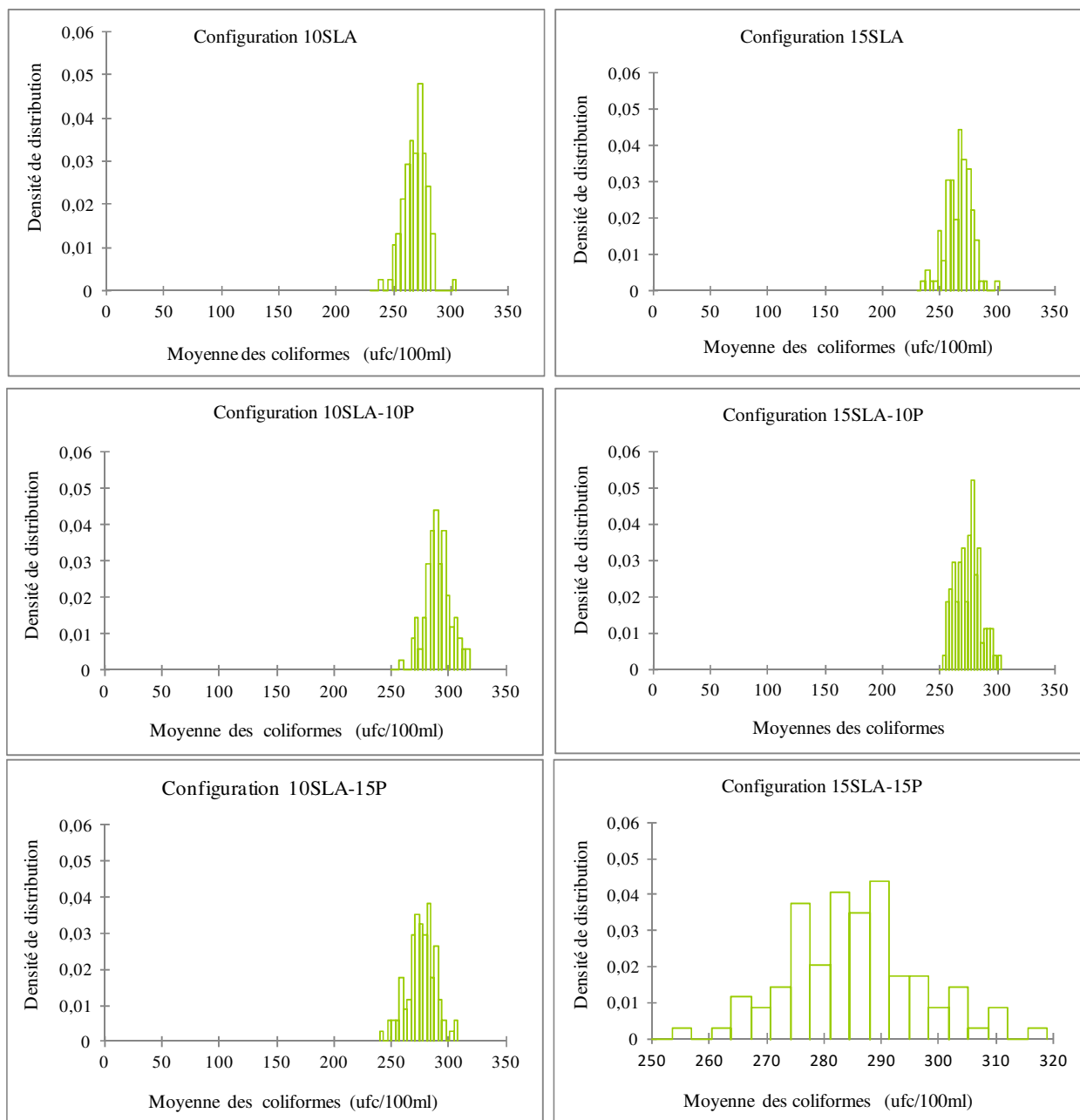


Figure 10. Distributions du taux moyen de coliformes dues à la variabilité des sols argileux et en fonction de la pente.

Au niveau des moments statistiques des différentes configurations, les résultats sont présentés dans les tableaux 10 et 11 respectivement pour les sols sableux et sols argileux.

Tableau 10. Moments statistiques du taux de coliformes en considérant les incertitudes du sol sableux.

| | Configuration pour les sols loameux sableux SLS | | | | | |
|-------------------|--|--------------|------------------|------------------|------------------|------------------|
| | 10SLS | 15SLS | 10SLS-10P | 15SLS-10P | 10SLS-15P | 15SLS-15P |
| μ | 204,3 | 206,1 | 210,6 | 218,4 | 208,3 | 210,5 |
| σ | 584,6 | 589,2 | 321,7 | 325,7 | 321 | 323,6 |
| Médiane | 141,1 | 144 | 147,5 | 149,3 | 135,4 | 136,4 |
| COV | 2,8 | 2,8 | 1,5 | 1,4 | 1,5 | 1,5 |
| Q _{2.5} | 14,4 | 13,5 | 10,1 | 9,8 | 9,5 | 10,1 |
| Q _{97.5} | 386,9 | 388,6 | 390 | 395 | 389,2 | 395,6 |

Tableau 11. Moments statistiques du taux de coliformes en considérant les incertitudes du sol argileux

| | Configuration pour les sols loameux argileux SLA | | | | | |
|-------------------|---|--------------|------------------|------------------|------------------|------------------|
| | 10 SLA | 15SLA | 10SLA-10P | 15SLA-10P | 10SLA-15P | 15SLA-15P |
| μ | 270,4 | 265,1 | 259,8 | 272,5 | 276,2 | 283,4 |
| σ | 831,8 | 852 | 865,6 | 827,8 | 881,2 | 891 |
| Médiane | 176,3 | 175,3 | 173,4 | 177,8 | 178,5 | 175,3 |
| COV | 3,1 | 3,2 | 3,3 | 3,1 | 3,2 | 3,1 |
| Q _{2.5} | 23 | 22,3 | 23,1 | 24,4 | 24,1 | 24,3 |
| Q _{97.5} | 709 | 710,5 | 700,9 | 702,7 | 703 | 700 |

En ce qui concerne les moments statistiques, il ressort qu'au niveau de **SLS** (tableau 10), les deux premières configurations **10SLS** et **15SLS** ont des valeurs moyennes de coliformes réduites au minimum, et les dispersions les plus importantes. Par exemple, pour la configuration **10SLS**, la moyenne est de 204 ufc/100 ml, alors que cette valeur atteint 218 ufc/100ml lorsqu'on augmente la pente de 10 %. Pour ces deux configurations les COV atteignent des valeurs de 280 % tandis que pour les configurations prenant en compte la pente, ces valeurs sont de 150 %.

Au niveau des **SLA** (tableau 11), il apparaît que la dispersion des moyennes des coliformes pour les six configurations est très élevée, atteignant les valeurs de COV supérieures à 300 %. Ces résultats montrent que les valeurs élevées de pente et de teneur en argile augmentent non seulement la moyenne du taux de coliformes, mais aussi sa variabilité. L'augmentation uniquement des teneurs en argile augmente les valeurs des écarts-types sans trop augmenter la variabilité des taux moyens de coliformes. À ce niveau, les résultats révèlent que

l'impact de la variabilité de la teneur en argile n'est significatif que lorsqu'on ajoute l'effet de la pente. Mais, cet effet est négligeable lorsque, seule, la variabilité de la teneur en argile est considérée. À titre d'exemple, lorsque l'on augmente la teneur en argile de 10 et 15 % de leur valeur de référence, le tableau 11 montre que la moyenne des coliformes diminue de 2 %. Alors que, pour une variation simultanée de la teneur en argile et de l'inclinaison de la pente (**15SLA-15P**), on observe une augmentation de 7 % par rapport à la valeur de référence. Notons enfin que, même si l'augmentation des pentes augmente la variabilité de la moyenne des coliformes, cette augmentation n'a pratiquement pas d'effet sur la dispersion de la moyenne des taux de coliformes.

Pour l'ensemble des configurations relatives aux deux variables pédologiques, les largeurs des intervalles de confiance sont relativement importantes (différence entre $Q_{97.5}$ et $Q_{2.5}$). Par exemple pour la configuration **10SLS** (tableau 9), l'intervalle de crédibilité à 95 % est large d'environ 372,4 ufc/100 ml, tandis que pour la même configuration au niveau des sols argileux (**10SLA**) cette valeur vaut presque 690 ufc/100 ml (tableau 10).

Une analyse comparative de l'ensemble des résultats (tableaux 10 et 11) montre que les incertitudes sont présentes aussi bien pour les sols argileux que les sols sableux, mais avec une amplitude variable d'une configuration à une autre. Pour les sols sableux (tableau 10), les incertitudes existent, mais elles sont moins importantes. Pour les configurations **10SLS-10P**, **15SLS-10P**, **10SLA-15P** et **15SLS-15P**, il n'y a aucune variation au niveau du coefficient de variation. Ceux-ci restent presque constants autour de la valeur 150 %. Ces résultats montrent que les incertitudes dues à la variation des sols sableux n'ont pas une influence importante sur les taux de coliformes, et que l'incertitude sur la réponse du modèle est moins ressentie. Quant aux sols argileux, les incertitudes sont importantes lorsque, seule, la variable **SLA** est considérée. En outre, bien qu'elles restent toujours élevées, elles augmentent légèrement lorsqu'on ajoute la pente dans la simulation (tableau 11). La même tendance est observée au niveau des écarts-types.

De façon globale, les résultats confirment à nouveau l'hypothèse d'une adhésion des coliformes aux particules des sols argileux. Mais qu'une pente optimale est la condition favorable pour faciliter le transport des coliformes des zones très éloignées des cours d'eau. Cette brève analyse montre que, dès lors qu'une source de variabilité des teneurs en argile, même peu importante pour le résultat final, est négligée, il y a une surestimation au niveau des quartiles, ce qui peut entraîner une grande erreur au niveau de la prévision de la contamination microbienne.

La modélisation se doit donc de prendre en compte toutes les sources de variabilité des sols argileux, même celles qui semblent avoir un impact mineur sur la distribution des coliformes.

4.8.2.2. Variables d'occupation du sol : terre agricole et zone urbaine

L'incertitude due à l'occupation du sol a été traitée de la même façon que celle des caractéristiques pédologiques. Les calculs ont été effectués en simulant successivement les distributions des taux de coliformes correspondant à chaque valeur de l'occupation du sol et en combinant les distributions obtenues avec la pente et les caractéristiques pédologiques. Pour chaque variable d'occupation du sol, trois configurations spatiales ont été étudiées, à savoir :

- deux configurations (2) à partir des variations de 10 et 15 % des valeurs de référence des types d'occupation du sol. Ces configurations sont notées : (config1) **10TA** et (config2) **15TA**; (config1) **10ZU** et (config2) **15ZU** respectivement pour désigner une variation de 10 et 15 % pour les terres agricoles (**TA**) et zone urbaine (**ZU**).
- Une configuration qui prend en compte les effets simultanés des deux types d'occupation du sol, les deux caractéristiques pédologiques. Cette configuration est nommée: PEDO + OS.

Pour chaque configuration, les histogrammes de distributions sont présentés sur les figures 10 et 11. Les moments statistiques des distributions sont reportés dans les tableaux 12 et 13. Seuls, les résultats les plus importants sont présentés. L'analyse de la figure 11 montre que la distribution des moyennes des coliformes en fonction de la variation de **TA** présente des formes plus ou moins en cloche avec des pics bien marqués. Ceux-ci se situent autour d'une valeur approximativement égale à 300 ufc/100 ml. De plus, ces pics marqués non centrés à zéro traduisent une distribution non symétrique, ce qui explique la différence entre la moyenne et la médiane dans le tableau 12. Cette asymétrie des distributions a été confirmée par une investigation supplémentaire faite en utilisant les tests de Kolmogorov avec un niveau de confiance de 95 %. Ces tests refusent l'hypothèse de normalité et de log-normalité de chacune des distributions au seuil de 5 %. Pour la variation de **ZU** (figure 12), les histogrammes de distributions obtenus, bien qu'ayant des allures comparables, paraissent plus ou moins différents. Ils sont semblables à ceux de **TA** et présentent aussi des pics relativement important autour de $x = 300$ ufc/100 ml.

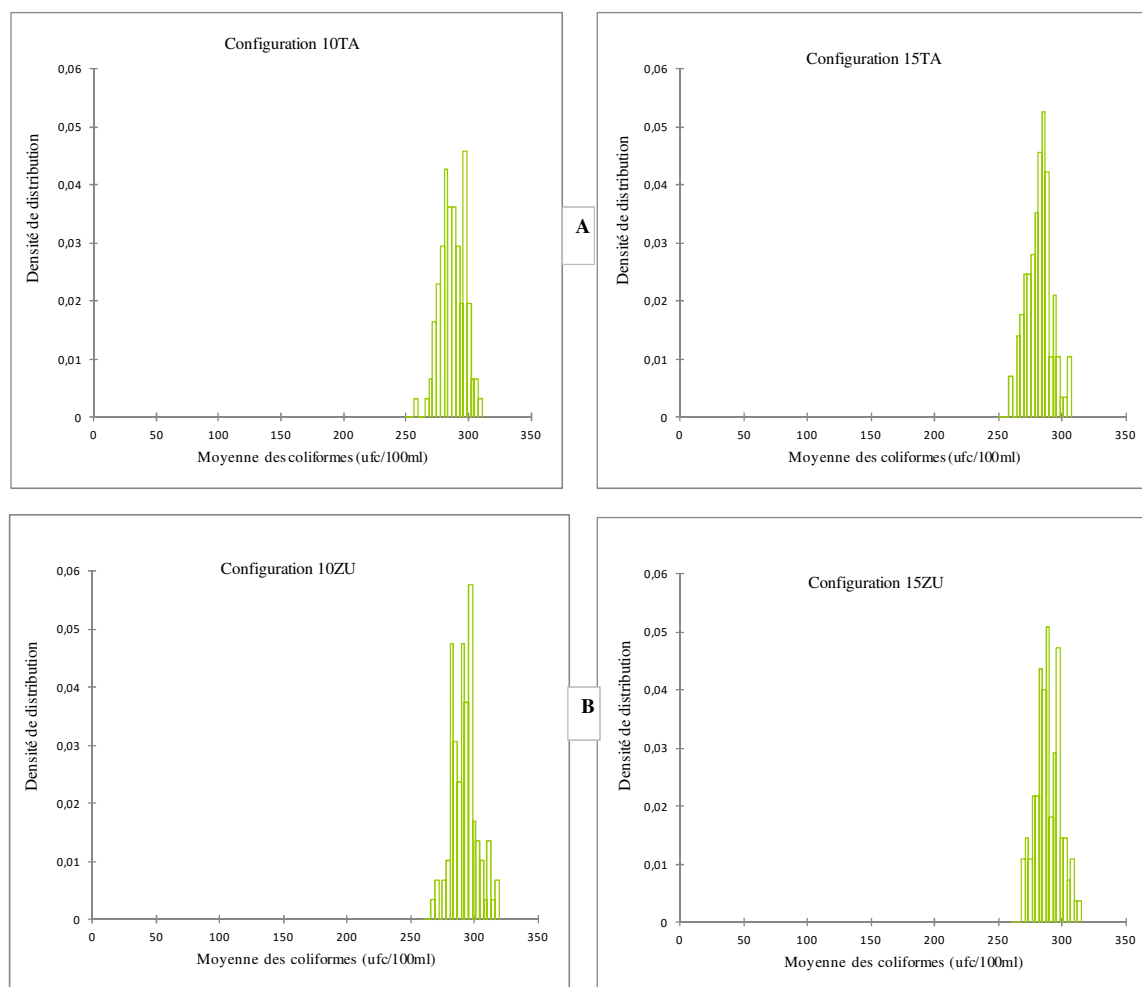


Figure 11. Distributions du taux moyen des coliformes dues à la variation de terre agricole - TA (A) et de zone urbaine - ZU (B)

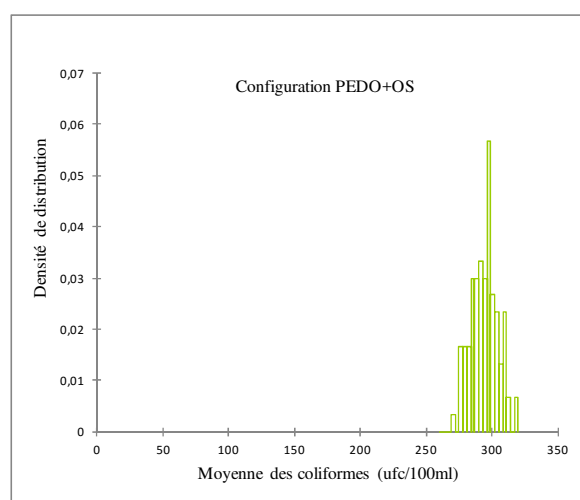


Figure 12. Distributions du taux moyen de coliformes en considérant simultanément les variables pédologiques et d'occupation du sol

Les résultats des moments statistiques de chacune des configurations sont présentés dans les tableaux 12 et 13.

Tableau 12. Moment statistiques de la distribution de coliformes en considérant la variation de deux variables d'occupation du sol et celle de la pente.

| | 10TA | 15TA | 10ZU | 15ZU | 10TA-10P | 10TA-15P | 10ZU-10P | 10ZU-15P |
|-------------------|-------|-------|-------|-------|----------|----------|----------|----------|
| μ | 306,8 | 315,6 | 300,7 | 304,3 | 304,1 | 315 | 305,7 | 306 |
| σ | 84,1 | 88,1 | 87 | 85,6 | 158,1 | 168 | 175,1 | 177,7 |
| Médiane | 126,7 | 146,6 | 149,6 | 144,5 | 142,1 | 132,3 | 145,2 | 140,8 |
| COV | 0,3 | 0,3 | 0,3 | 0,3 | 0,5 | 0,5 | 0,5 | 0,5 |
| Q _{2,5} | 205,7 | 207,6 | 302,8 | 315,4 | 400,3 | 345,4 | 356,2 | 359,7 |
| Q _{97,5} | 504,9 | 490,6 | 480,4 | 524,1 | 512,4 | 523,9 | 510,2 | 524,1 |

Tableau 13. Moment statistiques de la distribution de coliformes en considérant les incertitudes des deux variables pédologiques et les deux variables d'occupation du sol

| | PEDO + OS |
|-------------------|-----------|
| μ | 305,9 |
| σ | 48,1 |
| Médiane | 141,9 |
| COV | 0,2 |
| Q _{2,5} | 352,7 |
| Q _{97,5} | 404,8 |

L'examen détaillé des moments statistiques relatives aux deux variables d'occupation du sol (tableau 12) montre que même si les valeurs moyennes de coliformes varient légèrement avec l'augmentation des surfaces des deux types l'occupation du sol (**TA** et **ZU**), les COV ne suivent pas cette tendance, ils restent constants et égaux à environ 30 %. Par ailleurs, on peut observer à partir du même tableau que l'augmentation des quartiles est plus importante dans le cas **ZU** par rapport au cas **TA**. Par exemple, pour une variation de 15 % de la valeur initiale de **ZU**, Q_{97,5} passe de 480 à 524 ufc/100 ml, soit une augmentation de 8 %. Une tendance inverse s'observe au niveau de **TA**.

Au niveau de la variation de la pente, les résultats (tableau 13) montrent que les moyennes de distributions de coliformes connaissent des augmentations au fur et à mesure que

l'inclination de la pente est importante. À titre d'exemple, pour **TA** avec une augmentation de 15 % de la pente, les taux moyens de coliformes sont plus élevés, atteignant les valeurs de 315 ufc/100 ml. La même tendance s'observe dans le cas de **ZU** dont les valeurs moyennes atteignent 306 ufc/100 ml pour une augmentation de 15 % de la pente. En analysant les résultats des COV, on observe que la pente a une influence importante pour les deux variables d'occupation du sol. Par exemple pour la configuration **10TA**, la dispersion est de 30 %. Pour la même configuration avec une combinaison de la pente (15 %), la dispersion atteint 50 %.

En ce qui concerne la configuration combinant les variables pédologiques et d'occupation du sol, (tableau 13), il ressort que l'ajout des variables pédologiques donne lieu à une diminution de l'écart-type et de la COV. Nous constatons aussi qu'une différence apparaît au niveau des largeurs des intervalles de confiance dont les valeurs sont d'environ 50 ufc/100 ml, tandis ces valeurs restent élevées (> 100) pour les autres configurations.

La comparaison entre les configurations relatives à l'occupation du sol, montre que, même si on observe une légère différence entre les valeurs moyennes, la dispersion est plus grande au niveau des configurations où les variables d'occupation du sol sont combinées à la variation de la pente. Cette dispersion dont les valeurs atteignent 50 % semble stable pour chacune des variables, mais elle diminue considérablement (20 %) lorsqu'on introduit les variables pédologiques (tableau 13).

L'ensemble de ces résultats montre que la variabilité du taux de coliformes obtenue en considérant les incertitudes du sol et de l'occupation du sol est plus petite que celle obtenue par les deux variabilités calculées séparément (tableaux 12 et 13). De ces résultats, on peut conclure qu'il est nécessaire de prendre en compte toutes les incertitudes des caractéristiques du sol et des types de sol dans un seul calcul pour aboutir à des résultats précis. Ces constats mettent l'accent sur la complexité de la contamination microbienne. En principe de nombreuses variables sont en interaction et les inter-corrélations dépendent des valeurs de chaque variable qui peut avoir un effet stimulant. C'est le cas de la pente qui peut être assimilée à une variable aléatoire qui induit une différence de comportement entre les variabilités de la contamination microbienne.

4.7.2.3. Proximité des cours d'eau aux zones agricoles

Tout comme les caractéristiques pédologiques et d'occupation du sol, nous avons étudié la variabilité de la contamination microbienne due à la variation des distances séparant les cours

d'eau des zones agricoles. Pour étudier l'effet de la variabilité des distances sur la variabilité du taux de coliformes, nous avons dans un premier temps, recommencé la simulation telle que décrite précédemment au niveau des caractéristiques pédologiques et d'occupation du sol. Suivant le même principe, trois différentes configurations ont été testées.

- La première configuration a consisté à faire varier de 10 et 15 % la longueur initiale de la distance et en conservant les valeurs initiales de toutes les autres variables. De façon similaire aux notifications des variables précédemment étudiées, deux cas relatifs à cette configuration ont été étudiés. Il s'agit de **10DIST**; **15DIST** qui désignent respectivement une variation de 10 et 15 % de la distance initiale.
- La deuxième configuration a consisté à prendre simultanément en compte les variations des valeurs de référence des distances (10 et 15 %) et celles de la pente (10 et 15 %). Notons que lorsque, dans la simulation, on ajoute la pente, l'index P est utilisé. Par exemple **10DIST-10P** pour signifier qu'au niveau de la distance, la simulation consiste à faire varier de 10 % sa valeur initiale et de 10 % l'inclinaison de la pente.
- La troisième configuration prend simultanément en compte l'ensemble des variables spatiales, c'est à dire les deux variables pédologiques, les deux variables d'occupation du sol, la distance et la pente. Elle est notée **PEDO+OS+DIST**.

Les histogrammes des configurations sont présentés sur la figure 12; les moments statistiques sont présentés dans le tableau 14.

Une analyse de la figure 12 montre que les histogrammes liés à la variation des distances et de la pente génèrent des distributions de formes multimodales. Par exemple, pour le cas **10DIST**, on peut constater que la distribution de coliformes présente un mode principal autour de $x = 165$ ufc/100 ml et un mode secondaire autour de $x = 163$ ufc/100 ml. Pour ces configurations, on rejette encore l'hypothèse de normalité pour la distribution de coliformes. Les tests de Kolmogorov fournissent des valeurs inférieures à *p-value*. On conclut à nouveau le rejet de l'hypothèse de la normalité des distributions.

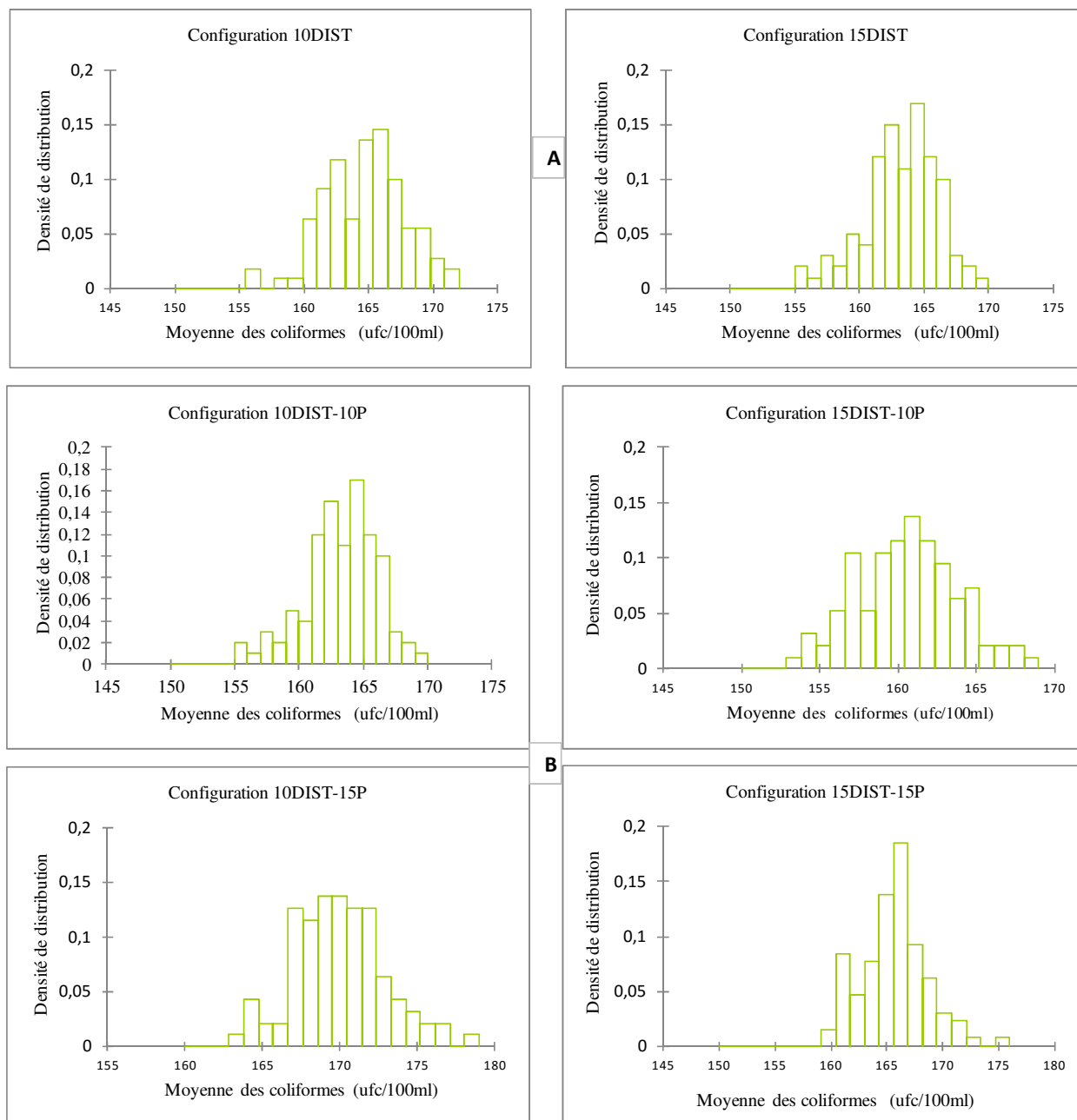


Figure 13. Distributions du taux moyen de coliformes dues à la variation de **DIST** (A) et en fonction de la pente (B)

Tableau 14. Moment statistiques de la distribution des coliformes en considérant : la variation de la distance et de la pente

| | Configuration 1 | | Configuration 2 | | | | Configuration 3 |
|-------------------|-----------------|--------------|-------------------|-------------------|-------------------|-------------------|--------------------|
| | 10DIST | 15DIS | 10DIST-10P | 15DIST-10P | 10DIST-15P | 15DIST-15P | PEDO+OS+DIS |
| μ | 129,6 | 139,3 | 122,7 | 114,1 | 104,8 | 115,1 | 314,4 |
| σ | 381,3 | 353,5 | 182,1 | 183 | 181,5 | 280,2 | 229,3 |
| Médiane | 191,3 | 184,1 | 114,5 | 110,4 | 100,9 | 101,7 | 274,6 |
| COV | 2,9 | 2,5 | 1,4 | 1,6 | 1,7 | 1,4 | 0,7 |
| Q _{2,5} | 0,4 | 0,9 | 0,4 | 0,5 | 0,9 | 0,9 | 28 |
| Q _{97,5} | 532,7 | 543,1 | 512,2 | 506,3 | 510,3 | 531,1 | 519,1 |

Si l'on s'intéresse aux moments statistiques des trois types de configurations relatives à la distance (tableau 14), on peut constater que les valeurs médianes semblent dépendantes de la variation de la distance et de la pente. Ceci est vrai entre les cas **10DIST** et **15DIST** et entre **10DIST-10P** et **15DIST-10P** à travers lesquels on peut observer une légère baisse des médianes en fonction de l'augmentation des proportions de la distance. En plus, pour les configurations **10DIST** et **15DIST** les médianes sont supérieures aux moyennes avec des écarts relatifs de plus de 30 % et des fortes dispersions : respectivement 294 % et 250 %.

Au niveau de la configuration 3 (PEDO+OS+DIST), l'analyse des résultats montre une diminution considérable des COV. Comparativement aux configurations précédentes, c'est sur cette dernière que les valeurs de COV les plus petites ont été enregistrées. Par exemple, pour la configuration **10DIST**, la valeur de COV passe de 290 % à 70 % lorsque, seule, la distance est considérée et que la distance est combinée aux variables pédologiques et d'occupation du sol (configuration 3 tableau 14). Ces résultats laissent présager que la détermination précise des incertitudes liées à la distance est très importante pour obtenir de résultats fiables du taux de coliformes. Cependant, cette variabilité due à la distance est essentiellement influencée par la variabilité totale, autrement dit à la combinaison des autres composantes : occupation du sol, la pente et les caractéristiques pédologiques.

Les résultats obtenus des deux premières configurations relatives à la variable **DIST (10DIST, 15DIST)** sont bien en accord avec les processus de rétention et de dégradation des polluants. L'une des raisons potentielles qui expliquent ces résultats est

que, contrairement aux caractéristiques pédologiques et d'occupation du sol, la distance 0-10 km est la distance maximale nécessaire pour apporter le plus d'informations sur la contamination microbienne. Au fur et à mesure que l'on s'éloigne de cette limite, les valeurs générées par la simulation deviennent aléatoires, permettant à n'importe quelle abscisse d'être atteinte. Les investigations faites sur les configurations de la distance permettent de croire que les valeurs des distances introduites dans la simulation en dehors du rayon d'influence 0-10km peuvent produire des sorties probablement erronées. L'extrapolation de la distance est donc à proscrire. Cela implique également qu'il faut veiller à ce que les tirages aléatoires des données d'entrée se fassent bien dans cet intervalle. À ce niveau de l'analyse, il semble que les conditions de simulations n'ont pas permis l'expression de la distance, qui semble ne pas jouer un rôle majeur dans la variabilité des taux de coliformes. Bien que les effets de la distance soient significatifs, il n'est pas pour l'instant évident d'en proposer une réelle interprétation. Nous espérons qu'avec la prise en compte des interactions, la combinaison de cette variable avec d'autres variables explicatives permettront d'affiner les résultats.

4.8.3. Analyse de sensibilité

Comme nous l'avons expliqué précédemment en présentant les principes de la méthode de Sobol, il est possible d'utiliser la variance pour estimer la contribution de chaque variable aléatoire à la variabilité du taux de coliformes. Dans cette partie, on définit pour chacune des variables spatiales sélectionnées un indice de sensibilité de premier ordre qui prend en compte la seule contribution de la variable aléatoire et un indice de sensibilité totale qui tient compte de la contribution de la variable aléatoire et de toutes ses contributions croisées avec les autres variables aléatoires. La figure 14 montre l'effet de la variabilité des variables sur l'indice de Sobol.

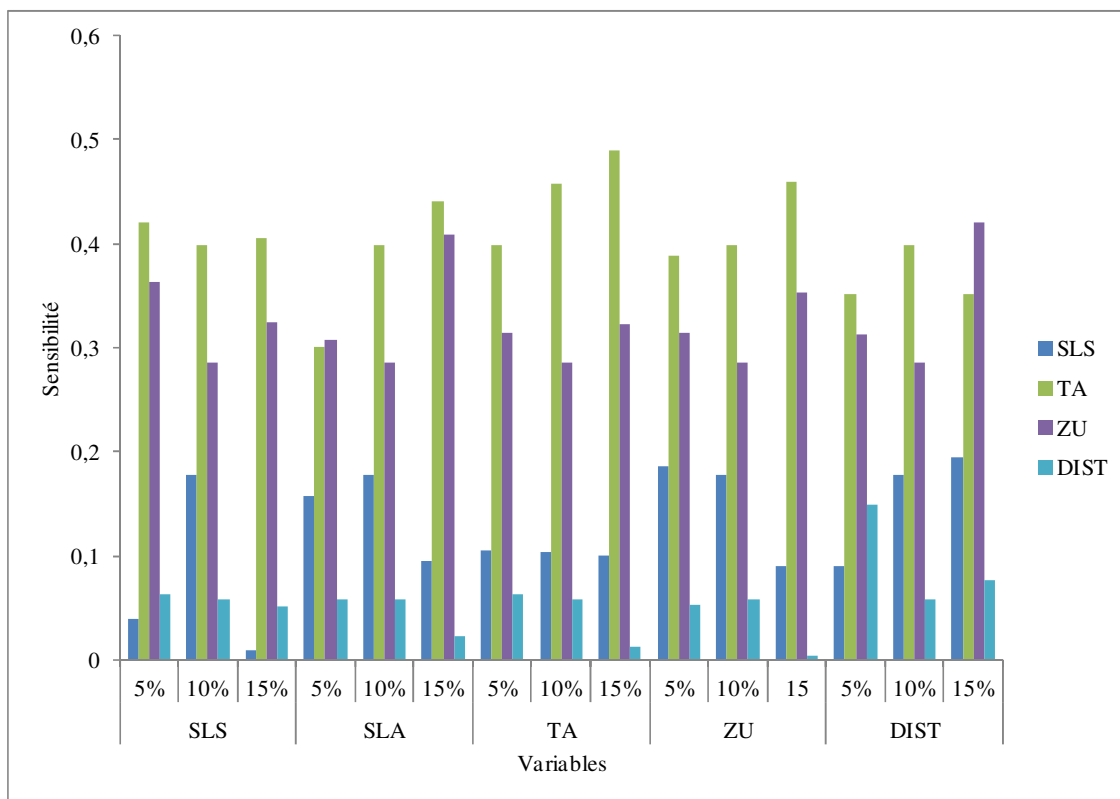


Figure 14. Effet de la variation des variables sur les indices de Sobol

Une analyse de la figure 13 montre que le taux d'*E. coli* est en premier lieu influencé par la variation par la surface agricole **TA** avec une valeur de sensibilité moyenne égale à 0,37 suivie de la zone urbaine **ZU** dont les valeurs moyennes atteignent 0,29. Cependant, les teneurs en sable et la distance ont les indices de sensibilité les plus faibles avec des valeurs inférieures à 0,15. De manière générale, les indices de sensibilité de premier ordre sont tous inférieurs à 0,50. De plus, nous remarquons que, l'augmentation et/ou la diminution des valeurs de référence d'une variable aléatoire conduit à l'augmentation et/ou la diminution de son indice de Sobol, conduisant par conséquent à la diminution/augmentation de l'indice de Sobol des autres variables aléatoires. Cela est particulièrement vrai pour les variables aléatoires ayant le plus grand poids dans la variabilité de la contamination microbienne. À titre d'exemple, pour **TA**, lorsque l'on augmente de 5, 10 et 15 % sa valeur de référence, l'indice de sensibilité augmente respectivement de 0,39, 0,45 et 0,49 tandis que les indices de sensibilité des sols argileux diminuent de 0,16, 0,10 et 0,10. Pour les deux variables restantes (distance

et sol sableux) leurs contributions demeurent toujours moins importantes en raison des petites valeurs de leurs indices de Sobol. Bien que faibles, les indices de sensibilité calculés pour une variation de 5 % de la distance ont des valeurs assez élevées soit une moyenne de 0,07 contre 0,03 pour une variation de 15 % de la valeur initiale de la distance. Ces valeurs décroissent vite lorsque la distance augmente. Ces résultats semblent indiquer que la distance n'a pas d'importance significative sur la variance des concentrations en coliformes. Mais elle intervient beaucoup dans le calcul de sa valeur. Cette variable peut être fixée à la valeur moyenne de son intervalle d'incertitude sans modifier significativement la concentration en coliformes.

Comme le démontrent les indices de premier ordre, l'influence individuelle des variables sur le taux moyen des coliformes est relativement faible. Cependant, l'influence des interactions paramétriques à travers les indices de sensibilité totaux est relativement importante. Nous pouvons le constater sur la figure 14 qui présente les résultats des indices de sensibilité totaux.

Dans l'ensemble (figure 14), nous constatons une augmentation des valeurs des indices totaux comparativement aux indices de sensibilité de premier ordre. Les différents résultats montrent que les indices de sensibilité totaux aux valeurs élevées proviennent des variables déjà importantes individuellement et sont significativement dominées par des effets d'interaction. Mais, relativement à la hiérarchie obtenue au niveau des indices de premier ordre, l'ordre d'importance des variables est permuté. En effet, la variable **DIST** située respectivement en dernière position pour l'analyse de sensibilité de premier ordre se retrouve en première position lorsqu'elle est en interaction avec d'autres variables. De même, la variable **SLA** située en troisième position se place en deuxième position ce qui augmente de façon considérable son influence sur les taux moyens de coliformes. L'occupation du sol vient en troisième position.

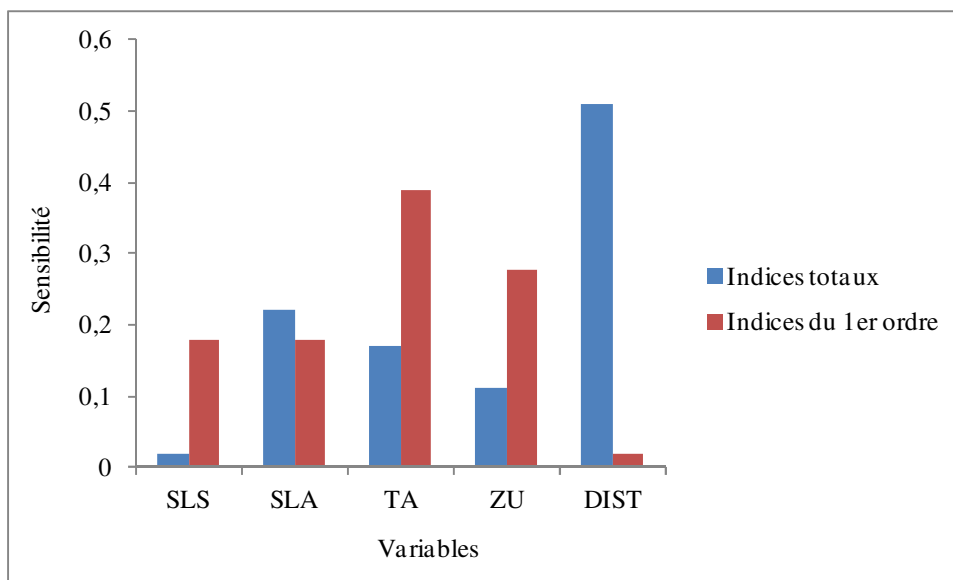


Figure 15. Indices de sensibilité du premier ordre et totaux

Au regard de l'ensemble des résultats de la sensibilité (figure 14), il reste à comprendre pourquoi certains effets des interactions induisent un comportement différent. Au niveau de la distance, il n'est pas aisé d'expliquer cette différence de comportement. Mais nos hypothèses laissent croire que cette augmentation de l'indice de sensibilité pourrait être due à la forte extension des zones agricoles dans une partie du BV du lac Massawippi, diminuant les distances entre les zones agricoles et le réseau hydrologique. D'autre part, il est fort probable que la morphologie du BV, les propriétés physiques du sol associées aux faibles distances dans cette partie du bassin peuvent être favorables au transport des coliformes. Dans ce contexte, une pente assez élevée combinée aux faibles distances va accélérer les mouvements des coliformes en agissant sur la direction, l'importance et la vitesse de l'écoulement de surface et du drainage de sub-surface (Moore *et al.*, 1991, 1998; Etzelmüller *et al.*, 2003; Gabet *et al.*, 2003). Ces deux conditions réunies peuvent avoir pour conséquence l'augmentation de l'effet des interactions entre la distance et les autres variables spatiales. En effet, le lien entre la distance et la concentration des coliformes dans les eaux a rarement été décrit. Mais, les quelques modèles de corrélation supposent dès le départ que dans le cas des bassins plats, la distance n'est pas pertinente pour représenter la concentration des coliformes.

Cependant, au niveau des terrains pentus, une corrélation significative est observée entre 0 et 10 km.

Au niveau des sols argileux, leurs effets d'interaction avec d'autres variables restent parmi les plus élevés (figure 14). De façon générale, ces résultats s'avèrent conformes aux principes mentionnant que la variabilité de la contamination microbienne est d'une part très sensible à cette variable et d'autre part, elle est sujette à des rétroactions positives avec d'autres variables. En principe, l'effet d'un sol argileux sur la contamination microbienne est en partie influencé par la distance, sans compter les effets de nombreuses autres variables. D'ailleurs, des auteurs (Crane *et al.*, 1983; Gagliardi et Karns, 2002) soulignent que l'influence du type de sol sur la contamination microbienne ne peut s'expliquer, si l'on ne prend pas en compte les conditions pédologiques. En particulier, le déplacement des coliformes est fortement dépendant de leur adhésion aux particules de sol, surtout pour les premiers centimètres du sol. La texture fine du sol empêche alors l'infiltration de l'eau et permet le déclenchement du ruissellement transportant ainsi les coliformes vers des cours d'eau. Les travaux de Jiang *et al.* (2002) et Islam *et al.* (2005) présentent de nombreux résultats et des modèles pour expliquer l'influence des sols sur le transport des coliformes. Dans leur revue bibliographique, ces auteurs identifient comme facteurs pédologiques explicatifs de la concentration des coliformes les facteurs suivants : la texture et la structure du sol, le taux d'humidité, Le pH et l'apport d'éléments nutritifs. Cependant, il s'avère qu'en l'état actuel de la connaissance scientifique, les différents facteurs du milieu naturel sont confondus dans les études scientifiques et il est donc très difficile de tenir des conclusions générales de ces travaux.

En ce qui concerne l'occupation du sol, les résultats sont assez proches de ceux observés pour les sols argileux, à la seule différence que l'influence de ces variables ne s'explique pas significativement par ses effets d'interaction, ces derniers étant moins importants que son effet principal. À cet égard, contrairement aux idées généralement répandues, il n'existe pas une culture à risque, ni un système de production à risque (Sebillotte *et al.*, 1990); tout dépend du contexte géographique, pédoclimatique et hydrologique. Pour qu'il y ait pollution diffuse des eaux de surface, deux conditions doivent être réunies : la présence d'un élément sous une forme transférable, à l'intérieur

ou à la surface d'un sol, et l'apparition d'un élément déclenchant (le plus souvent un épisode pluvieux ou une crue).

4.9. Conclusion partielle

L'analyse de sensibilité des variables spatiales sur *PdlP* a permis de vérifier dans quelle mesure chacune d'entre elles influence la concentration en coliformes des eaux de surface. L'intérêt d'un tel exercice était de spécifier une hiérarchisation des variables d'entrée spatialisées sur lesquelles des incertitudes avaient préalablement été définies. L'analyse détaillée de l'incertitude a permis de mettre en évidence l'existence de relations entre la variabilité de la contamination microbienne et l'incertitude qui entache les différentes variables spatiales et leur contribution. De l'analyse des résultats, nous pouvons tirer que l'incertitude de la contamination microbienne dépend directement de la variabilité des sols argileux, de la distance et des terres agricoles. Au niveau des SLA, la dispersion moyenne des taux des coliformes est la plus élevée (300 %) et sa distribution est très étalée. Les résultats obtenus ont montré par ailleurs une nette amélioration du coefficient de variation dès lors que la distance a été introduite dans l'ensemble des configurations. Celle-ci n'a pas semblé présenter de tendance aux niveaux des moyennes de coliformes. Par contre, elle s'est révélée suffisamment marquée pour permettre une diminution de l'écart-type et du COV, avec néanmoins des niveaux de confiance faibles.

Les indices de premier ordre de l'analyse de Sobol ont montré que parmi les facteurs les plus susceptibles d'influer la contamination microbienne, **TA** est le premier facteur important dans l'évaluation du taux de coliformes. C'est donc sur cette variable que l'attention devra se porter pour améliorer sa description dans le contexte de prévision d'une contamination microbienne. Ensuite, la deuxième variable la plus importante est **ZU** avec des parts de sensibilité d'environ 30 %. Par ailleurs, les estimations des indices totaux sont meilleures que celles des indices de premier ordre, ce qui signifie que l'impact des interactions paramétriques est significatif pour la modélisation de la contamination microbienne. Ce comportement est plus remarquable au niveau des indices de sensibilité du premier ordre de la variable **DIST** dont les valeurs ne sont que de 0,12. Cependant, au niveau des indices de sensibilité totaux, la variabilité de cette variable est

plus importante avec un comportement inverse, puisque son indice de sensibilité total devient le plus important.

Au vu de ces résultats, plusieurs points concernant la sensibilité du modèle *PdlP* aux variables spatiales peuvent être soulevés après analyse des indices de sensibilité totaux :

- La répartition du taux de coliformes due à la variabilité des variables est structurée en fonction d'un gradient sur la variable **DIST** d'une part et le type de sol d'autre part. Ce double gradient laisse présager une relation univoque entre les concentrations des coliformes, la distance et le type de sol.
- À superficie agricole et inclinaison de pente égales, plus la variable « sol argileux » est importante plus le taux de coliformes est élevé. Ce comportement est logique puisque cette variable a une influence directe sur des processus de transport telles l'advection-dispersion et l'adsorption-désorption, dont l'ensemble est fonction de la trajectoire de l'écoulement de l'eau et la nature du sol (Mayer *et al.*, 1997; Hendry *et al.*, 1997 et 1999). Ainsi, plus le sol est imperméable plus le ruissellement et le lessivage seront importants, avec des temps de transit courts.
- Une diminution de la distance se traduit par une diminution du temps de transit des coliformes entre les milieux agricoles et les eaux de surface. Là encore, ce comportement est cohérent puisque la distance permet de réguler la durée de transport des taux de coliformes. Plus cette variable est de faible valeur, plus le transport des coliformes depuis les sous-bassins est rapide, ce qui peut augmenter leur présence dans les cours d'eau.

Bien que la méthode de Sobol offre plusieurs avantages et semble prometteuse pour la hiérarchisation des variables spatiales, il convient d'être prudent avant toute généralisation des résultats. Ceux qui sont proposés sont pour l'instant propres au BV du lac Massawippi à la fois en termes d'occupation du sol et de type de sol.

Chapitre 5. Variabilité temporelle de la contamination microbienne des eaux : une approche géomatique à l'échelle d'un Bassin versant agricole

La dynamique temporelle des épisodes de contamination microbienne des eaux de surface implique une grande variabilité et une multitude de processus hydrodynamiques; biotiques et physiologiques. Ces derniers induisent une complexité structurelle difficile à expliquer à l'aide d'indicateurs ou de simples statistiques multi-variées. Les chapitres 2 et 3 de cette thèse ont mis en évidence l'importance des facteurs spatiaux sur de la contamination microbienne des eaux de surface. De ces études, il a été possible de quantifier l'influence de chacune de ces variables environnementales, d'analyser leur sensibilité et de cartographier la probabilité des risques de contamination. Toutefois, ces deux chapitres, ne permettent pas d'évaluer la variabilité temporelle des épisodes de contamination, puisque les caractéristiques de la pluie ne sont pas prises en compte. Or, lorsque l'on s'intéresse à l'apparition des épisodes de contamination microbienne, la prise en compte de variables spatiales et temporelles est nécessaire.

Dans la présente partie de cette thèse, il s'agit de mettre en œuvre une modélisation de la variabilité temporelle de la contamination microbiologique du BV du lac Massawippi, à partir du modèle ArcView Soil Water Assessment Tool (AVSWAT; DiLuzio *et al.*, 2005), de tester sa pertinence et sa complémentarité par rapport à des études antérieures réalisées sur ledit BV.

5.1. Introduction

5.1.1. Contexte

Les précipitations constituent un aspect important influant sur le transfert des flux de bactéries dans les cours d'eau. Dans ce contexte, des études (Crowther *et al.*, 2001; Hébert, 2001; Passerat *et al.*, 2011) ont démontré qu'il existait une liaison entre la pollution bactérienne des eaux et les variables caractérisant la pluie, notamment l'intensité et la durée. L'analyse de ces caractéristiques est donc un élément constitutif de l'étude de la prévision temporelle des flux de bactéries dans les cours d'eau. Elle permet de positionner pour chaque nouvel épisode pluvieux sur une échelle d'intensité (faible, moyenne, forte, exceptionnelle) la quantité de bactéries susceptible de porter préjudice au

cours d'eau. D'importants travaux de recherches ont été réalisés afin de quantifier l'influence de la pluie sur la contamination chimique ou microbiologique des eaux de surface (Kay *et al.*, 2005a; Guber *et al.*, 2007; Grimaldi *et al.*, 2006). Ces travaux ont mis en évidence le rôle important de la pluie dans le transport par ruissellement des contaminants vers les milieux récepteurs. En effet, en tant qu'élément déclencheur, il est bien évident que la pluie par la force de son énergie occasionne le décollement des particules, tandis que le ruissellement génère une mise en suspension qui favorise le déplacement des polluants. Bien que l'importance et le rôle de la pluie dans le transport des bactéries soient mis en exergue, les interactions entre les précipitations, les hétérogénéités spatio-temporelles des usages du sol et la dégradation des eaux demeurent complexes. Ces interactions, souvent mal connues et généralement variables dans le temps et l'espace sont difficiles à expliquer à l'aide d'indicateurs ou de simples statistiques multivariées. La méconnaissance des interactions impliquées dans le processus de la contamination des eaux, la sous-représentativité des phénomènes expliquant la présence des particules dans l'eau et de manière plus récurrente, les erreurs liées à la spatialisation sont source de dérives et d'inexactitudes dans les estimations statistiques des flux de bactéries. Ainsi, dans la plupart des modèles utilisés, la contrainte de représentativité des origines des phénomènes régissant la présence des bactéries dans l'eau demeure importante non seulement pour une représentation acceptable des phénomènes observés mais aussi pour des prévisions fiables.

5.1.2. Problématique

La littérature abonde sur les modèles issus du couplage entre SIG / modèle hydrologique/ modèle de la qualité de l'eau (Bioteau *et al.*, 2002; Gomez. 2002; Arnold *et al.*, 2005; Boithias, *et al.*, 2011). L'intégration des SIG dans des modèles de la qualité de l'eau présente de nombreux avantages tant du côté scientifique que de la gestion. Dans un premier temps, les SIG facilitent le paramétrage effectif des modèles au niveau de la restitution spatiale des résultats pour différentes simulations. Par ailleurs, ils sont particulièrement adaptés pour gérer la question du passage d'une échelle à une autre, suite aux manipulations des différents niveaux d'appréhension de ce type d'outils. Ces atouts

font des SIG un support incontournable pour résoudre les problèmes environnementaux liés à la qualité de l'eau.

Pour la contamination des milieux aquatiques par des polluants chimiques ou biologiques, deux grands types de modèles fonctionnant à l'échelle du BV sont utilisés. Il s'agit des modèles événementiels et des modèles continus qui prennent en compte la discrétisation temporelle. Des exemples d'application de ce type de modèles sur les bassins versants réels témoignent de leur capacité à simuler correctement la qualité des eaux (Werner *et al.*, 2006; Sciuto et Diekkrueger, 2010).

On peut citer l'exemple du modèle BASINS (Better Assessment Science Integrating point et Nonpoint Sources, EPAa, 1996) qui a été développé par l'Agence américaine de protection de l'environnement. Le but de ce modèle est de simuler le devenir et le transport de l'azote, du phosphore et des coliformes fécaux (Lahlou *et al.*, 1998). Il s'agit d'un système informatique impliquant le couplage entre ArcView, les modèles hydrologiques HSPF/SWAT (Hydrologic Simulation Program-Fortran, EPA, 1996b; Soil and Water Assessment Tool; Arnold *et al.*, 2012; Arnold et Fohrer, 2005) et le modèle de la qualité des eaux QUAL2E (Brown et Barnwell, 1987). L'utilisation de cet ensemble de modèles reconnus et validés font du modèle BASINS l'un des modèles les plus adaptés et prisés en Amérique du Nord pour l'aide à la décision dans le cas de la gestion des BVs (Imhoff *et al.*, 2007).

Dans le même contexte, le modèle HSPF a été développé. Ce modèle s'appuyant sur un SIG a pour objectif de simuler le taux de ruissellement et les concentrations de contaminants, en réponse aux modifications de la couverture du sol et des précipitations. Deux ensembles distincts de processus sont ainsi représentés: (1) les processus qui déterminent le devenir et le transport des polluants à la surface ou dans le sous-sol d'un BV, et (2) les processus se produisant dans le réseau hydrographique. Cette particularité lui permet d'être applicable à de grands BVs et de prendre aussi en compte des sources de pollutions ponctuelles (Augusto et Leite, 2010). Cependant, plusieurs de ces modules sont dépourvus de réalisme physique (coefficient de mélange dans les réservoirs conceptuels) (Frere *et al.*, 1997). Par ailleurs, vu son incapacité à intégrer les pratiques de gestion agricole, ce modèle a tendance à sous-estimer les sédiments et les nutriments. En outre, il s'agit d'un modèle complexe. Il nécessite une quantité importante de données

pour caractériser les systèmes agricoles simulés et les conditions climatiques.

AVSWAT (ArcView, Soil and Water Assessment Tool) quant à lui simule la dynamique des flux des polluants dans les eaux à l'échelle d'un BV (Arnold *et al.*, 2005; Diluzio *et al.*, 2005). Ce modèle correspond au modèle SWAT couplé à ArcView. Son objectif est de prévoir l'impact des pratiques agricoles sur les BVs caractérisés par des hétérogénéités spatiales (Neitsch *et al.*, 2002; Lomakine, 2005). En 2000, un sous-modèle microbien a été développé et ajouté à SWAT pour simuler le devenir, le transport et la mortalité des bactéries fécales (Sadeghi et Arnold, 2002; Neitsch *et al.*, 2005; Diluzio *et al.*, 2012). L'ajout de ce module lui a permis d'être utilisé comme un outil permettant de traiter la contamination fécale de l'eau causée par des sources ponctuelles et non ponctuelles.

Au Québec, bien que la contamination microbienne des eaux de surface soit une préoccupation majeure pour les autorités de santé publique, peu d'études utilisent des méthodes adéquates permettant de comprendre cette situation. Or, dans ce contexte, il devient essentiel d'utiliser un modèle opérationnel et spécifique aux bactéries fécales, afin de pouvoir estimer de façon temporelle les apports fécaux se déversant dans les cours d'eau. Un outil réaliste et opérationnel est alors utile aux gestionnaires des bassins versants du Québec afin de comprendre et analyser l'impact des caractéristiques de la pluie sur la contamination fécale des eaux de surface en période estivale.

5.2. Objectifs

L'objectif principal de ce chapitre est de caractériser, par une approche géomatique, la dynamique temporelle des transferts d'*E. coli* à l'échelle d'un BV à vocation agricole. Il s'agit de prédire de manière temporelle la charge bactérienne véhiculée lors de la pluie dans les eaux récréatives, en tenant compte de la durée, de l'intensité et de la hauteur d'un épisode pluvieux. Cette contribution implique une démarche méthodologique de modélisation hydrologique centrée sur la mise en place d'une base de données physiographiques d'un BV couplée à une analyse des interactions entre les caractéristiques de la pluie et les hétérogénéités spatio-temporelles des usages du sol.

Les objectifs spécifiques consistent à :

- caractériser, à l'aide du modèle AVSWAT, la dynamique d'*E. coli* à l'échelle du BV, à la suite d'un « *flush* » pluviométrique, phénomène central à la prévision de la qualité microbienne des eaux récréatives;
- identifier et mettre en évidence les facteurs temporels de contrôle du transfert d'*E. coli*, et comprendre l'impact des interactions entre la variabilité spatio-temporelle des pluies et les hétérogénéités spatiales sur la contamination microbienne des eaux.

5.3. Description du modèle AVSWAT et son fonctionnement

AVSWAT, est un système de modélisation de la qualité de l'eau reliant le modèle agro-hydrologique SWAT aux SIG ArcView. Le but principal d'AVSWAT est l'évaluation des interactions entre les phénomènes climatiques, la végétation, les sols, la topographie et les activités agricoles sur les eaux de ruissellement. Il intègre également la simulation de l'érosion des terres, les flux de nitrate et phosphore et de bactéries fécales en fonction des épisodes pluvieux (Diluzio *et al.*, 2001; Neitsch *et al.*, 2005; Sadeghi et Arnold, 2002). La composante SIG du système, en plus des fonctions traditionnelles d'acquisition de données, de stockage, d'organisation et d'affichage, met en œuvre des méthodes analytiques de gestion avec une flexibilité accrue pour améliorer la caractérisation de la qualité de l'eau d'un BV. L'utilisation conjointe de l'extension du SIG ArcView dans SWAT permet alors de fournir une interaction efficace entre le modèle et les bases de données des paramètres associés. Cela permet finalement de simplifier les évaluations de la qualité de l'eau, tout en maintenant et en augmentant leur fiabilité.

5.3.1. Structure d'AVSWAT

Dans son fonctionnement, AVSWAT subdivise le BV en sous-bassins eux mêmes subdivisés en unités de réponse hydrologique (URHs). Ces dernières sont des zones homogènes en termes de réponse hydrologique et sont construites par le recoupement de l'occupation du sol et des unités pédologiques. Les différents bilans (eau, sédiments, nutriments, bactéries) sont calculés séparément pour chaque unité et additionnés dans chaque sous-bassin. Une fois que les valeurs ont été déterminées au niveau des sous-bassins, AVSWAT détermine les relations spatiales entre les sous-bassins et

incorpore les données calculées au réseau hydrographique global (Neitsch *et al.*, 2005; Lang, *et al.*, 2011). Le principe de délimitation des URHs est illustré à la figure 15.

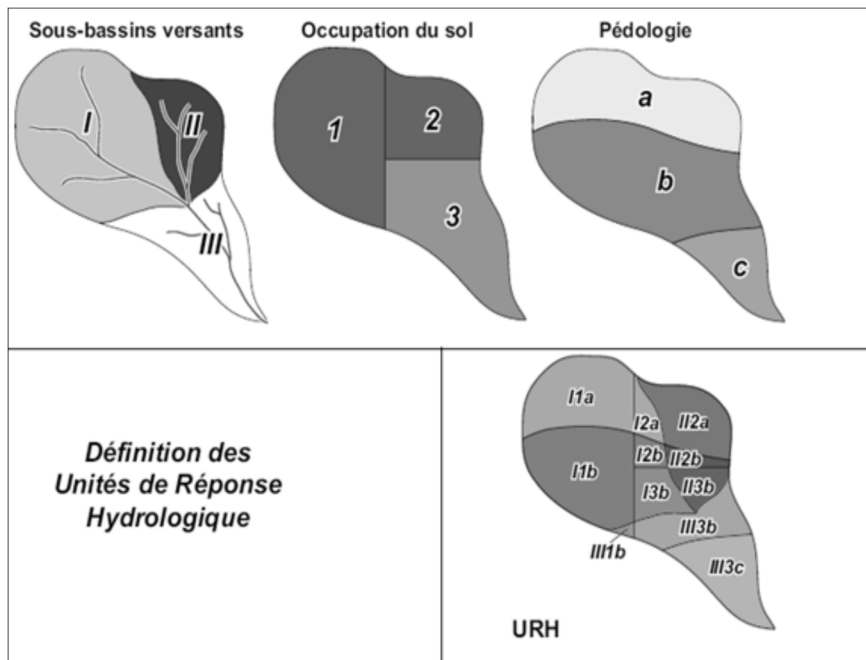


Figure 16. Discretisation spatiale dans SWAT (adapté de Rollo, 2012)

5.3.2. Modélisation des processus hydrologiques

La modélisation dans AVSWAT peut être séparée en deux parties : la première étape est la phase terrestre du cycle hydrologique. Celle-ci contrôle la quantité d'eau dans le cours d'eau principal de chaque SBV. La deuxième étape concerne le transport aquatique qui peut être défini comme le mouvement de l'eau, des sédiments, des bactéries à travers le réseau hydrographique du bassin jusqu'à l'exutoire (Didier Haguma, 2013).

Dans la phase terrestre, plusieurs composantes du bilan hydrologique sont modélisées. Il s'agit entre autre du ruissellement de surface, l'écoulement sub-surface ou latéral, l'écoulement souterrain ainsi que le transport des sédiments et des éléments nutritifs vers le réseau hydrographique. La simulation de cette phase terrestre, suivant un pas de temps journalier, est basée sur le bilan en eau suivant l'équation suivante :

$$SW_t = SW_0 + \sum_{i=1}^t (R_{day} - Q_{surf} - E_a - w_{seep} - Q_{gw}) \quad \text{Éq. 21}$$

où :

S_{wt} : quantité finale d'eau dans le sol (en mm)
 S_{w0} : quantité initiale d'eau dans le sol (en mm)
 R_{day} : précipitation totale (en mm)
 Q_{surf} : ruissellement total (en mm)
 E_a : évapotranspiration totale (en mm)
 W_{seep} : quantité d'eau du sol entrant dans la zone non saturée (en mm)
 Q_{gw} : quantité d'eau retournant dans le sol (en mm).

Les détails sur les processus impliqués dans le cycle hydrologique peuvent être trouvés dans Neitsch *et al.* (2005). Compte tenu de la spécificité de notre étude, nous décrirons uniquement la modélisation du ruissellement de surface et le transport des bactéries vers le réseau hydrographique.

5.3.2.1. Le ruissellement de surface

Le ruissellement de surface constitue la part de la pluie qui n'est pas absorbée par le sol et qui s'écoule dans le sens de la pente. Ce phénomène apparaît lorsque : (1) l'intensité des pluies est supérieure à l'infiltrabilité de la surface du sol ; (2) la pluie arrive sur une surface partiellement ou totalement saturée par une nappe (ruissellement par saturation). Son importance dépend de l'intensité des précipitations et de leur capacité à saturer la couche superficielle du sol, de la perméabilité des sols et de leur saturation en eau.

Le calcul du ruissellement de surface dans AVSWAT est simulé au choix selon les méthodes de Green et Ampt et celle du numéro de courbe du « Soil Conservation Service » (SCS, 1972). Dans le premier cas, il s'agit d'un modèle conceptuel d'infiltration et de précipitation. Il permet de représenter le refus à l'infiltration par dépassement de la capacité d'infiltration ou bien sur des sols saturés à l'aide d'une représentation simple du processus d'infiltration. Cette méthode a été utilisée dans un grand nombre de problèmes décrivant le mouvement de l'eau dans la zone non saturée, et dans des modèles hydrologiques (Freyberg *et al.*, 1980). Cependant, le fait que ce modèle repose sur des hypothèses simplificatrices limite son usage à certains cas d'infiltration. Il s'agit plus particulièrement des sols initialement secs et de texture grossière. Par ailleurs, dans les zones humides les variations de la conductivité hydraulique sont négligées (Morel-Seytoux, 1974).

Dans le deuxième cas, la méthode repose sur une caractérisation empirique établie entre la hauteur de précipitation en excès (ou production de ruissellement) et la hauteur de pluie. Elle permet de prendre en compte la capacité d'infiltration supposée décroissante au cours d'un épisode pluvieux. La relation de base du modèle SCS s'écrit comme suit (SCS, 1972) :

$$Q_{surf} = \frac{(R_{day} - I_a)^2}{R_{day} - I_a \times S} \quad \text{Éq. 12}$$

Le paramètre I_a correspond à la perte en eau initiale via les phénomènes d'interception, d'infiltration et de stockage de surface. Il est estimé à partir de l'absorption potentielle et d'un paramètre empirique α ajusté à 0,2 (SCS, 1972). La hauteur d'eau initiale ne participant pas à l'écoulement s'obtient comme suit :

$$I_a = 0,2 \times S \quad \text{Éq. 22}$$

S désigne la capacité maximale d'infiltration du sol (mm) et est directement reliée au numéro de courbe (CN). Elle est déterminée comme suit :

$$S = 25,4 \times \left(\frac{1000}{CN} - 10 \right) \quad \text{Éq. 23}$$

où la constante 25,4 de l'équation permet de transformer le résultat en mm; CN désigne le numéro de courbe (sans unité). Ses valeurs varient entre 0 (infiltration totale) et 100 (infiltration nulle). Plus celle-ci est élevée, plus le ruissellement est important. Cette valeur est fixée selon les informations hydrologiques, d'occupation du sol et pédologiques.

5.3.2.2. Phase aquatique : Transport et mortalité des bactéries

Une fois qu'AVSWAT a déterminé les débits, la phase suivante consiste au transport des sédiments, des nutriments, des pesticides et des coliformes dans la rivière. Comme énoncé précédemment, nous détaillons essentiellement le transport des bactéries compte tenu des objectifs de notre travail.

Plusieurs processus de transfert de bactéries fécales sont en œuvre à l'échelle d'un BV. Dans AVSWAT, trois voies de transfert vers les cours d'eau sont différenciées. Il

s'agit du transport particulaire, du transport libre et hypodermique. Le ruissellement de surface entraîne avec lui la fraction particulaire et la partie de la fraction dissoute. La mise en place de la quantité de coliformes ruisselés en mode libre s'accompagne du phénomène d'érosion. Lorsque les coliformes entrent dans le milieu aquatique, ils sont soumis à des facteurs édaphiques qui influencent leur survie. Leur durée de vie va donc dépendre des conditions physico-chimiques de l'eau, notamment la température et le pH de celle-ci. Considérant ces facteurs, la mortalité des coliformes dans AVSWAT est simulée à l'aide de l'équation de *Chick* avec un coefficient de mortalité selon la température :

$$C_t = C_0 \times e^{-K_{20}t\theta^{(T-20)}} \quad \text{Éq. 24}$$

où:

C_t : Concentration en coliformes au temps t (cfu/100ml)

C_0 : Concentration initiale en coliformes (cfu/100ml)

K_{20} : Taux de mortalité de 1^{er} ordre à 20 °C (jour-1)

t : temps d'exposition (jours)

θ : facteur d'ajustement de la température

T : température (°C)

Le module relatif aux coliformes dans AVSWAT est basé sur des processus régissant le transfert et le devenir des coliformes fécaux dans les cours d'eau. Différentes formulations mathématiques implémentées dans ce module répondent à la grande variété des phénomènes décrivant le comportement des coliformes au sol et dans les milieux aquatiques. De ce fait, ce module prend en compte:

- la quantité des coliformes solubles et adsorbés;
- le transport des coliformes sous forme libre, particulaire et hypodermique;
- les coefficients de partitionnement des coliformes entre les phases dissoute et adsorbée;
- la mortalité des coliformes au sol et dans les cours d'eau.

5.4. Méthodologie

Cette partie présente les procédures détaillées de la méthodologie développée pour la dynamique temporelle de la contamination fécale des eaux sur le BV du lac Massawippi avec AVSWAT. La démarche adoptée respecte les étapes de ce modèle (Arnold, 2012) et est illustrée à la figure 16. Nous détaillons d'abord le choix du modèle AVSWAT pour notre étude. Ensuite nous présentons les bases de données nécessaires pour l'implémentation de ce modèle sur le BV du lac Massawippi. Suivront ensuite, le choix des paramètres à caler et les différentes stratégies de paramétrisation utilisées. Les performances en calage sont discutées. Suivra une analyse de la sensibilité des paramètres dans le but d'évaluer la robustesse relative au modèle AVSWAT. Enfin le modèle est validé afin d'évaluer la robustesse des jeux de paramètres obtenus lors de calibration pour l'ensemble du modèle.

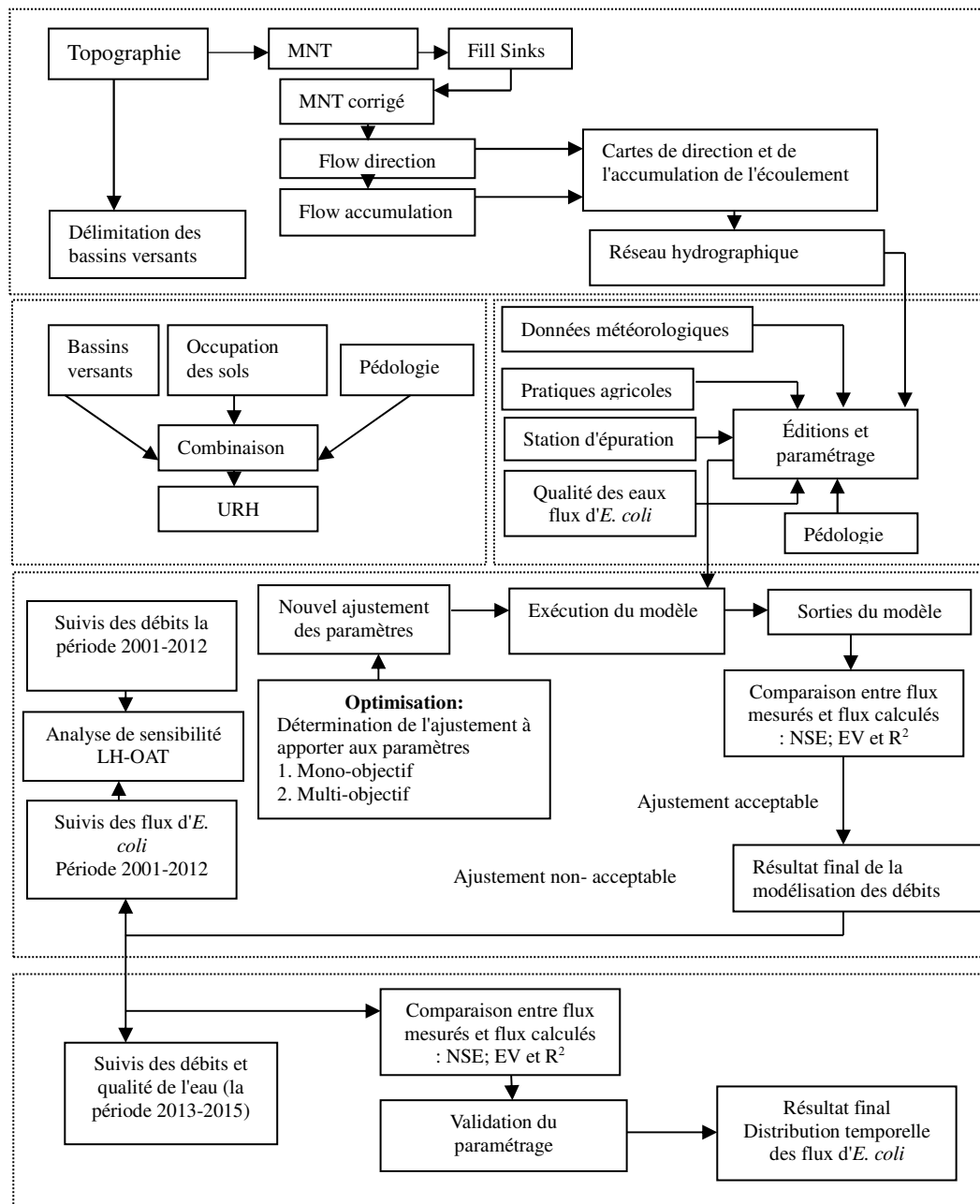


Figure 17. Méthodologie de la dynamique temporelle de la contamination des eaux

5.4.1. Choix du modèle AVSWAT

Après avoir pris connaissance d'un large ensemble de modèles, nous avons choisi d'utiliser AVSWAT pour la modélisation temporelle de la contamination fécale des eaux de surface au niveau du BV du lac Massawippi. Les principales raisons de cette sélection étaient son utilisation prédominante dans les BVs agricoles, son aspect interdisciplinaire et son efficacité de calcul de la mortalité d'*E. coli* (Guber, *et al.*, 2007). Au niveau du Québec, peu de recherches relatives à la modélisation des risques de contamination microbienne ont été effectuées avec AVSWAT. Or au vu de ses nombreuses capacités, l'utilisation de ce modèle événementiel, a été validée à l'échelle internationale pour la compréhension des risques de contamination microbienne des eaux de surface (Baffaut et Benson, 2003; Kang, *et al.*, 2006). Son principal avantage réside dans l'incorporation en une seule simulation des processus terrestres et aquatiques. Mais surtout, son aptitude à reproduire avec un réalisme acceptable le comportement d'un BV face à un risque de contamination fécale des eaux. Au niveau du module hydrologique, la possibilité de choisir la méthode du numéro de courbe permet d'adapter le modèle à une large gamme de conditions hydrologiques. Au niveau de l'utilisation, le téléchargement d'AVSWAT est libre sur internet; son code source ouvert et accessible constitue aussi un atout non négligeable. D'un point de vue plus opérationnel et pour une facilité d'utilisation et d'adaptation, son application et sa validation à d'autres sites européens et nord-américains sont des preuves rassurantes de la fiabilité de ce modèle.

5.4.2. Implémentation d'AVSWAT sur le bassin versant du lac Massawippi

5.4.2.1. Délimitation du bassin versant et extraction du réseau hydrographique

L'application du modèle AVSWAT sur un secteur consiste au préalable en la délimitation du BV et en l'extraction du réseau hydrographique. Pour ce faire, nous avons utilisé les procédures automatiques du SIG Arc/Info. La première étape a consisté aux traitements des dépressions et des zones plates présentes sur le MNT. Pour cela la fonction *FILL* de Arc/info a été utilisée. Cette méthode appliquée à un fichier en mode maillé, permet de combler les dépressions qui piègent l'eau en forçant l'eau à s'écouler dans une même direction à l'intérieur des zones plates. Cette fonction crée ainsi des zones horizontales

dans les vallées qui se marquent dans les profils en long des cours d'eau. Après les comblements des dépressions, la seconde étape a consisté à calculer les directions d'écoulements à partir du MNT sur l'ensemble de la zone d'étude.

Les méthodes de calcul de l'écoulement à partir de MNT abondent dans la littérature (O'Callaghan et Mark, 1984; Depraetere et Moniod, 1991; Chorowicz *et al.*, 1992; Tarboton *et al.*, 2009). La plus robuste et couramment utilisée est fondée sur l'algorithme D8. Cette méthode d'écoulement unidirectionnelle en 8 connexités, noté D8, définit le sens d'écoulement en chaque pixel à partir des valeurs altimétriques de ses voisins immédiats et choisit le pixel en aval suivant la descente maximale (O'Callaghan et Mark, 1984, Jenson et Dominique, 1988). Pour la présente étude, c'est cet algorithme qui a été choisi, *via* la fonction *Direction de flux (Flow direction)* pour générer les écoulements. La *Direction de flux* se calcule en recherchant la direction de la descente la plus raide ou la pente maximale de chaque cellule. La distance est calculée entre les centres des cellules. Par conséquent, si la taille de cellule est 1, la distance entre les deux cellules orthogonales est 1 et la distance entre deux cellules diagonales est 1,414, soit la racine carrée de deux. Si la pente maximale vers plusieurs cellules est la même, le voisinage est agrandi jusqu'à ce que la descente la plus raide soit trouvée. Lorsqu'une direction de descente la plus raide est trouvée, la cellule en sortie est codée avec la valeur représentant cette direction (Jenson et Domingue. 1988). Pour calculer le flux accumulé en chaque pixel à partir de la carte d'écoulement, nous avons utilisé la fonction *Accumulation de flux (Flow accumulation)*. En effet, les flux d'accumulation définissent l'amont du réseau hydrographique en nombre de cellules drainées dans chaque cellule. Cette fonction permet de calculer le flux accumulé sous la forme d'une pondération cumulée de toutes les cellules s'écoulant dans chaque cellule en pente descendante du raster en sortie. Une fois les cartes de direction et d'accumulation de l'écoulement déterminées, nous avons procédé à l'extraction du réseau hydrographique. Cette étape a été réalisée à partir d'un seuil d'apparition de l'écoulement de manière à prendre en compte le plus grand nombre de cours d'eau en tête de BV. Pour cela nous avons choisi, après avoir essayé différentes valeurs entre 0,1 et 0,5 km², une valeur intermédiaire de 0,2 km².

5.4.2.2. Discrétisation spatiale et temporelle

Une fois le BV délimité et le réseau hydrographique extrait, l'étape suivante a consisté en l'intégration des cartes d'occupation du sol avec celle de la pédologie dans le modèle. Cette étape permet d'intégrer ces différentes cartes aux bases de données internes du modèle. Celles-ci regroupent l'ensemble des valeurs alphanumériques nécessaires aux simulations. Il s'agit entre autres des périodes d'épandage de fumier, de la rotation des cultures des précipitations, des températures. Ces valeurs alphanumériques sont par défaut alimentées par les propriétés de différents types de sols nord-américains et par de nombreuses catégories d'occupation du sol. Pour cela, il a donc été nécessaire de ré-classifier les représentations spatiales pour les associer aux entrées disponibles dans la base de données.

Au cours de cette étape, différentes méthodes de délimitation des URHs sont disponibles en fonction du degré de simplification. Afin d'aboutir à une discrétisation fine, nous avons choisi la délimitation multiple intégrant l'ensemble des types de sols et des classes d'occupation répertoriées sur les différents bassins (FitzHugh et Mackay, 2000).

Au niveau de la discrétisation temporelle, le modèle a été appliqué avec un pas de temps journalier sur une échéance temporelle de cinq ans, de 2009 à 2014, divisée en deux périodes :

- une période de calage (2009-2011),
- une période de validation (2012-2014).

Cette discrétisation temporelle s'applique aussi bien pour la simulation des débits que celle des flux de coliformes.

5.4.2.3. Optimisation des paramètres

Toute modélisation de la contamination fécale des eaux doit avant tout partir d'une bonne optimisation du modèle. Cette phase a pour but de trouver un jeu de paramètres qui reproduisant correctement le comportement du bassin. Elle est d'un intérêt pratique pour une bonne modélisation de la qualité de l'eau, car elle permet de définir un critère pour évaluer la similitude entre les comportements observés et simulés du bassin.

La procédure de calage proposée dans AVSWAT peut-être soit manuelle soit

automatique. Dans le premier cas, l'ajustement des paramètres est effectué par essai-erreur et requiert l'exécution manuelle de plusieurs simulations consécutives. Il s'agit d'une tâche laborieuse, complexe et difficilement reproductible (Boyle *et al.*, 2000; Moradkhani et Sorooshian, 2008; Pechlivanidis *et al.*, 2011). Par ailleurs, un autre inconvénient est la grande subjectivité d'un critère d'ajustement visuel (Jackson et Aron, 1971).

Afin d'éviter ce problème de subjectivité liée à l'approche manuelle, l'approche automatique est utilisée. Il s'agit de l'utilisation d'algorithmes d'optimisation qui, de façon objective et reproductible, explorent l'espace des solutions (jeux de paramètres) afin de minimiser la valeur de la fonction objectif. Dans le cadre de cette étude, nous avons utilisé l'algorithme Shuffled Complex Evolution - University of Arizona (SCE UA).

SWAT2005 inclut une procédure automatique de calage multi-objectif et d'analyse d'incertitude, appelée Parasol (Parameter Solutions method), développée par van Griensven et Bauwens (2003). Cette procédure de calage est basée sur l'algorithme « Shuffled Complex Evolution- Uncertainty Analysis » ou SCE-UA. Pour le calage des débits nous avons testé les calages mono-objectif et multi-objectif en utilisant tour à tour NSE, R^2 et le pourcentage de biais. Une fois la calibration des paramètres influant sur la réponse hydrologique a été jugée complète, nous avons procédé à la calibration des paramètres permettant de caractériser le devenir et la mortalité des coliformes en utilisant également les calages mono et multi-objectifs.

5.4.2.4. La sensibilité du jeu optimal à la fonction objectif

Préalablement au calibrage des paramètres nous avons procédé à une étude de la sensibilité. En effet, le modèle agro-hydrologique AVSWAT contient des centaines de paramètres qui représentent les processus hydrologiques et de la qualité des eaux. Avec un aussi grand nombre de paramètres, l'analyse de sensibilité a consisté à identifier les paramètres influençant le plus les débits et le flux de coliformes. Cette étude quantitative est importante dans l'élaboration de modèle. Elle permet à travers les étapes de calibration et de validation à diminuer l'incertitude liées aux valeurs des paramètres d'entrée, et à améliorer les performances prédictives du modèle AVSWAT (Arnold et

Fohrer 2005, Holvoet *et al.*, 2005; 2008; 2007).

La méthode de sensibilité utilisée dans AVSWAT est une approche hybride de simulation d'Hypercube latin et d'échantillonnage d'un paramètre à la fois *via* Latin Hypercube – (*One factor At a Time*, ou LH-OAT). Il s'agit d'une méthode d'analyse de sensibilité locale qui étudie comment des perturbations autour d'une valeur des paramètres se répercutent sur la valeur de la sortie. Cette méthode présente l'avantage de la robustesse de *Monté Carlo* tout en optimisant le nombre de calculs (van Griensven *et al.*, 2006). La méthode LH-OAT subdivise la distribution de chaque paramètre en N intervalles égaux, et échantillonne au hasard les valeurs du paramètre dans chaque intervalle. Dans le cas de cette étude, nous avons modifié alternativement chaque paramètre de +15 % et -15% autour de sa valeur initiale. Dans la littérature de nombreux auteurs attestent que ce pourcentage est acceptable par rapport aux imprécisions probables liées aux paramètres d'entrée du modèle AVSWAT (Favis-Mortlock et Smith, 1990).

5.4.2.5. Évaluation des performances du modèle

L'optimisation des paramètres a permis d'obtenir un jeu de paramètres optimum pour la période de calage. Suite à cette optimisation, nous avons procédé à l'évaluation de la performance du modèle AVSWAT à partir des trois indices suivants :

1- La valeur du coefficient de détermination (R^2) : elle rend compte de la proportion totale des variances des valeurs observées qui forment la valeur explicative du modèle. Elle mesure la justesse de l'adéquation entre les valeurs prédites et les valeurs observées. Les valeurs de ce coefficient vont de $-\infty$ à 1. Plus, ces valeurs sont proches de 1, meilleures sont les prédictions du modèle. Le coefficient R^2 est calculé selon la formule suivante (Tolson et Shoemaker, 2004) :

$$R^2 = \frac{[\sum_{i=1}^n (Qs_i - \overline{Qs}) (Qm_i - \overline{Qm})]^2}{\sum_{i=1}^n (Qs_i - \overline{Qs})^2 \times \sum_{i=1}^n (Qm_i - \overline{Qm})^2} \quad \text{Éq. 25}$$

où

R^2 : coefficient de détermination

Qm_i : valeurs observées

QS_i : valeurs simulées

\overline{Qm} : moyenne des valeurs observées

Q_S : moyenne des valeurs simulées

2. Le critère de Nash et Sutcliffe (NSE; 1970) : ce critère, très largement utilisé en hydrologie, varie dans l'intervalle $[-\infty, 1]$. Il caractérise la proportion de la variance initiale des débits expliquée par le modèle. Une valeur de NSE inférieure à zéro signifie que le modèle de débit constant qui est égal à la moyenne de valeurs observées explique mieux le comportement du bassin que le modèle proposé. Une valeur de NSE supérieure à 0,75 est considérée comme bonne, alors qu'une valeur comprise entre 0,75 et 0,36 est satisfaisante (Ahl *et al.*, 2008). Si le NSE est négatif ou très proche de zéro, les prédictions du modèle sont alors considérées comme inacceptables (Santhi *et al.*, 2001). Le critère Nash-Sutcliffe est calculé selon l'équation ci-dessous :

$$NSE = \frac{1 - \sum_{i=1}^n (Qm_i - QS_i)^2}{\sum_{i=1}^n (Qm_i - \overline{Qm})^2} \quad \text{Éq. 26}$$

où

Qm_i : valeur mesurée

QS_i : valeurs simulées

\overline{Qm} : moyenne des valeurs observées

n : nombre de simulations.

3. Le pourcentage de biais (PBAIS) mesure la tendance moyenne des données simulées à être au-dessus ou au-dessous des données observées. La valeur optimale du PBAIS de 0 indique une exactitude de la simulation du modèle. Les valeurs positives indiquent une sous-estimation du biais, et les valeurs négatives indiquent une surestimation du biais (Santhi *et al.*, 2001). Ce critère a été choisi en raison de sa capacité de démontrer clairement la faible performance du modèle. Celui-ci est calculé selon l'équation ci-dessous :

$$PBAIS = \left[\frac{\sum_{i=1}^n (V_o - V_s) \times 100}{\sum_{i=1}^n (V_o)} \right] \quad \text{Éq. 27}$$

où

PBAIS: : pourcentage de biais

V_o : valeurs observées

V_s : valeurs simulées

5.4.2.6. Validation du modèle

La validation constituait la dernière étape du processus. Il s'agissait de tester la robustesse du modèle avec les paramètres optimaux issus du calage en période de validation. Pour cela, nous avons estimé les valeurs moyennes des paramètres calés en les appliquant sur les évènements qui n'ont pas servi au calage.

5.5. Résultats et interprétation

Nous présentons les résultats en fonction des étapes méthodologiques effectuées. La première partie présente les résultats des analyses de sensibilité pour les débits et les coliformes fécaux. La deuxième partie présente l'analyse des performances du calage. Cette analyse porte particulièrement sur la comparaison entre les résultats de la simulation et les données observées disponibles après ajustement suivant les deux stratégies de calage. Enfin, la dernière partie de cette section s'articulera sur l'influence des scénarios temporels sur l'ensemble des processus de mobilisation et du transport temporel des coliformes fécaux dans les eaux.

5.5.1. Analyse de sensibilité

Les figures 17, 18, 19 et 20 illustrent la sensibilité du débit et des taux de coliformes en fonction de l'augmentation et la diminution (+ 15 et -15 %) de différentes valeurs des paramètres. Ces résultats ont permis d'identifier précisément les paramètres pour lesquels il est nécessaire de fournir des efforts pour une meilleure estimation du débit et des taux de coliformes.

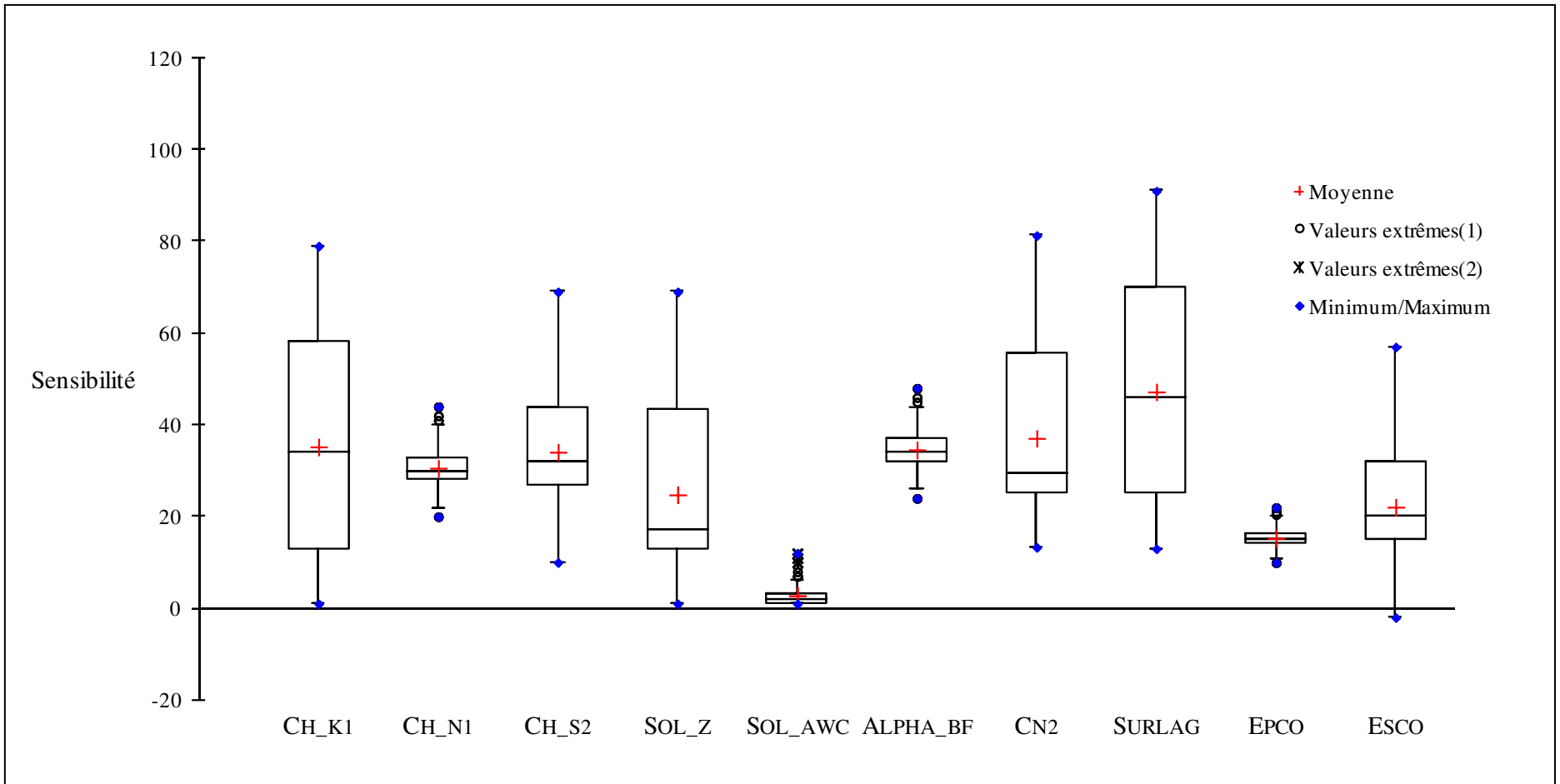


Figure 18. Sensibilité du débit à une augmentation des paramètres (+15 %)

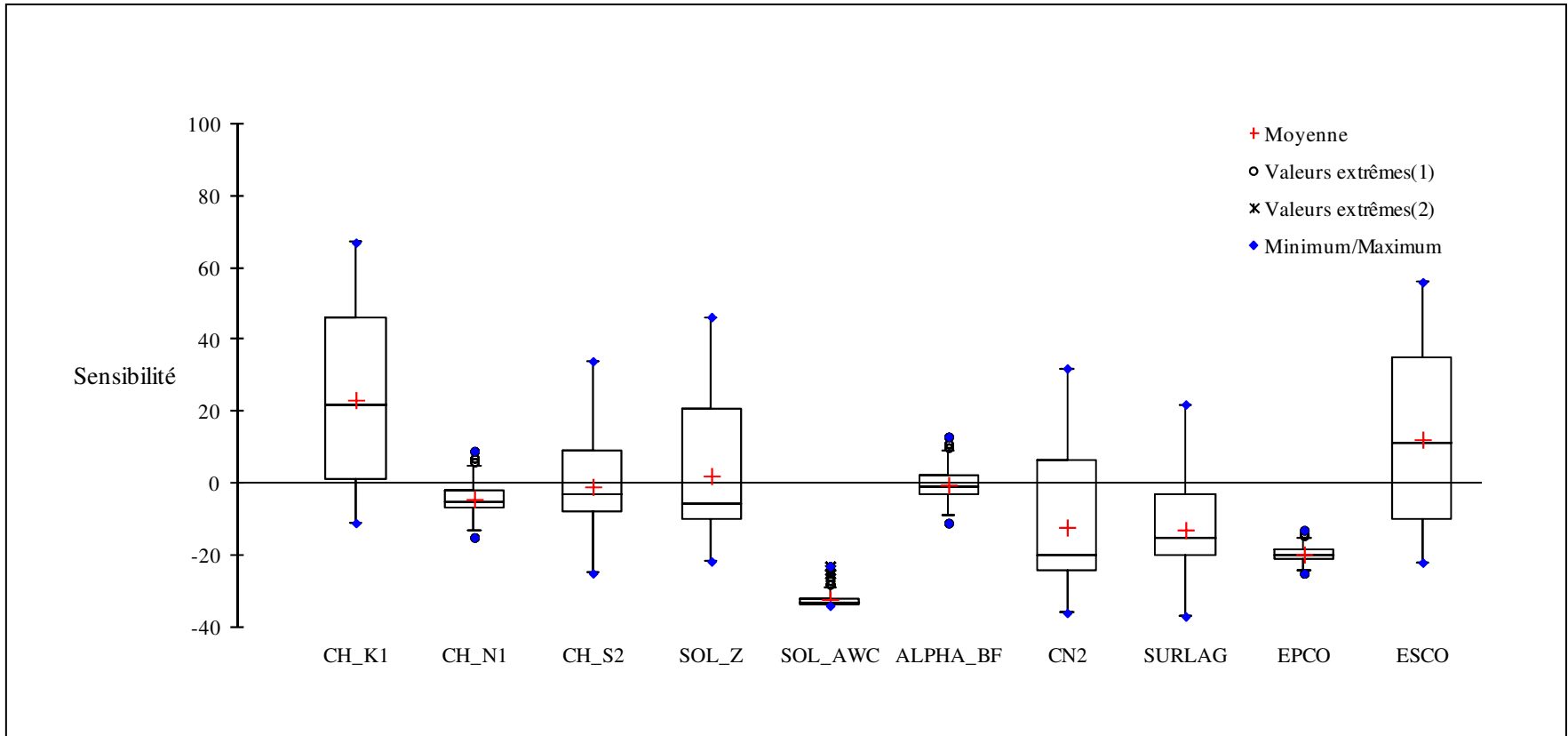


Figure 19. Sensibilité du débit à la diminution des paramètres (-15%)

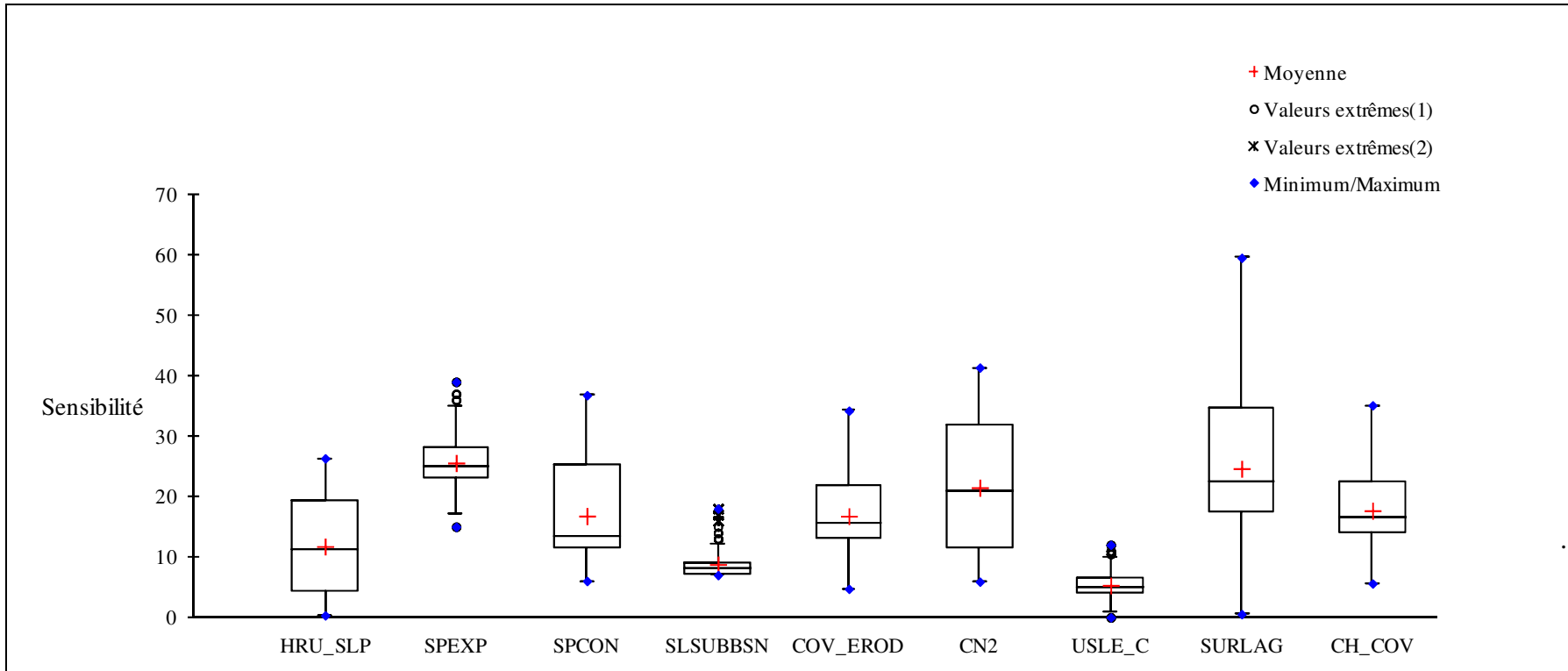


Figure 20. Sensibilité des flux de coliformes à une diminution des paramètres (-15 %)

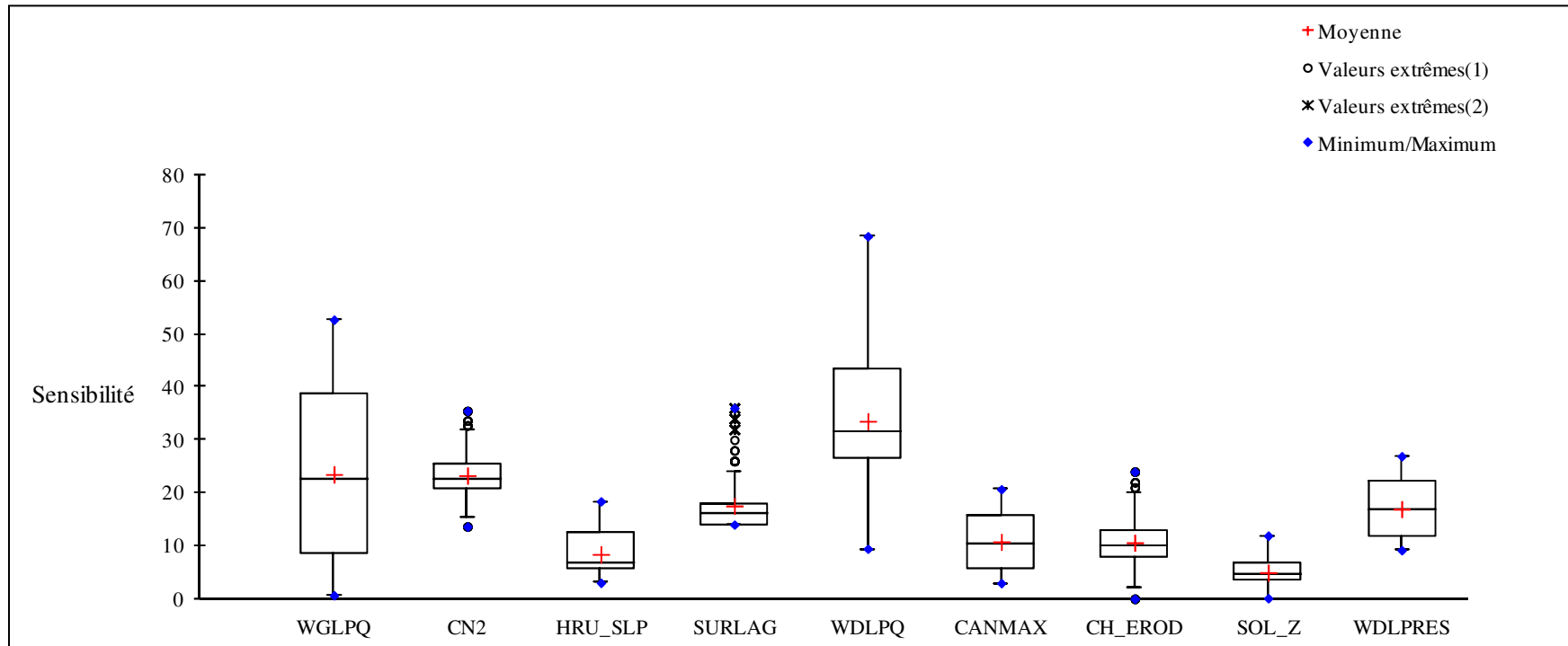


Figure 21. Sensibilité des flux de coliformes à une augmentation des paramètres (15%)

L'analyse de sensibilité sur les débits (figure 17) a mis en évidence que parmi les trente et un (31) paramètres analysés (annexe 9), seulement dix ont montré une sensibilité au débit. Il s'agit en premier des paramètres relatifs au ruissellement de surface notamment le coefficient de délai du ruissellement (SURLAG) et le coefficient de détermination de la répartition des précipitations entre l'écoulement de surface et l'infiltration (CN2). Une augmentation de 15 % de chacun de ces deux paramètres occasionne une variation significative de la sensibilité atteignant respectivement 91 et 81 %. Le constat est similaire à la diminution de 15 % des valeurs de ces deux paramètres (figure 18). Dans ce cas on observe cependant des variations négatives du débit de plus de -40 %. Dans l'ensemble SURLAG et CN2 sont deux paramètres très importants qui influencent significativement la réponse hydrologique sur le BV du lac Massawippi. Après les paramètres relatifs au ruissellement, le deuxième groupe de paramètres affectant le débit se compose des paramètres physiques des canaux tels que la conductivité hydraulique, le coefficient de Manning et le facteur de couverture du cours d'eau (CH_K, CH_N, CH_COV). Ces derniers ont également des impacts importants sur les variations des débits. Leurs contributions aux variations du débit peuvent varier jusqu'à plus de 50 %. Enfin, le troisième groupe de paramètres affectant le débit est constitué

- des facteurs de compensation du calcul d'évaporation notamment le facteur limitant la reprise par évaporation dans le sol et le facteur limitant la reprise d'eau du sol par les plantes (EPCO et ESCO);
- du facteur de débit de base (ALHA_BF); de la profondeur du sol (SOL_Z).

Une analyse de sensibilité des flux de coliformes (figure 19) montre qu'en diminuant simultanément les valeurs des paramètres des débits, les résultats mettent en évidence que SURLAG et CN2, constituent là encore la principale source de variation des flux de coliformes. La sensibilité du modèle à ces deux paramètres varie entre 59 à 41 % respectivement. Ces résultats mettent en exergue le rôle prépondérant du ruissellement dans l'acheminement des coliformes dans les eaux. Le deuxième groupe dont l'influence sur les coliformes est également grande comprend les paramètres relatifs au transport des sédiments dans le cours d'eau. Il s'agit entre autres des paramètres linéaire et exponentiel régissant la charge sédimentaire maximale dans les cours d'eau

(SPCON et SPEXP), et des facteurs d'érodabilité et de couverture des cours d'eau (CH_EROD et CH_COV). Les résultats révèlent que la contribution de ces paramètres sur la variation des flux de coliformes peut atteindre des valeurs dépassant 30 %. Les caractéristiques de la pente du BV notamment son inclinaison, (HRU_SLP), la longueur maximale des ruisseaux (SLSUBBSN) et les caractéristiques apparaissent aussi sensibles. Leurs contributions tournent respectivement autour de 26 et 18 %. Il apparaît également que la variation du facteur d'USLE (USLE_C) joue un rôle non négligeable dans la variation des flux de coliformes. Cet effet se fait sentir significativement à partir d'un certain degré et est plus marquant si le stock de matières en suspension est grand.

La même analyse en augmentant les paramètres relatifs au devenir et à la mortalité des coliformes (figure 20), montre que les deux premiers paramètres relatifs au ruissellement dont l'influence est beaucoup plus importante pour les deux précédentes optimisations constituent la seconde source de variation du modèle avec des contributions de moins de 40 %. Par ailleurs, les facteurs de mortalité et de croissance des bactéries solubles et adsorbées (WDLPQ et WGLPQ) sont ceux qui possèdent les plus fortes valeurs de sensibilité, pour la variation du flux de coliformes (les paramètres relatifs à la mortalité et au devenir des bactéries sont présentés en annexe 10). Leur augmentation engendre une variation importante des flux de coliformes dont les valeurs sont proches de 70 %. Le facteur de croissance des bactéries dans les eaux récréatives (WDLPRES) a également une forte influence sur la variation du flux de coliformes. Son effet est généralement plus important si on lui associe une grande valeur (deux semaines). Elle atteint une sensibilité proche de 30 %. Cependant, la variation du flux de coliformes avec ce paramètre n'est pas linéaire et reste faible lorsque la valeur de demi-vie des bactéries est moins d'une semaine. Il est également intéressant de noter que les flux de coliformes sont influencés par les paramètres d'érodabilité (CH_EROD) et la profondeur du sol (SOL_Z) avec des valeurs de sensibilité moyenne avoisinant 10 %. Des résultats similaires sont obtenus par Shrestha *et al.* (2007); Zhang, *et al.* (2010) qui soulignent les interactions entre l'érosion et la fonction de transfert des coliformes dans les eaux.

En définitive, si les matières en suspension constituent un moyen de transport pour les bactéries, la part de matières érodées en surface qui ruissellent sur le chemin vers l'exutoire apparaît ainsi importante pour l'estimation du flux de coliformes. Cet aspect

joue un rôle suffisant pour que le paramètre CH_EROD joue un rôle qui peut devenir comparable à celui du ruissellement.

Que ce soit le débit ou le flux de coliformes, les analyses révèlent que les paramètres relatifs au ruissellement apparaissent toujours les plus influents de chacune des variables (débits et flux de coliformes). En effet, le formalisme du modèle AVSWAT pour l'élaboration de la contamination microbienne par des coliformes fixe d'abord la réponse hydrologique du BV. Celle-ci est déterminée à partir de SURLAG et CN, puisque ces derniers permettent d'augmenter la génération des flux de drainage profond et de ruissellement. En fait, une augmentation de ces deux paramètres s'accompagne d'une augmentation du débit de base et par conséquent du volume total simulé. Cet accroissement du ruissellement peut alors entraîner une érosion, favorisant ainsi le transport des sédiments et des bactéries. Par ailleurs, ces deux paramètres régissent la répartition des précipitations entre l'écoulement de surface et l'infiltration. Ces effets importants expliquent sans doute les influences, nettement plus élevées de ces deux paramètres comparativement aux autres. Ces derniers paramètres interviennent indirectement sur la présence des bactéries dans les eaux, en accélérant leur déplacement tout en affectant positivement ou négativement leur nombre dans les eaux.

De ces résultats, il ressort également que les taux de croissance et de mortalité des bactéries adsorbées jouent un rôle déterminant sur le flux de coliformes acheminés vers l'exutoire du BV. Mais, cet aspect fondamental est principalement contrôlé par les paramètres de ruissellement. En revanche, la sensibilité obtenue confirme que la profondeur du sol a une influence modérée sur la partition entre infiltration et ruissellement et donc sur le transfert des coliformes dans les eaux. En d'autres termes, on retrouve logiquement que, seuls, les paramètres de ruissellement, d'érodabilité et de mortalité contrôlent grandement le flux de coliformes atteignant les cours d'eau.

5.5.2. Analyse des performances suivant les deux stratégies d'optimisation

Les tableaux 15 et 16 présentent les performances du modèle AVSWAT en fonction des deux stratégies de calage (mono-objectif et multi-objectif) selon les trois critères de performances : NSE, le coefficient de détermination (R^2) et le pourcentage de biais (P.Biais).

Tableau 15. Résultats de l'optimisation uni-objective avec différentes combinaisons de paramètres

| | Optimisation mono-objectif | | | |
|----------------|----------------------------|------------|---------------------|------------|
| | Phase de calage | | Phase de validation | |
| | Débits | Coliformes | Débits | Coliformes |
| NSE | 0,53 | 0,56 | 0,52 | 0,39 |
| R ² | 0,57 | 0,49 | 0,34 | 0,26 |
| P.Biais | -36 | +37 | +47 | -53 |

Tableau 16. Résultats de l'optimisation multi-objective avec différentes combinaisons de paramètres

| | optimisation multi-objectif | | | |
|----------------|-----------------------------|------------|---------------------|------------|
| | Phase de calage | | Phase de validation | |
| | Débits | Coliformes | Débits | Coliformes |
| NSE | 0,43 | 0,35 | 0,70 | 0,63 |
| R ² | 0,39 | 0,45 | 0,65 | 0,58 |
| P.Biais | -29 | -52 | -23 | +19 |

L'analyse de l'ensemble des résultats (tableaux 15 et 16) permet de constater que les meilleures performances en calage sont obtenues au niveau de l'optimisation multi-objective. Aussi bien pour les débits que pour le flux de coliformes, ces observations concernent essentiellement la période de validation pour laquelle les critères NSE et R² affichent les valeurs comprises entre 0,70 et 0,65 et entre 0,63 et 0,58 respectivement pour les débits et le flux de coliformes. Le pourcentage de biais quant à lui atteint les valeurs les plus basses, -23 et +19 % pour les débits et le flux de coliformes respectivement. De telles valeurs sont acceptables, et témoignent d'une simulation satisfaisante (figure 21 et 22). Ces valeurs expliquent une bonne prise en compte des débits de transport des coliformes, et le facteur indispensable d'une bonne représentation de la mortalité et du devenir des coliformes dans le réseau hydrographique. Malgré les meilleures performances obtenues du calage multi-objectif, les débits simulés par le modèle en période de validation pour les mois de janvier demeurent légèrement surestimés (figure 21).

Au niveau des flux de coliformes, pour certains mois comme juillet, août et septembre, le modèle sous-estime les flux de bactéries de l'ordre de +19 % à +35 % (figure 22). Ces résultats du modèle laissent supposer qu'une partie des bactéries n'est pas simulée ou que certains prélèvements n'ont pas été effectués au moment des pics, ce qui peut conduire à une analyse sous-évaluée. Par ailleurs, en phase de calage on observe une surestimation des flux des bactéries avec des valeurs de biais de pourcentage qui atteignent (-52). Ces résultats laissent supposer une mauvaise évaluation de la contribution aux flux de bactéries due à une mobilisation très rapide ou relativement importante causée par un débit intense. Il peut également s'agir des événements pluvieux après épandage qui entraînent plus de bactéries dans le réseau hydrographique. Hormis ces surestimations, la simulation du flux des bactéries apparaît corrélée de manière satisfaisante aux mesures avec les valeurs de coefficient de détermination de 0,58 en phase de validation.

Au niveau de l'optimisation mono-objective (tableau 14) nous pouvons constater que lors de la période de calage, les critères de Nash atteignent parfois des valeurs supérieures à 0,50 pour certaines simulations. Ces résultats peuvent être qualifiés de satisfaisants. Mais ces derniers, ne garantissent pas toujours de meilleures performances en validation (figures 21 et 22). Par exemple, les performances des débits et des coliformes déjà meilleures en période de calage avec les jeux de paramètres identifiés par l'efficacité de Nash et le coefficient de détermination se dégradent en période de validation (tableau 14). En principe, il semble que le calage mono-objectif n'arrive pas à trouver l'optimum à partir des valeurs aléatoires de paramètres. Ainsi dans la majorité des cas, la méthode multi-objective est plus performante que la méthode mono-objective en calage.

Après analyse des résultats à la fois sur les sites de calibration et les sites de validation, nous avons constaté une augmentation globale des résultats des simulations pour la phase de validation en optimisation multi-objectif. Ces performances pourraient être le fait d'un plus grand nombre de données utilisées pendant la validation, de plus les paramètres des coliformes et débits sont simultanément pris en compte lors du calage. Cela pourrait s'expliquer de la façon suivante : la correction des paramètres relatifs au débit permet d'ajuster et augmenter les bilans au niveau des taux de coliformes, ce qui revient à influencer

sur les valeurs de Nash. Or, ce dernier est a priori, impacté par les changements de critères lors du calage. Ainsi, il paraît logique de constater la même tendance sur la correction des taux de coliformes lors de la validation avec un nombre de données plus élevé.

De façon générale, il ressort de cette analyse que les deux méthodes d'optimisation obtiennent des performances différentes au niveau de la validation, avec résultats plus satisfaisants pour la méthode multi-objective. Par ailleurs, la validation utilisant NSE comme fonction objective apparaît la plus robuste comparativement au critère R^2 qui ne permet pas de respecter le bilan hydrologique et conduit à de mauvaises simulations des flux de bactéries; il ne peut alors être utilisé comme seule fonction, pour la validation des simulations. En somme, il ressort que le modèle AVSWAT à travers l'optimisation multi-objective conduit à une meilleure représentation des caractéristiques hydrodynamiques du devenir des bactéries au vu de leur performance vis-à-vis des critères Nash et du pourcentage de biais.

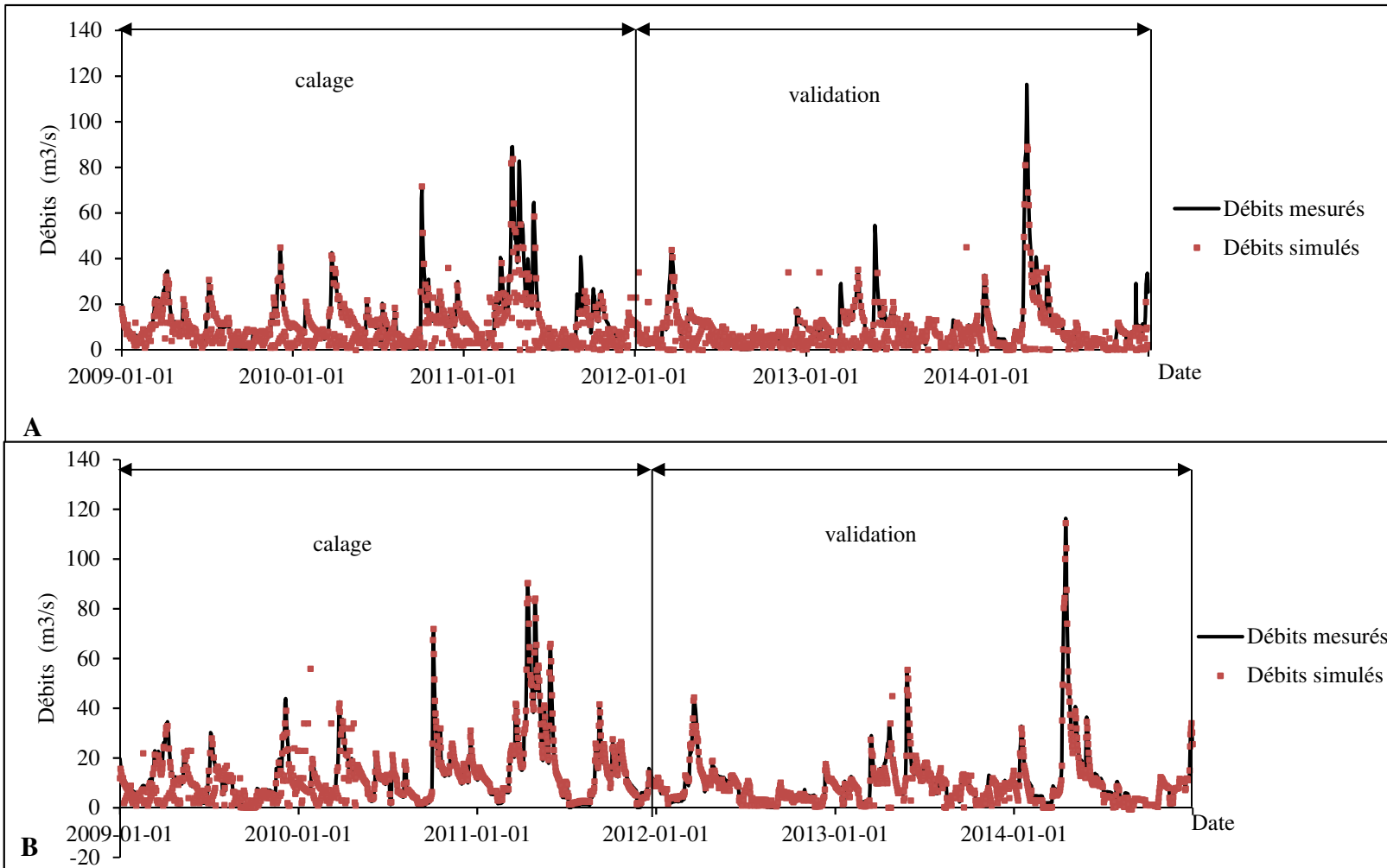


Figure 22. Résultats de calage mono-objectif (A) et multi-objectif (B) pour les débits

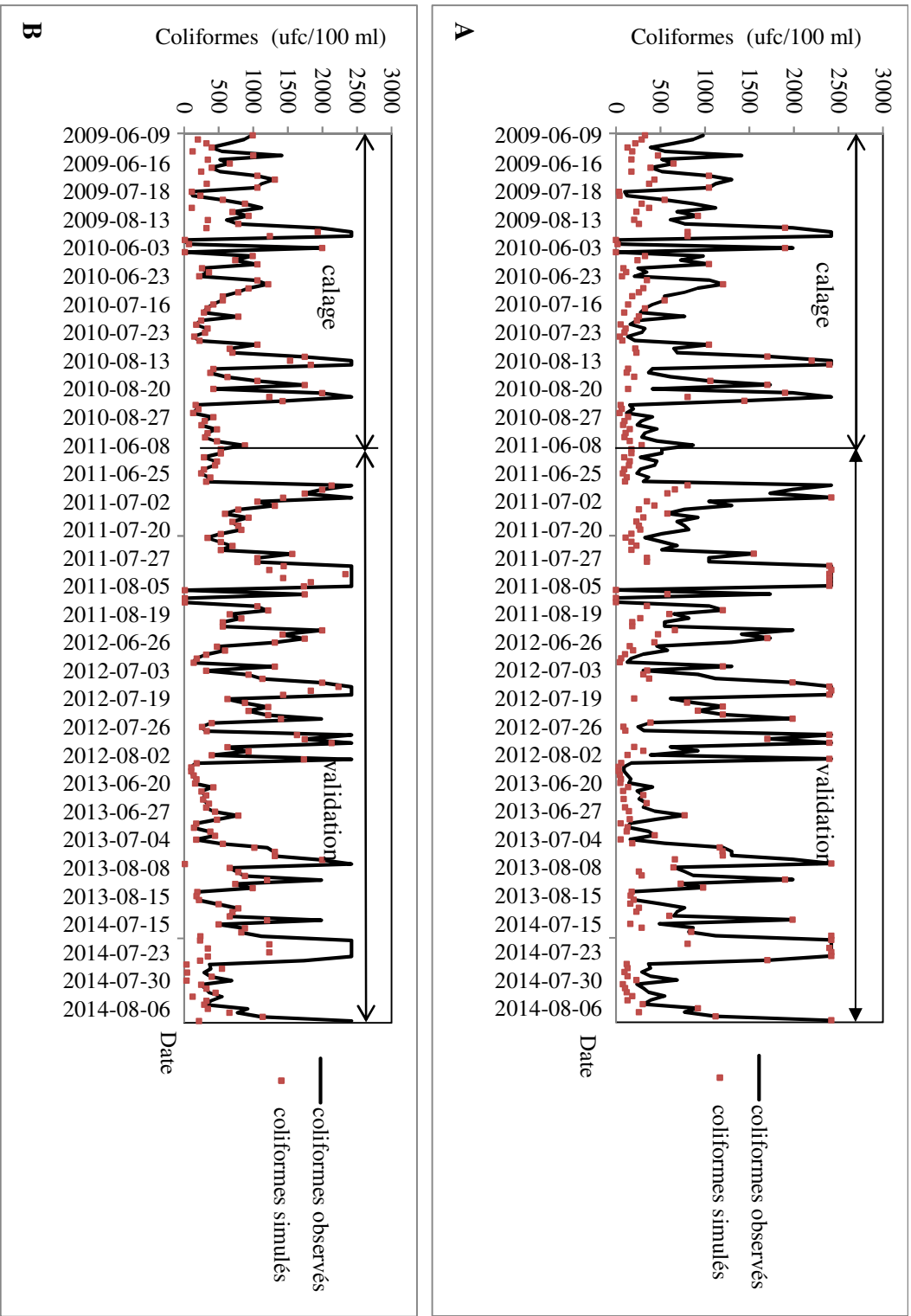


Figure 23. Résultats de calage mono-objectif (A) et multi-objectif (B) pour les coliformes

5.5.3. Prédications du modèle en fonction des scénarios temporels

Une fois les débits et les flux de coliformes bien simulés, nous avons cherché à analyser la variabilité temporelle des flux de coliformes. L'attention a été particulièrement portée sur les mois de mai, juin, juillet et août, périodes où les plages des lacs sont ouvertes pour la baignade estivale. Pour ce faire, les données pluviométriques du modèle ont été modifiées. Il ne s'agissait pas des scénarios de changement climatique, car seules, la durée, la hauteur moyenne et l'intensité moyenne des pluies ont été modifiées. Afin d'appréhender l'influence de ces caractéristiques pluvieuses sur la contamination microbienne des eaux, nous avons réalisé trois scénarios différents.

- Le premier scénario permet d'étudier l'influence de la hauteur de la pluie sur la contamination microbienne. Pour ce faire, on a considéré une pluie de durée moyenne en faisant varier les hauteurs de 10, 50 et 100 mm tout en gardant les intensités faibles (inférieur à 50 mm/h);
- Le deuxième scénario permet d'étudier l'influence de la durée de la pluie sur la contamination microbienne. Pour ce faire, on a considéré trois pluies de faible, moyenne et longue durée. Ces pluies sont communément caractérisées par des hauteurs inférieures à 30 mm et des intensités inférieures à 50 mm/h. Les durées des pluies sont de 30 min; 2h et 8 h;
- Le troisième scénario permet d'étudier l'influence de l'intensité de la pluie sur la contamination microbienne. Pour ce faire, on a fait varier les intensités de la pluie de 10, 50 et 100 mm/h, pour une pluie d'une durée moyenne et de hauteur faible.

Au départ, nous émettons l'hypothèse que les scénarios reposent sur une comparaison des flux de coliformes obtenues avec un stock uniforme de fumier et un coefficient de partition égal. Les résultats des scénarios sont illustrés sur les figures 23 à 25.

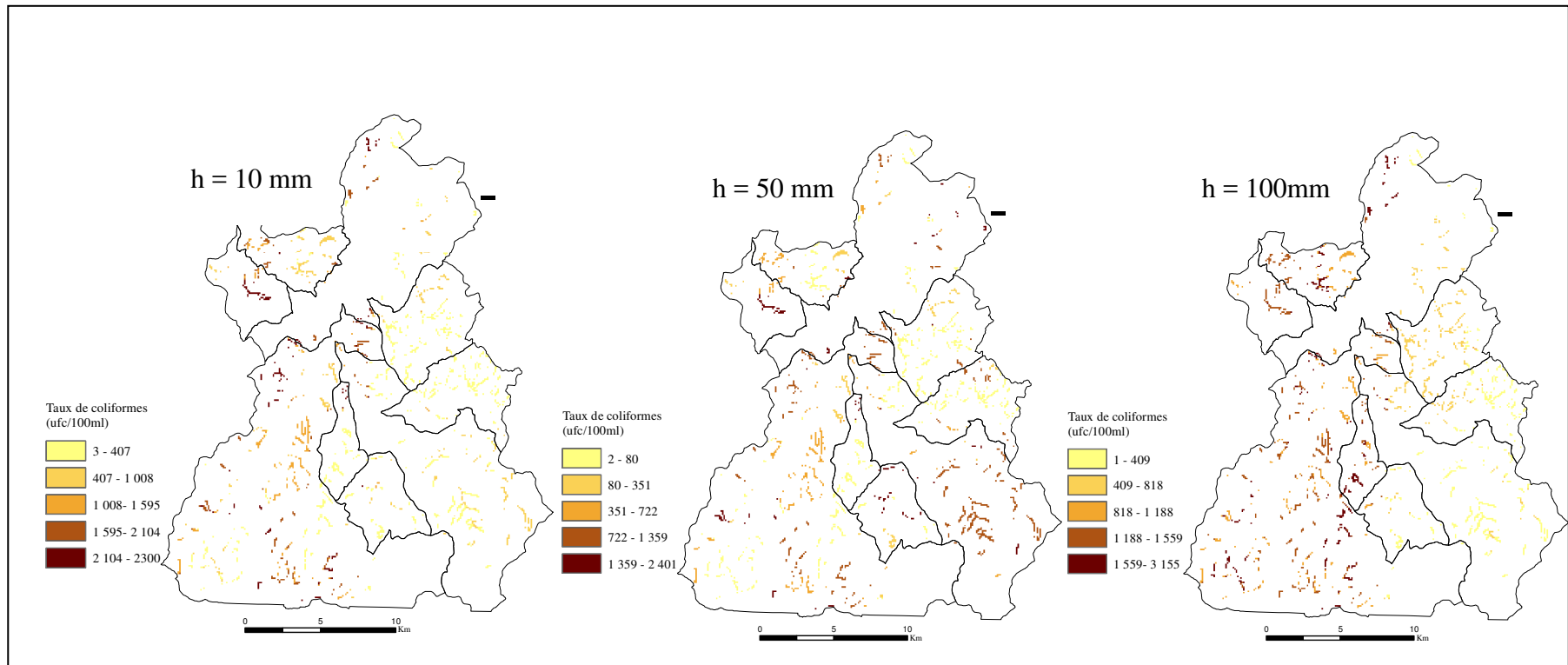


Figure 24. Distribution des flux de coliformes en fonction de hauteur de la pluie

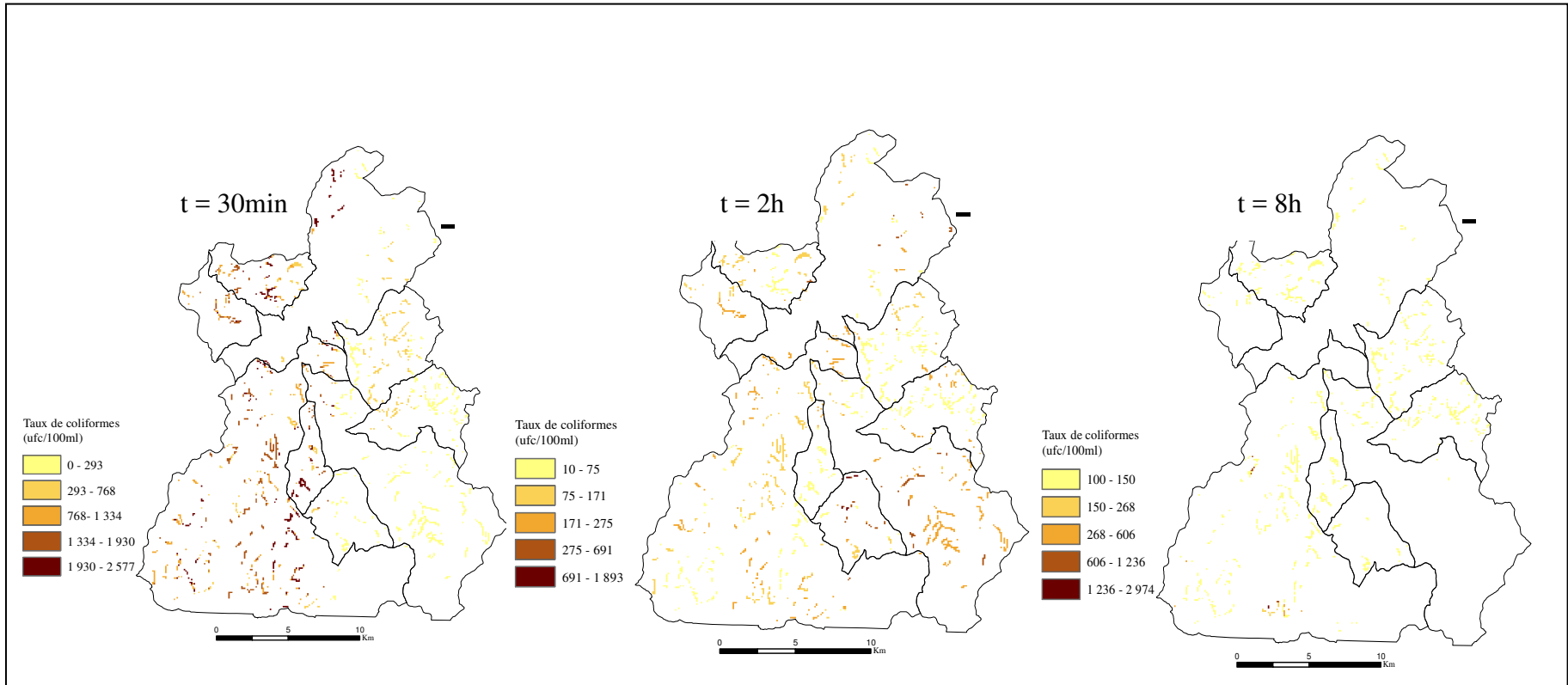


Figure 25. Distribution des flux de coliformes en fonction de la durée de la pluie

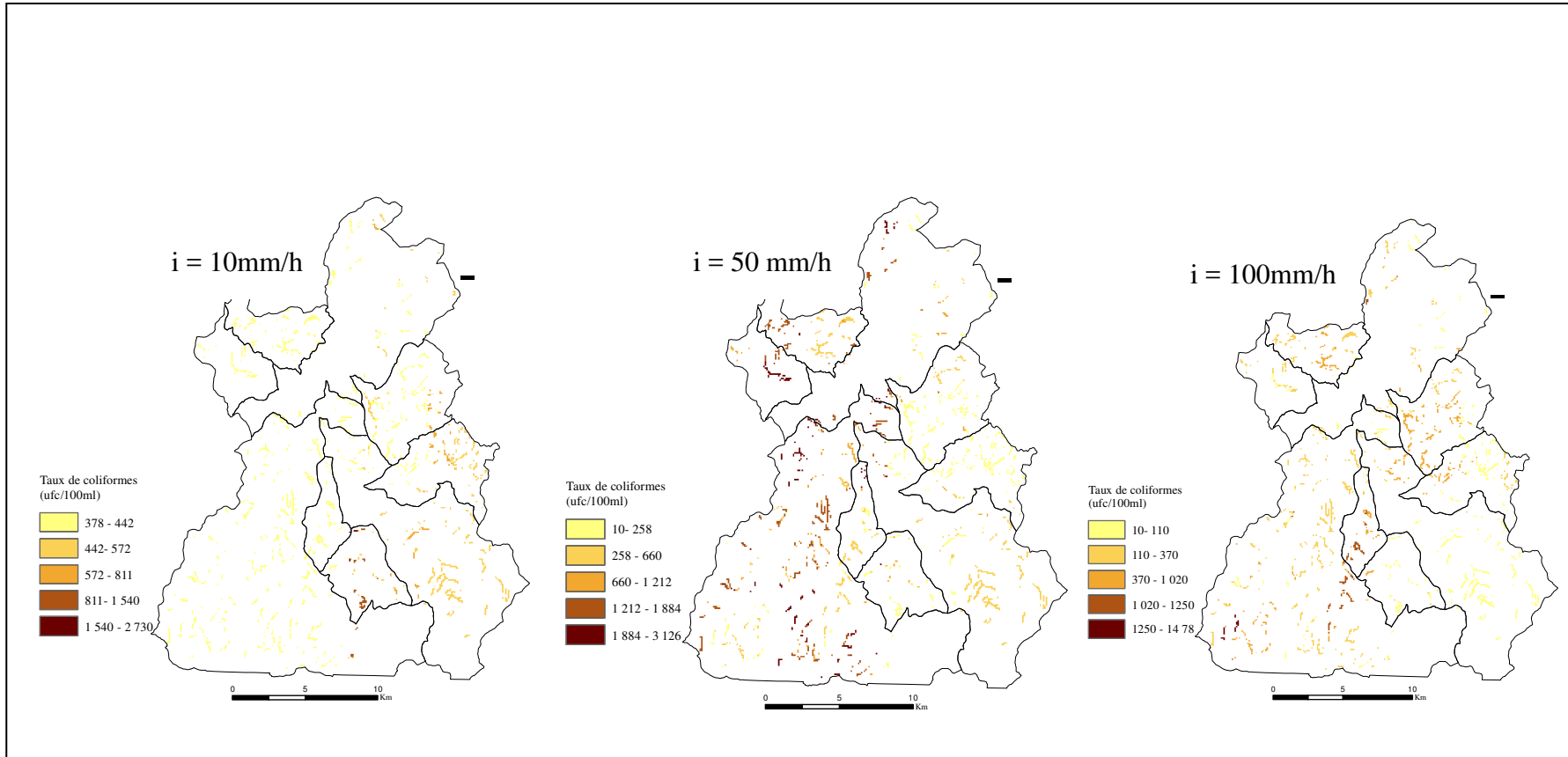


Figure 26. Distributions des flux de coliformes en fonction de l'intensité de la pluie

La figure 23 présente la variation du taux de coliformes en fonction de la hauteur de la pluie. Les résultats montrent que sur le BV du lac Massawippi, les taux de coliformes varient entre 1 et 3 155 ufc/100 en fonction de la hauteur de pluie. Plus celle-ci augmente plus les taux de coliformes atteignent de grandes valeurs. On peut voir sur la carte (figure 23) que les taux de coliformes les plus élevés ont été obtenus pour le cas où la hauteur de la pluie est égale à 100 mm. Cependant, les résultats de l'analyse de corrélation font apparaître de grands contrastes entre le SBV de la Tomifobia et le reste des SBVs du lac Massawippi (figure 26). Par exemple, le SBV de la Tomifobia présente davantage une augmentation fréquente des coefficients de corrélation en fonction de la hauteur contrairement aux autres SBV. Au niveau de ce SBV, on constate une évolution temporelle du taux de coliformes marquée par une corrélation assez nette entre les hauteurs de pluie et la contamination microbienne. Par exemple pour une hauteur de 50 mm, le coefficient de corrélation est de 0,33; lorsque la hauteur cumulée de la pluie est de 100 mm, la valeur de ce coefficient atteint 0,65.

À l'opposé, la hauteur de la pluie ne montre aucune corrélation directe avec le taux de coliformes sur les autres SBV. Pour ces derniers, les valeurs des coefficients de corrélation oscillant entre 0,12 et 0,42 révèlent une situation stratifiée et relativement irrégulière.

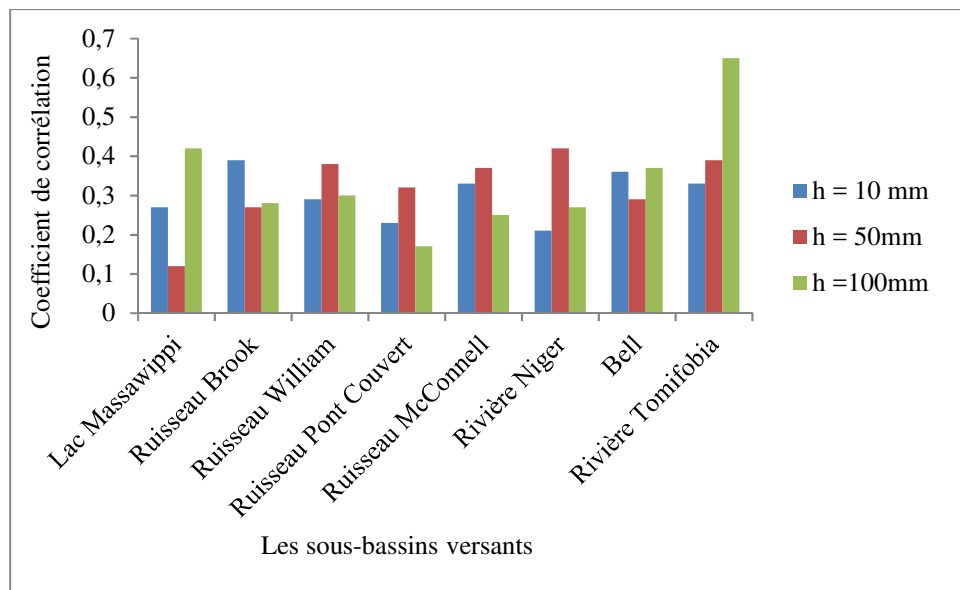


Figure 27. Variation des coefficients de corrélation en fonction de la hauteur de la pluie par sous-bassin versant.

La figure 24 présente la variabilité des taux de coliformes en fonction de la durée de la pluie. L'analyse des trois graphiques montre clairement que les taux de coliformes décroît en fonction de la durée de la pluie pour l'ensemble des BV. Cette variation visuelle corrobore les corrélations obtenues entre la durée de la pluie et les flux de coliformes sur les SBV (figure 27). De façon générale, les coefficients de corrélation varient entre 0,21 et 0,61 en fonction des SBV avec une moyenne globale de l'ordre de 0,4. L'ensemble des résultats fait apparaître que les valeurs de coefficient de corrélation les plus basses sont obtenues au niveau du SBV de la rivière Niger avec une corrélation moyenne d'environ 0,28. Mais cette valeur n'est pas très différente de celles des autres SBV dont le niveau de corrélation est de l'ordre de 0,3.

Par ailleurs, les valeurs de coefficient de corrélation les plus élevées ont été obtenus au niveau des SBV de la Tomifobia et du ruisseau McConnell avec des corrélations moyennes de 0,49 et 0,50 respectivement. Au niveau de ces deux SBV, les valeurs maximales de coefficients de corrélation ont été atteintes pour le scénario $t = 2h$. Au-delà de deux heures pour la majorité des SBV, il n'existe plus de relation significative entre la durée de la pluie et le taux de coliformes.

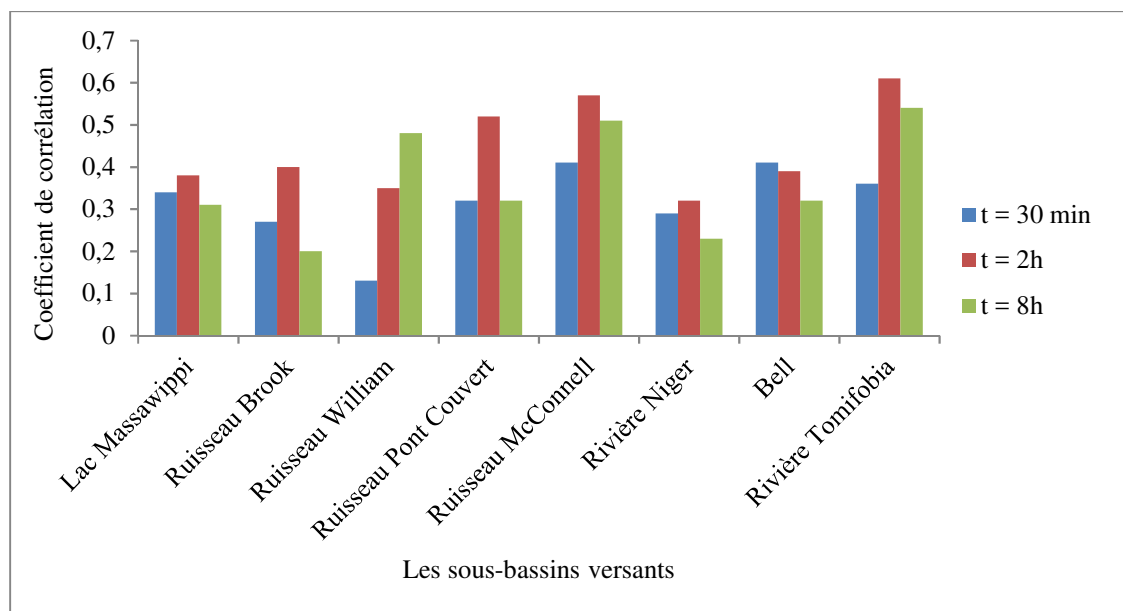


Figure 28. Variation des coefficients de corrélation en fonction de la durée de la pluie par bassin versant

En ce qui concerne l'influence de l'intensité de la pluie sur les taux de coliformes, les résultats sont présentés à la figure 25. L'analyse de ces résultats montre que le taux de coliformes varie entre 10 et 3126 ufc/100 en fonction de l'intensité de la pluie. Les taux moyens les plus élevés sont atteints lorsque $i = 50$ mm/h.

L'analyse des coefficients de corrélation (figure 28) montre que c'est au niveau du scénario $i = 10$ mm/h que les valeurs de coefficient de corrélation les plus faibles sont enregistrées. Bien que majoritairement faibles, quelques SBV se sont démarqués par des valeurs de coefficients de corrélation élevées. C'est le cas des SBV de la rivière Niger et celui de la Tomifobia dont les coefficients de corrélation sont les plus élevés avec des valeurs de 0,33 et 39 respectivement.

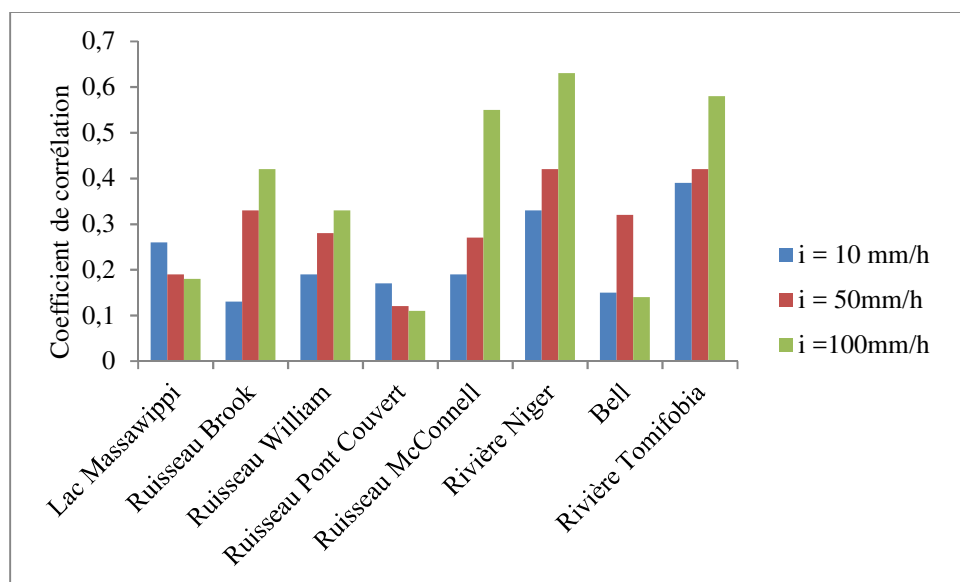


Figure 29. Variation des coefficients de corrélation en fonction de la durée de la pluie par sous-bassin versant

En comparant ces résultats ($i = 10$ mm/h) avec ceux obtenus par le scénario $= 50$ mm/h nous constatons que les valeurs des coefficients de corrélation augmentent, mais restent faibles pour quelques SBV. Pour ce scénario ($i = 50$ mm/h), le coefficient de corrélation moyen n'est que de 0,29. Par ailleurs, nous remarquons une légère différence entre les résultats obtenus du scénario $i = 50$ mm/h et ceux du scénario $i = 100$ mm/h. Au niveau de ces deux scénarios, on observe une augmentation significative des coefficients de corrélation en fonction de l'intensité de pluie. Ceci est vrai pour les SBV de la Tomifobia, de la rivière Niger, et celui du Ruisseau McConnell (figure 28). Au niveau de ces SBV, l'effet de l'intensité de la pluie se fait fortement ressentir. Les coefficients de corrélation atteignent leurs valeurs maximales respectivement 0,58,

0,63 et 0,55, et elles sont significatives ($p\text{-value} < 0,05$). Partant de ces cas, nous constatons qu'une intensité de 100 mm/h a un effet positif sur les taux de coliformes en augmentant les valeurs des coefficients de corrélation.

De ces résultats, nous pouvons retenir que, parmi les différents facteurs qui jouent un rôle dans le transport des coliformes, les plus importants concernent l'intensité de la pluie. Plus l'intensité des pluies est petite, plus le transport des bactéries est moindre dans une échelle limitée et plus les taux de coliformes sont petits et inversement. Les valeurs de coefficients de corrélation montrent que l'intensité des pluies joue un rôle plus décisif que la hauteur totale et la durée des pluies. D'ailleurs, on peut constater que les intensités de pluies égales à 10 mm/h mm sont sans grande importance du point de vue de l'effet de transport des coliformes. C'est à partir de 50 mm/h que se trouve la limite inférieure des intensités des pluies permettant la corrélation entre l'intensité de la pluie et le taux des coliformes. La corrélation est encore plus significative lorsque l'intensité de la pluie est 100 mm/h. Cependant, il existe une variabilité parmi les différents BV au point de vue de la corrélation entre les caractéristiques de la pluie et les taux moyens de coliformes. C'est-à-dire que pour chaque BV, il existe une hauteur, une durée et une intensité de pluie nécessaires pour un transport plus remarquable des coliformes correspondant à une valeur déterminée. Cette proportionnalité des variables caractéristiques des pluies et les taux de coliformes varient en fonction des conditions naturelles du bassin, des propriétés physiques, de la surface totale du bassin et de la longueur moyenne des bassins.

5.6. Conclusion partielle

Le contexte actuel de la contamination microbienne des eaux récréatives au Québec met en évidence le besoin d'outils de modélisation pour appréhender la dynamique temporelle des polluants bactériens dans les eaux. L'objectif de la présente étude a donc été de caractériser par une approche géomatique la dynamique temporelle des coliformes à l'échelle d'un BV à vocation agricole. Pour ce faire, nous avons abordé le sujet de façon méthodique par l'utilisation d'une modélisation couplant les composantes temporelles aux taux de coliformes fécaux. La synthèse des principaux résultats de sensibilité a révélé un système principalement contrôlé par les paramètres relatifs au ruissellement. Le processus d'érodabilité a également une grande influence sur le transport des coliformes. Les deux méthodes d'optimisation pour la calibration des

paramètres ont, dans l'ensemble, fourni de bons résultats. La démarche de comparaison de différentes stratégies de calage a révélé que l'optimisation multi-objective est celle qui a donné les meilleurs résultats selon l'analyse des critères Nash et du pourcentage de biais. Elle a réussi à mieux reproduire la dynamique relative aux débits et aux coliformes. Cependant, pour l'optimisation mono-objective, en calage, le modèle donne des bonnes réponses pour les débits et flux de coliformes, mais en période de validation, le modèle n'est pas capable d'aboutir à une bonne représentation des débits et des coliformes.

L'analyse des scénarios permettant de répondre à l'objectif de ce travail révèle une dépendance assez étroite entre l'intensité de la pluie et la contamination microbienne des eaux. Avec les faibles intensités, les coliformes lessivés sont principalement ceux présents sur le bassin et adsorbés par les matières en suspension. À mesure que l'intensité augmente, à partir d'un certain seuil de débit, le transport des coliformes augmente et la concentration des coliformes mesurée à l'exutoire du bassin augmente. Après une certaine intensité, le flux devient presque constant ou nul. Ceci traduit un épuisement progressif du stock de coliformes disponible et facilement mobilisable. Cette proportionnalité semble dépendre des caractéristiques géométriques qui composent les sous-bassins.

De façon générale, ce travail constitue une des rares approches de modélisation au Québec en termes d'utilisation du modèle AVSWAT permettant une modélisation réaliste de la dynamique temporelle de la contamination microbienne des eaux récréatives. L'avantage de cette étude est que cette approche est basée en priorité sur une idée de prévision événementielle et de gestion du risque, plutôt que sur une simple estimation spatio-temporelle des flux de coliformes. La méthodologie que nous avons mise en œuvre permet d'aller au-delà des simples corrélations entre les caractéristiques d'une pluie et la survenue d'un épisode de contamination microbienne. Cette approche qui est souvent utilisée pour la modélisation de la qualité chimique des eaux de surface, a été mise en œuvre dans le cas d'une contamination microbiologique des eaux en période de pluie. Après l'introduction dans le modèle des caractéristiques d'un événement pluvieux, des cartes de variabilité temporelle de la contamination microbienne ont été obtenues afin de prendre en compte l'influence des caractéristiques de pluies sur la contamination microbienne des eaux de surface. Un tel modèle, avec une représentation temporelle des processus régissant la contamination des eaux, permet d'améliorer la compréhension des processus en jeu. L'utilisation de paramètres temporels ajoute beaucoup à la compréhension du

transport des coliformes dans les eaux. Compte tenu du peu d'information disponible dans la bibliographie concernant le devenir des bactéries fécales, les relations entre les caractéristiques de la pluie et du flux de coliformes pourraient permettre d'élaborer des contraintes sur la réponse événementielle attendue d'un BV.

Chapitre 6. Discussion générale

Cette discussion intègre et synthétise les principaux résultats obtenus en liaison avec les trois objectifs de ce travail à savoir : (i) la modélisation spatiale de la contamination microbienne des eaux récréatives par une approche probabiliste *PdlP*; (ii) l'analyse spatiale de la sensibilité du modèle *PdlP*; (iii) la variabilité temporelle de la contamination microbienne des eaux : une approche géomatique à l'échelle d'un BV agricole. Ces trois objectifs seront revus successivement afin que la compréhension des principales forces et limites méthodologiques des approches mises en œuvre aident à évaluer leur validité et mieux appréhender la variabilité spatiale de la contamination microbienne.

6.1. La modélisation spatiale de la contamination microbienne des eaux récréatives par une approche probabiliste *PdlP*

La variabilité spatiale de la contamination microbienne des eaux récréatives constitue un enjeu pour la préservation de la santé des baigneurs et l'évaluation des risques encourus. L'utilisation conjointe du modèle probabiliste *PdlP* et des SIG était bien adaptée pour mieux analyser la variabilité spatiale selon les caractéristiques agro-pédo-climatiques. Les résultats montrent que la distance entre les cours d'eau et les zones agricoles est un facteur déterminant dans l'estimation de la contamination microbienne. Une analyse plus détaillée des preuves de ce facteur montre que le contraste le plus élevé est celui de la classe moyenne 5-10 km avec une valeur de $C = 0,87$. Mais dans l'ensemble, les résultats confirment la dynamique interactive entre la superficie agricole, la pente, la distance ainsi que la pluie dans la variabilité spatiale de la contamination microbienne des eaux.

Les meilleures concordances (les valeurs d'ASC) entre les cartes simulées et observées sont détectées au niveau des bassins versants de Yamaska-Nord et Yamaska Sud-Est. Cette concordance spatiale semble être une fonction de la superficie agricole. Une analyse plus approfondie montre une assez-bonne similarité entre ces deux bassins versants. On y retrouve des superficies agricoles de 28 et 25 % respectivement pour les sous-bassins versants de Yamaska-Nord et Sud- Est. La prédiction du modèle *PdlP* pour ces deux bassins présente de meilleures performances descriptives avec les valeurs d'ASC supérieures à 70 %. Toutefois, pour d'autres BV tels que Savail et Pot Au Beurre, cette concordance s'est révélée être assez différente et médiocre. Ces tendances, qui ont été déjà observées dans des études antérieures (Ascough *et*

al., 1999; Aspinall *et al.*, 2000; Aubertot *et al.*, 2005) sont principalement dues au microrelief, à la localisation des mesures, et surtout à la variation spatiale de la capacité d'infiltration (texture, stabilité structurale) du sol. Ces études soutiennent que d'un BV à un autre, le transport des bactéries dépend plus des propriétés hydrodynamiques des sols, de leur capacité à retenir l'eau ainsi que de leur sensibilité au ruissellement.

Beauduin *et al.* (2004) et Balestrat (2006) corroborent le fait que le transport et le devenir des contaminants dans l'environnement sont significativement affectés par leur participation aux réactions de sorption. Par exemple, ces auteurs ont démontré que dans un sol argileux, le temps moyen nécessaire à une réduction de 95 % des populations de bactéries est de 53 jours lorsque le sol est saturé en eau, de 38 jours lorsqu'il est à la capacité au champ et de 22 jours lorsqu'il est à une humidité de 50 % de la capacité au champ. Dans ce type de sol, les bactéries restent donc longtemps à la surface et sont de ce fait plus sensibles à l'entraînement par l'eau, donc à la contamination microbienne des eaux.

Sans minimiser l'effet des types de sol qui n'intervient pas dans notre modèle, les résultats de cette étude rejoignent sur certains points ceux de quelques études antérieures à l'effet que l'occupation du sol et les facteurs morphologiques sont les plus importants dans le processus de variabilité spatiale de la contamination microbienne des eaux. Même si elles doivent être interprétées avec nuance, étant donné les limites méthodologiques qu'elles semblent comporter, les études de Wilkinson *et al.* (1995), Crowther *et al.* (2001), Gardi (2001), Collins et Rutherford (2004), George *et al.* (2004), Srinivasan *et al.* (2005) et Kay *et al.* (2007) suggèrent aussi que les usages anthropiques, les caractéristiques morphologiques du BV ainsi que les conditions météorologiques sont les paramètres qui pilotent la variabilité spatiale d'un contaminant sur un BV. Toutefois, nous sommes d'avis que les conditions pédologiques peuvent améliorer les prévisions spatiales de la variabilité de la contamination microbienne, sans en être la principale cause.

En termes de paramètres clés affectant la variabilité spatiale de la contamination des eaux sur un BV, les résultats de cette étude sont également en accord avec ceux de Bousquet et Didier (1999) et Houdart *et al.* (2005). Bien qu'il soit difficile de comparer nos approches, ces études antérieures proposent de privilégier des modèles intégrés en s'appuyant en particulier sur les systèmes multi-agents (SMA). Ces derniers offriraient probablement de meilleures qualités de prédiction. Ces auteurs mentionnent que la simulation multi-agents ne se contente pas d'intégrer

des paramètres environnementaux puis d'analyser la réponse du modèle en fonction des scénarii. Elle participe pleinement au processus de recherche du fonctionnement des systèmes complexes.

Bien que les caractéristiques spatiales soient d'une importance capitale à la prédiction de la variabilité spatiale de la contamination microbienne, certains auteurs comme Garcia *et al.* (2005) recommandent que celles-ci doivent toujours être associées avec la mortalité des bactéries pour obtenir des résultats plus performants. Dans ce contexte, Camesano et Logan (1998) et Simoni *et al.* (1998) ont développé la théorie de filtration basée sur un modèle de rétention irréversible de 1^{er} ordre afin de décrire les variations spatiales des micro-organismes. Ce modèle tient compte de la concentration bactérienne initiale introduite dans le sol, de la concentration bactérienne mesurée à l'instant t dans le sol, de la porosité de la couche de sol, de la distance parcourue par l'eau, du coefficient de filtration et de l'efficacité de collision.

Devant la complexité des mécanismes en jeu, une modélisation réaliste de la contamination microbienne devrait considérer l'action combinée de divers paramètres physico-chimiques et biologiques qui interagissent avec les bactéries. Techniquement réalisable (Garcia-Lara *et al.* 1991), cette approche reste expérimentale car l'interprétation biologique est délicate. Sinon, cette modélisation ouvrirait la voie vers le suivi du bon état des eaux récréatives.

Au vue de l'ensemble d'autres variables pouvant influencer la variabilité spatiale de la contamination microbienne, des compléments doivent être apportés au modèle *PdIP*, tant sur les processus de mortalité des bactéries fécales que sur leurs composantes biologiques avant qu'il devienne un outil opérationnel. Une des conséquences majeures de l'imperfection de ce modèle réside dans le fait qu'il ne peut pas, avec une bonne qualité, reproduire tous les paramètres impliqués dans le processus d'une contamination microbienne, avec une combinaison unique de paramètres. La prise en compte de tous ces paramètres n'est pas possible, vu sa conception initiale. Une telle démarche serait contradictoire avec l'utilisation d'un nombre limité de paramètres.

Jusqu'à maintenant, rares sont les modèles de la qualité de l'eau qui ont été capables de reproduire parfaitement les variations de contamination microbienne des eaux récréatives avec un taux de précision de 80 %. Or, le calage et la validation du modèle *PdIP* a permis de sélectionner un jeu de paramètres permettant de simuler la distribution de la contamination microbienne avec des prédictions supérieures à 70 %. Ce modèle se révèle ainsi compétitif par rapport aux modèles plus élaborés en termes de capacité à simuler la variabilité spatiale de la

contamination microbienne des eaux récréatives.

Bien qu'encore limité, le modèle probabiliste *PdIP* a donné satisfaction sur d'autres bassins que ceux utilisés lors du calage. Le but n'était pas ici de reproduire parfaitement les concentrations de coliformes observées, mais de savoir cibler les zones potentiellement à risque. Dans ce cadre, notre modèle permet d'ores et déjà l'exploration et la spatialisation du phénomène de contamination des eaux récréatives. À ce titre, il pourrait être intégré à une démarche de surveillance des eaux récréatives pour identifier et cibler les secteurs à risque ou les zones contributives à la contamination microbienne.

Grâce à l'utilisation de la méthode probabiliste, les apports de cette étude sont doubles. Un premier apport d'ordre méthodologique a montré que l'approche bayésienne hiérarchique utilisée, est capable d'estimer la dynamique de la contamination microbienne. Il s'agit d'une démarche nouvelle. Les modèles de ce type constituent une approche connue en épidémiologie, mais cette approche a rarement été utilisée pour l'évaluation de la contamination microbienne des eaux récréatives. Elle offre la possibilité d'analyser le rôle de chaque paramètre considéré et les relations entre ces paramètres dans un contexte donné. Cette étude présente également un apport qualitatif permettant d'avoir une idée des probabilités que présente une zone à contaminer un cours d'eau.

Cette étude exploratoire ouvre la porte à de nombreuses perspectives. Pour la première fois, seuls, les aspects spatiaux sur les BV dont la limite est de 1 000 km² ont été considérés. Il serait intéressant d'utiliser les aspects temporels en lien avec les changements climatiques. Une autre perspective serait d'étendre la recherche à des régions plus importantes. Cet exercice déterminera si les mêmes facteurs interviennent dans l'explication de la contamination des eaux quelle que soit l'échelle. Dans cette perspective, la transposabilité sur d'autres BV du Québec s'avère une méthode efficace pour valider notre modèle et juger de sa performance du point de vue de son applicabilité pour la prévision de la contamination microbienne.

6.2. Analyse spatiale de la sensibilité du modèle *PdIP*

L'analyse de sensibilité est une information importante permettant de savoir quelles variables contrôlent le plus la réponse d'un modèle, ou celles à privilégier en phase de calibration. C'est dans ce contexte que la deuxième partie de nos travaux a consisté à explorer le potentiel de l'analyse de sensibilité globale basée sur la variance pour évaluer l'impact des incertitudes

associées à chaque variable explicative sur la réponse simulée par *PdLP*. Cet exercice a permis d'identifier les variables ayant le plus d'influence sur la variabilité de la contamination microbienne simulée.

Les indices de premier ordre de l'analyse de Sobol ont montré que parmi les variables explicatives les plus susceptibles d'affecter la contamination microbienne, **TA** est la première variable d'importance dans l'évaluation du taux de coliformes. C'est donc la variable sur laquelle l'attention devra portée pour améliorer sa description dans le contexte de prévision d'une contamination microbienne. Ensuite, la deuxième variable la plus importante est **ZU** avec une sensibilité d'environ 30 %. Par ailleurs, les estimations des indices totaux sont meilleures que celles des indices de premier ordre, ce qui signifie que l'impact des interactions paramétriques est significatif pour la modélisation de la contamination microbienne. Au niveau des indices de Sobol de premier ordre, on constate qu'ils sont tous inférieurs à 0,50. Cela signifie que la part de variabilité induite par chaque variable a un effet négligeable sur la sortie du modèle. Fixée à une valeur quelconque dans son domaine de variation, cette variable est sans conséquence sur la prédiction du modèle.

De plus, nous remarquons que, l'augmentation ou la diminution des valeurs de référence d'une variable aléatoire conduit à l'augmentation ou la diminution de son indice de Sobol, conduisant par conséquent à la diminution/augmentation de l'indice de Sobol des autres variables aléatoires. Ces résultats laissent croire que l'influence individuelle des variables explicatives sur le taux moyen des coliformes est relativement faible. Cependant, l'influence des interactions paramétriques à travers les indices de sensibilité totaux est relativement importante.

En comparant les résultats des indices de premier ordre et les indices de sensibilité totaux, cette étude montre que la sensibilité du modèle *PdLP* vis-à-vis de ces variables explicatives, fluctue et est assez relative. Elle peut subir des changements importants d'un jeu de données à un autre. Par exemple, dans le cas de la distance, il est difficile d'imaginer qu'elle ait une si petite influence sur la variabilité de la contamination microbienne en tenant compte de l'indice de Sobol de premier ordre. À ce niveau de l'analyse, nous aurions pu écarter cette variable, pourtant elle se retrouve en première position lorsqu'elle est en interaction avec d'autres variables environnementales. Cette faible sensibilité de premier ordre peut ainsi apporter des problèmes d'identifiabilité importants et donc d'une dépendance critique des combinaisons de variables identifiées en fonction du critère d'estimation et de la série de données (Francos *et al.*, 2003;

Kavetski *et al.*, 2006). Ce qui signifie que les paramètres ayant les meilleurs indices de premier ordre n'apportent pas forcément les meilleures contributions à l'ensemble du modèle. Toutefois, nous ne pouvons pas clairement identifier la raison pour laquelle les combinaisons basées sur les indices de premier ordre sont moins performantes que celles basées sur les indices totaux.

Nous suggérons que, les causes peuvent être des variables explicatives mal identifiées, des défauts structurels du modèle ou des processus importants qui n'auraient pas été pris en compte. Une autre explication serait que dans certains cas, une variable ne pourrait être utilisée qu'en de rares occasions (des processus qui se produisent après un certain seuil d'écoulement ou des événements extrêmes), et ne pas être utile dans la période de calibrage. Dans de tels cas, la variable n'a pas d'effet sur les résultats du modèle. Par conséquent, cette variable pourrait être considérée comme de moindre influence et pourrait se voir attribuer de faibles valeurs d'indice de sensibilité. Toutefois, si un événement est survenu et qui justifierait l'utilisation de la variable, l'analyse doit être entreprise avec la nouvelle période et cet événement extrême. Cela garantirait que le modèle ne se limite pas artificiellement par une variable d'une influence limitée.

Par ailleurs, au cours de cette étude, le choix de l'analyse de sensibilité par les indices de Sobol a été fait sans tenir compte des inconvénients connus de cette approche. En principe, la méthode de Sobol est basée sur la forte hypothèse que les variables ne sont pas corrélées. Pour plusieurs études (Obled, 1994; Kavetski, 2006), cette approche s'est avérée meilleure en donnant de bons résultats pour la hiérarchisation des variables explicatives dans des modèles hydrologiques. Or cette hypothèse de non-corrélation ne semble pas être vérifiée dans la plupart des modèles hydrologiques et de la qualité de l'eau.

Au cours des dernières années, quelques études ont abordé cette problématique (Pappenberger *et al.*, 2005; Chastaing *et al.*, 2014). Il ressort de ces études que la principale limite du calcul des indices de Sobol est qu'elles nécessitent au moins 7 à 8 paramètres pour identifier de façon raisonnable les variables qui influencent la sortie du modèle. Par conséquent, notre modèle *PdIP* constitué de cinq variables aurait été trop simpliste pour s'adapter globalement et parfaitement aux différents indices de Sobol et fournir des relations adéquates convenables à toutes les variables explicatives. Au niveau des indices totaux, les résultats apparaissent relativement cohérents en ce qui concerne l'influence de la dynamique interactive sur la sortie de *PdIP*. Néanmoins, il y a des raisons de croire qu'une bonne précision et de meilleurs résultats globaux pourraient être obtenus si cette méthode s'appliquait sur des modèles

de la qualité de l'eau plus complexes ayant un grand nombre de variables explicatives.

D'une autre part, nous sommes en accord avec les études de Favis-Mortlock et Smith (1990) qui mentionnent que les résultats d'une analyse de sensibilité doivent être manipulés avec précaution. En effet, la sensibilité d'un modèle à une variable donnée dépend de l'importance, durant la période choisie, du processus que cette variable affecte, elle-même liée aux contraintes environnementales et, dans une moindre mesure, aux conditions initiales. En d'autres termes, la formulation d'un processus, valable pour un type de situation en particulier, ne l'est pas forcément pour d'autres situations. Huijbregts *et al.* (1998) ont pu suggérer que dans ce cas, il est préférable de modifier la structure avec de nouvelles hypothèses, mais il faut alors recommencer toutes les démarches d'identification et de validation. Ces auteurs mentionnent qu'il est préférable dans ce cas d'ajouter d'autres variables et d'adapter le modèle à un jeu de données issues d'une situation particulière plutôt que d'identifier un jeu de variables plus ou moins valable pour une variété de situations.

Bien que cela soit possible, à quoi cela servira puisque de nombreux modèles agro-hydrologiques sont confrontés à un problème de sur-paramétrisation. L'intégration de plusieurs variables complique inutilement les modèles agro-hydrologiques. Cette sur-paramétrisation se répercute inévitablement sur l'emploi du modèle dans un cadre prévisionnel. En ce sens, Heuvelink *et al.* (2010a et b) suggèrent que l'exploration des interactions entre les variables en chaque situation particulière est préférable à une recherche aveugle d'un jeu de paramètres plus ou moins valable pour une variété de situations. Cela permet de vérifier la robustesse des modèles et de mieux évaluer la confiance dans leurs résultats. La connaissance de cette interaction permettrait également de connaître les variables du modèle qui expliquent la plus grande part de la variance du signal de sortie sur un support spatiale donné en fonction d'un changement particulier.

Grâce à l'utilisation de la méthode de Sobol pour l'analyse de sensibilité des variables spatiales, les apports de cette étude sont nombreux. Sous un angle méthodologique, la différence repose, non seulement sur l'étude des incertitudes liées aux variables d'entrée spatiales qui sont propagées dans le modèle, mais aussi sur l'influence de ces variables sur la variabilité totale de la contamination microbienne. Il s'agit d'une démarche nouvelle, évaluée par l'intermédiaire d'une analyse de sensibilité globale à travers la méthode de Sobol. Comme résultat important issu de ce travail de recherche, des histogrammes de distributions pour la densité du taux de coliformes

sont obtenus au niveau de chaque variable spatiale étudiée, ce qui est considéré comme étant une contribution originale. La méthode de Sobol s'est alors révélée être un outil très utile, en donnant de précieuses informations sur l'influence de chaque variable explicative sur la sortie du modèle *PdIP*. Elle a également permis de comprendre les différentes hiérarchies existantes pour les incertitudes étudiées en fonction de la morphologie du BV.

L'analyse de sensibilité constitue clairement un apport intéressant de la quantification d'incertitude puisqu'elle permet de mieux comprendre le rôle joué par une grandeur dans l'estimation du taux de coliformes. La méthodologie proposée peut alors être utilisée pour aider à comprendre comment les incertitudes vont affecter le comportement de la distribution spatiale du taux de coliformes et servir d'appui dans le futur pour améliorer la prédiction des risques de contamination microbienne. Il s'agit d'une méthode suffisamment simple qui pourra être utilisée dans d'autres modèles, y compris à des échelles plus grossières, utilisant des modèles de contamination microbienne tels que Hydrotel ou GIBSI. À notre connaissance, aucune démarche de ce type n'a encore été proposée.

Bien que les résultats obtenus montrent que l'analyse de sensibilité est une méthode assez fiable pour estimer la variabilité due aux variables spatiales, nous pouvons néanmoins relever quelques limites. Au cours de cette étude, la quantité des mesures et leur fréquence n'ont pas été prises en considération. Or dans ce contexte, plusieurs auteurs s'accordent sur le fait que si les mesures sont de mauvaise qualité, ou si l'hétérogénéité du milieu induit des mesures dont la variabilité est trop grande, le processus d'échantillonnage peut aboutir à un modèle incorrectement paramétré et par conséquent, des valeurs d'incertitudes assez élevées. Par ailleurs, dans le cas d'un échantillonnage préférentiel en été et dans certaines zones, les incertitudes peuvent aussi être grandes. Cela peut être une des conséquences d'un échantillonnage des plages mal réparti pour lequel plusieurs mois de l'année ne sont jamais échantillonnés, comme c'est le cas pour des données que nous avons utilisées. Les mesures sur lesquelles étaient basées nos calculs sont certes nombreuses, mais globalement mal réparties car la stratégie d'échantillonnage n'était pas régulière pour certains SBV (trois à dix fois par mois) et pour d'autres SBV les mesures sont quotidiennes s'échelonnant sur un mois. Alors, passer d'un calcul aux mesures journalières à un calcul mensuel irrégulier pourra également augmenter l'intervalle de confiance, dû à un biais des moyennes et des écarts-types. Dans ce cas, la question de la pertinence de la fréquence proposée se pose.

6.3. Variabilité temporelle de la contamination microbienne des eaux : une approche géomatique à l'échelle d'un BV agricole

La modélisation de la dynamique temporelle des coliformes dans les eaux de surface à partir du modèle AVSWAT a permis d'appréhender l'influence des caractéristiques de la pluie. Au regard de l'ensemble des résultats obtenus, il est possible d'admettre que nous avons abouti à une représentation satisfaisante de la dynamique de flux de coliformes sur le BV du lac Massawippi suivant les caractéristiques pluviométriques. Diverses études ont mis en évidence l'influence des caractéristiques des pluies sur la contamination fécales. Notre étude est la première à mettre en évidence des cartographies temporelles en fonction de la hauteur de la pluie, sa durée ainsi que son intensité.

L'analyse de sensibilité a permis d'élargir la connaissance des paramètres les plus influents sur les débits et les flux de coliformes simulés. Les résultats obtenus ont permis de mieux caractériser et d'améliorer l'identification des paramètres. Au niveau du calage, les performances optimales sont assez différentes sur les deux périodes et suivant le type de calage : mono-objectif ou multi-objectif. Une comparaison des valeurs du pourcentage de biais et du critère NSE révèle que le calage multi-objectif apparaît plus performant que le calage mono-objectif. Cependant, les jeux de paramètres donnant une bonne simulation à l'égard de ces deux critères peuvent fournir des résultats insignifiants en ce qui concerne le coefficient de détermination. La réciproque ne se vérifie pas puisque les jeux de paramètres donnant une bonne simulation à l'égard du coefficient de détermination ne semblent pas donner de mauvais résultats sur le NSE et le pourcentage de biais. En effet, on remarque que, plus, le NSE est élevé, plus l'adéquation entre les mesures simulées et mesurées est bonne, alors qu'une bonne adéquation du coefficient de détermination ne conduit pas forcément à une bonne complémentarité entre les valeurs simulées et mesurées. Au Niveau du calage, l'apport du coefficient de détermination est donc faible.

Au niveau de l'influence des caractéristiques de la pluie, il ressort que le flux de coliformes est fortement dépendant de l'intensité de la pluie. Dans une étude similaire, Moulinet *et al.* (2004) ont montré la parfaite corrélation entre la variation de ces deux paramètres de la pluie et celle de la contamination microbienne. En été, les pluies de fortes intensités sont

nécessaires pour déclencher une crue; l'augmentation du débit est alors très importante et très rapide conduisant au transport des coliformes. Les caractéristiques du BV du lac Massawippi expliquent ce fonctionnement hydrologique. C'est un bassin *a priori* propice au ruissellement : on note la présence de fortes pentes; l'occupation du sol est majoritairement agricole et les cultures sont situées le plus souvent en bordure du cours d'eau.

La relation entre la durée de la pluie et les flux de coliformes semble confuse dans le réseau hydrographique. L'analyse des cartes de distribution montre bien que les taux de coliformes décroissent en fonction de la durée de la pluie pour l'ensemble des bassins versants. Cette décroissance est plus nette sur le BV de la rivière Niger où on note des taux de coliformes de moins de 100 ufc/100ml. Cela peut s'expliquer par des transferts très rapides et en grande quantité de coliformes pendant une longue durée de pluie. Bien que, les relations entre la durée de la pluie soient significatives pour $t = 2$ h, ces corrélations ne demeurent pas significatives au-delà de cette durée. En effet, ces résultats sont logiques et rejoignent les conclusions de plusieurs auteurs (Debieche, 2001) qui ont démontré que si la pluie est de courte durée et intense, elle produit beaucoup plus de coliformes que si elle est de longue durée. Par ailleurs, pour qu'une pluie de longue durée produise la même concentration de coliformes, obtenue avec une pluie de courte durée, il faut que la hauteur de la pluie soit de plus en plus grande. Cela s'explique par le fait que, les pluies importantes de courte durée et qui se répartissent sur une grande partie du BV, produisent des débits importants capables de déclencher le transport des coliformes à travers le réseau hydrographique. Par contre les pluies qui se répartissent sur un intervalle de temps long, produisent un débit moins important à cause de l'infiltration des eaux vers la nappe, et de l'écoulement progressif des eaux; par conséquent, le transport des coliformes est amoindri.

Au niveau des intensités de la pluie, on peut constater qu'avant la crue ($i = 10$ mm/h), il y a une absence de relation entre le flux de coliformes et les caractéristiques de la pluie, puisque la variation des coliformes peut se produire dans des conditions très variées. Ce résultat est contraire aux observations de plusieurs auteurs (López-Tarazón *et al.*, 2010) qui montrent que le flux des bactéries est très lié à la durée des pluies et au débit avant la crue, en lien vraisemblablement avec l'état d'humidité des sols. Par ailleurs, nos résultats rejoignent ceux plus nuancés de Zabaleta *et al.* (2010) qui observent ou non cette influence de l'intensité de la pluie avant la crue. En effet, les variations des flux de coliformes en fonction de l'intensité de la pluie sont surtout marquées entre les mois de mai et juin. Après les crues de mars et avril dues à la

fonte des neiges, les pluies des mois de mai et juin (faibles intensités) exportent moins les coliformes, ce qui va de pair avec des débits et des vitesses de montée plus faibles. Les débits de base plus faibles durant ces mois limitent aussi le ruissellement des coliformes dans le cours d'eau entre les crues. Les crues qui mobilisent le plus de coliformes se produisent aux mois de juin et juillet (50 à 60 mm/h) rarement au mois d'août. En cette période, les fortes corrélations entre l'intensité de la pluie et les flux de coliformes s'expliquent par l'épandage du fumier, l'érosion de parcelles labourées, due à une série de pluies survenues à cette période à risque.

Chapitre 7. Conclusion

S'inscrivant dans le cadre général de la mise en place d'une modélisation spatio-temporelle de la contamination microbienne des eaux récréatives, cette thèse propose une méthode scientifiquement innovante permettant de répondre à une question essentiellement multidisciplinaire combinant les notions de géomatique, de microbiologie et d'hydrologie. Cette étude exploratoire a permis d'étudier un phénomène environnemental complexe : la dynamique des coliformes fécaux dans les eaux, en cartographiant leur variabilité de manière spatialement et temporellement explicite. La démarche proposée pour ce travail de thèse s'est articulée autour de trois axes de recherche principaux.

Le premier axe a présenté la modélisation probabiliste de la contamination microbienne des eaux récréatives dans le but de hiérarchiser et cartographier les zones contributives aux risques d'atteinte à la qualité microbiologique des eaux. Pour ce faire, une approche bivariée basée sur le théorème de Bayes (théorie de l'évidence) combiné aux systèmes d'information géographique a été choisi. Nous avons donc évalué les stratégies de paramétrisation et validation de ce modèle pour sa transposition sur d'autres bassins versants. Deux échelles spatiales caractérisées par une variabilité des états de surface et une différence morphologique ont été utilisées. Cette approche est nouvelle et différente de celles classiquement employées dans le domaine d'estimation de la vulnérabilité des eaux de surface. À notre connaissance, aucune étude n'avait jusqu'à présent abordé de manière probabiliste l'effet des processus spatiaux qui contrôlent la dynamique de la contamination des eaux à l'échelle d'un BV à vocation agricole. En la considérant comme une méthode opérationnelle de gestion de la qualité de l'eau, nous avons évalué la possibilité d'une transposition vers d'autres sites aux caractéristiques environnementales semblables. La démarche mise en œuvre a permis d'aller au-delà d'une simple cartographie de la vulnérabilité des eaux. Elle a mis en exergue la possibilité de quantifier et de hiérarchiser la structure profonde des mécanismes de propagation des coliformes dans les eaux, c'est-à-dire le rôle de chacune des variables prédictives et les relations entre ces paramètres dans un contexte donné. Les résultats obtenus ont montré l'efficacité du *PdIP* à estimer la distribution *a posteriori* des zones de contamination les plus probables à la contamination microbienne.

L'étude du modèle a toutefois aussi montré l'existence de sous-estimations des flux de coliformes. Cela provenait majoritairement d'erreurs de paramètres calés trop faibles et moins

représentatifs de certains SBV. Malgré tout, le calage et la validation du modèle ont permis de sélectionner un jeu de paramètres permettant de simuler au mieux la distribution de la contamination microbienne. Il manquerait en fait quelques variables pour augmenter le taux de prédiction des simulations. Par exemple, s'il était possible de quantifier les apports d'eau provenant des réseaux de drainage artificiel en milieu agricole, il est clair que les estimations de contamination seraient grandement améliorées.

Au cours de ces travaux, le modèle *PdlP* a été développé pour étudier la probabilité des eaux récréatives à être contaminées par *E. coli*. Les résultats obtenus ne suffisent pas pour dépister des sources de pollution microbienne (animale ou humaine). Une piste de développement serait de déterminer la provenance de ces bactéries fécales afin de mieux déterminer la source de contamination microbienne, et assurer une protection efficace des bassins versants et de la santé publique.

Par ailleurs, il se pose une question plus fondamentale sur l'intérêt de la prise en compte des propriétés d'adsorption des sols pour l'estimation de la contamination des eaux de surface. Dans le modèle proposé, la propagation des bactéries dépend majoritairement des variables spatiales. Or, ces dernières ne sont sûrement pas les seules à avoir une influence sur le transport et la mortalité des bactéries. Les phénomènes de compétition et d'adsorption n'ont pas été abordés au cours de nos travaux, pourtant ils jouent un rôle important dans la survie des bactéries. Une perspective intéressante serait d'évaluer la contamination microbienne des eaux en fonction des bactéries adsorbées sur les particules de sol, et entraînées par érosion. Cela consisterait à traduire la compétition et l'adsorption sous forme de variables qui interviendraient dans la contamination des eaux au même titre qu'une variable spatiale.

Le deuxième axe a consisté en une analyse de sensibilité de *PdlP* pour mieux faire ressortir l'apport de chacune des variables spatiales utilisées comme intrant au modèle prédictif. Il s'agissait de quantifier le domaine de variabilité des sorties du modèle *PdlP*, en déterminant la part de variance due à l'occupation du sol et aux caractéristiques pédologiques. L'originalité de ce travail demeure dans la manière de quantifier et de propager les incertitudes des variables ainsi que dans la méthodologie d'analyse de sensibilité mise en place. De manière générale, dans le cadre des modèles de la qualité de l'eau, les analyses de sensibilité utilisées sont de type local. Ces dernières, ne permettent pas de tenir compte des interactions entre les paramètres et les effets non-linéaires. Pour notre étude, nous avons choisi les indices de Sobol, une méthode globale

d'analyse de sensibilité basée sur la décomposition de la variance avec un plan d'expérience un-à-la-fois. Ces indices permettent de bien explorer les valeurs des paramètres et de tenir compte des interactions et des effets non-linéaires. Les incertitudes sont représentées par le biais d'une distribution de probabilité dont les paramètres s'ajustent en fonction de l'information disponible. Il a été possible de prédire précisément les incertitudes des différentes variables en tout point du terrain. Cet exercice n'était pas une évaluation exhaustive des toutes les incertitudes du *PdlP*, mais tout simplement une identification de différentes sources d'incertitude sur la distribution des taux de coliformes. Les conclusions pratiques concernent l'identification des paramètres d'entrée, à savoir où concentrer les efforts de mesure, dans le cas d'un domaine présentant des caractéristiques pédologiques et d'occupation de sol, différentes afin de réduire les incertitudes des variables de sortie. En effet, nous avons montré que, les incertitudes sont présentes aussi bien pour les sols argileux que les sols sableux, mais avec une amplitude variable. Les incertitudes dues à la variation des sols sableux n'ont pas une influence importante sur les taux de coliformes, et l'incertitude sur la réponse du modèle est moins ressentie. Par ailleurs, dès lors qu'une source de variabilité des teneurs en argile, même peu importante pour le résultat final, est négligée, il y a une surestimation au niveau des quartiles, ce qui peut entraîner une grande erreur au niveau de la prévision de la contamination microbienne. La modélisation se doit donc de prendre en compte toutes les sources de variabilité des sols argileux, même celles qui semblent avoir un impact mineur sur la distribution des coliformes. Un second cas a montré que l'incertitude de la distance est très importante pour obtenir des résultats fiables du taux de coliformes. Cependant, l'incertitude due à la distance est essentiellement influencée par la variabilité totale, autrement dit par la combinaison d'autres variables telles que l'occupation du sol, la pente et les caractéristiques pédologiques. Les résultats obtenus sont encourageants et permettent de démontrer le potentiel des indices de Sobol pour l'analyse de sensibilité des variables et l'incertitude associée à ces derniers dans l'estimation des flux de coliformes par l'approche *PdlP*. Dans un premier temps, la richesse de l'information obtenue par les indices de Sobol offre un nouvel angle de vue permettant de mieux appréhender la relation entre les variables spatiales et les flux de coliformes et d'en évaluer les incertitudes. Cette étude a fourni une analyse complète de l'incertitude globale induite par les variables d'entrée ainsi qu'une mesure de la sensibilité des variables d'intérêt aux différentes variables. De cette façon, les indices de Sobol ont répondu aux objectifs préalablement établis de notre étude. Cette approche a ainsi permis de mesurer la

variabilité spatiale du champ de concentration des coliformes aux variables d'entrée spatiales en caractérisant les gammes de variations des facteurs sur lesquelles il faut agir pour que le flux de coliformes prédit par *PdlP* soit dans une gamme bien précise. Même si quelques différences peuvent être observées entre les indices de sensibilité (de premier ordre et totaux), ils fournissent des informations complémentaires, ce qui est un facteur de qualité pour le modèle. L'analyse de sensibilité réalisée dans cette étude a permis d'identifier les variables spatiales les plus sensibles à la contamination microbienne. Une étude d'analyse de sensibilité du modèle *PdlP* à des configurations différentes de variations temporelles pourrait améliorer davantage notre compréhension du transport des bactéries dans les eaux.

Le troisième axe de ce travail de thèse est de parvenir à une meilleure compréhension des dynamiques temporelles des coliformes vers les eaux de surface. La stratégie d'étude a consisté à analyser, par le biais du modèle agro-hydrologique AVWSAT, l'impact de la pluviométrie, sur les évolutions des taux de coliformes. Le modèle a été calé sur le BV de la Massawippi en utilisant des approches mono-objectif et multi-objectif. La performance du modèle a été meilleure après une optimisation multi-objective: les coefficients de Nash-Sutcliffe étaient plus élevés et les biais plus faibles. Le paramétrage du modèle a permis d'atteindre de bonnes simulations des débits et des coliformes fécaux sur de longues périodes hydrologiques. Une analyse paramétrique a montré que les taux de croissance et de mortalité des bactéries adsorbées jouent un rôle déterminant sur le flux de coliformes acheminés vers l'exutoire du BV. L'analyse a aussi montré que cet aspect fondamental est principalement contrôlé par les paramètres de ruissellement. Parallèlement, la profondeur du sol n'a démontré qu'une influence modérée sur la partition entre infiltration et ruissellement et donc sur le transfert des coliformes dans les eaux.

En termes de dynamique temporelle, des résultats encourageants fournissent un premier aperçu des relations entre la pluviométrie et le déplacement des coliformes à travers un réseau hydrographique. L'analyse des coefficients de corrélation a indiqué qu'il existe une certaine relation entre l'intensité des pluies et l'augmentation des taux de coliformes. En effet, plus l'intensité des pluies est faible, plus le transport des bactéries se produit dans une échelle limitée, et plus la charge polluante est moins significative. Ainsi, les intensités des pluies de 10 mm/h se sont avérées moins importantes pour déclencher le ruissellement des coliformes dans les réseaux hydrographiques. D'ailleurs, pour une intensité de 10 mm/h il y a une absence de relation entre le flux de coliformes et les caractéristiques de la crue, puisque la variation des coliformes peut se

produire dans des conditions très variées. L'analyse a démontré que ce n'est qu'à partir d'une intensité de 50 mm/h que se trouve la limite inférieure des pluies pouvant occasionner une contamination microbienne. Cette relation n'est pas nécessairement applicable d'emblée pour tous les BV. Elle varie d'un BV à un autre en fonction des conditions naturelles, des propriétés physiques et chimiques du sol, et de la surface totale du bassin.

Bien que le modèle offre un certain réalisme par rapport aux données mesurées, l'approche méthodologique adoptée présente des limites d'ordre conceptuel et de validation qu'il est important de souligner. En premier lieu, nous rappelons que la présente recherche souffre d'un manque de mesures des concentrations de coliformes fécaux en termes de fréquence et de répartition spatiale. La prise de mesures de la qualité de l'eau est répartie de façon inégale entre les différents BV. Cette inégalité a rendu notre tâche difficile pour valider le modèle. Une bonne validation nécessiterait plus de mesures quotidiennes bien étalées sur de longues durées. Il faudrait que les périodes de monitoring des plages soient régulières afin d'éviter les risques de mauvaise interprétation des résultats, et d'aboutir à des résultats exploitables et pertinents. Il convient donc de considérer les résultats obtenus comme des indicateurs de risque de transfert et non comme des grandeurs réelles.

Les perspectives de l'ensemble de notre étude sont diverses. Bien que cette thèse apporte des éléments de réponse originaux concernant la dynamique spatio-temporelle des coliformes, il reste de nombreuses questions à résoudre. Tout d'abord, il s'agit des aspects relatifs à l'utilisation de la méthode de Sobol dans la modélisation des risques de contamination. Ce travail peut être poursuivi dans différentes directions. Tout d'abord, au lieu de se limiter à modéliser l'effet des variables spatiales sur la contamination microbienne à l'aide des lois normale et uniforme, il pourrait être intéressant d'envisager d'autres types de distributions (Gamma, gaussienne, triangulaire) ou bien non paramétriques en se basant sur l'estimation des déciles de la distribution. D'autre part, l'approche proposée devrait être en mesure de considérer d'autres variables : la localisation des fosses, la nature de la culture, les dates d'épandage de fumier et lisier. De tels paramètres ont aussi un impact majeur sur la réponse de la contamination microbienne et devraient être pris en compte. Un point intéressant serait de déterminer leurs interactions avec les autres paramètres du modèle qui ne sont pas nécessairement évidentes. À long terme, nous prévoyons de calculer les probabilités de contamination microbienne en conditions prospectives, c'est-à-dire de réaliser des simulations avec le changement de couverture

du sol. De ce fait, les cartes de probabilités de contamination microbienne développées dans la première partie de la thèse peuvent être exploitées et combinées avec un certain nombre de situations représentatives, et définir une cartographie des incertitudes dues aux variables spatiales.

Du point de vue événementiel, il semble nécessaire dans un premier temps d'améliorer et d'affiner l'évolution temporelle des sensibilités des paramètres, pour juger de la vraisemblance de la paramétrisation des débits et des flux de coliformes. Les différences importantes qui ont été mises en évidence entre les combinaisons des variables lors des simulations de débits ou des flux de coliformes mettent en avant l'étude des sensibilités paramétriques temporelles des paramètres. En fait, les comparaisons qui ont pu être réalisées entre les simulations ont montré que pour une combinaison identique de paramètres, les résultats de simulations obtenus pouvaient être significativement différents. La sensibilité spatio-temporelle des paramètres restent donc un enjeu important dans la simulation du flux de coliformes. Une meilleure compréhension de cet aspect permettra de prendre en compte leur unicité et de les adapter aux procédures de régionalisation. Une autre perspective de recherche consisterait à analyser la quantification systématique des incertitudes sur ces données. Cela permettrait de les propager sur les prévisions de débit, étape indispensable pour apprécier justement les résultats d'un modèle. Cette confrontation permettra de lier le niveau de complexité du modèle avec la quantité de données nécessaire pour la mise en œuvre d'une meilleure modélisation de la dynamique des coliformes dans les eaux de surface.

Ce travail de thèse nous a ainsi permis de caractériser la contribution significative de la dynamique spatio-temporelle de la contamination microbienne des eaux, dans un objectif prévisionnel. Bien qu'il reste encore quelques points à approfondir, les premiers résultats sont encourageants.

Références bibliographiques

- Agence de la santé et des services sociaux de Chaudière-Appalaches. (2006): Profil de Santé environnementale Chaudière-Appalaches 1998-2005. 180 p.
- Agriculture et Agroalimentaire Canada. (2003): Statistiques sur l'agriculture biologique, 2003, Québec. <http://www.agr.gc.ca/misb/hort/org-bio/pdf/OrganicsQc2003-f.pdf>.
- Ahl, R. S.; Woods, S. W.; Zuuring, H. R. (2008): Hydrologic calibration and validation of SWAT in a snow-dominated rocky mountain watershed, Montana, USA. *Journal of the American Water Resources Association*, vol. 44, n° 6, pp. 1411-1430.
- Aitken, M. N. (2003): Impact of agricultural practices and river catchment characteristics on river and bathing water quality. *Water Science and Technology*, vol 48, n° 10, pp. 217-224.
- Alexander, D. D.; Weed, D. L.; Mink, P. J.; Mitchell, M. E. (2012): A weight-of-evidence review of colorectal cancer in pesticide applicators: the agricultural health study and other epidemiologic studies. *International Archives of Occupational and Environmental Health*, vol. 85, pp. 715-745.
- Andréassian, V. N.; Le Moine, C.; Perrin, M. H.; Ramos, L.; Oudin, T.; Mathevet, J.; Lerat, L.; Berthet. (2012): All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes*, vol. 26, n° 14, pp. 2206-2210.
- Arnold, J. G.; Allen, P. M. (1999): Automated methods for estimating baseflow and ground water recharge from streamflow records. *Journal of the American Water Resources Association*, vol. 35, n° 2, pp. 411-424.
- Arnold, J. G.; Fohrer, N. (2005): SWAT2000: Current capabilities and research opportunities in applied watershed modeling. *Hydrological Processes*, vol. 19, n° 3, pp. 563-572.
- Arnold, J. G.; Moriasi, D. N.; Gassman, P. W.; Abbaspour, K. C.; White, M. J.; Srinivasan, R.; Santhi, C.; Harmel, R. D.; Van Griensven, A.; Van Liew, M.; Kannan, W. N.; Jha, M.K. (2012): SWAT: Model Use, Calibration, and Validation. *Transactions of the ASABE*, vol. 55, n° 4, pp. 1491-1508.
- Arsenault, R. (2015): Équifinalité, incertitude et procédures multi-modèle en prévision hydrologique aux sites non-jauges. Thèse de doctorat, École de Technologie Supérieure, Montréal, 405 p.
- Ascough, J- C.; Rector, H- D.; Hoag, D-L.; McMaster, G- S.; Vandenberg, B- C.; Shaffer, M- J.; Weltz, M.-A.; Ahjua, L.-R. (1999): Multicriteria spatial decision support systems: overview, applications, and future research directions, in GIS and Multicriteria Decision Analysis. Eds J. Malczewski, John Wiley and Sons, New York, 392p.
- Aspinall, R.; Pearson, D. (2000): Integrated geographical assessment of environmental condition in water catchments: Linking landscape ecology, environmental modelling and GIS. *Journal of Environmental Management*, vol. 59, n° 4, pp. 299-319.
- Association pour la protection de l'environnement du lac St-Charles et des Marais du Nord (APEL) (2012) : Suivi des rivières du bassin versant de la rivière Saint-Charles, Campagne 2011, Association pour la protection de l'environnement du lac Saint-Charles et des Marais du Nord, Québec, 133 pages.

- Aubertot, J- N.; Barbier, J- M.; Carpentier, A.; Gril, J. J.; Guichard, L.; Lucas, P.; Savary, S.; Savini, I.; Voltz, M. (2005): Pesticides, agriculture et environnement. Réduire l'utilisation des pesticides et en limiter les impacts environnementaux. Expertise scientifique collective INRA, Cemagref, Paris, 902 p.
- Auzet, (1990): L'érosion des sols par l'eau dans les régions de grande culture : aspects aménagements. Centre d'études et recherches éco-géographiques, ministère de l'Environnement / ministère de l'Agriculture, 39 p.
- Baecher, G. B.; Lanney, N. A.; Einstein, H. H. (1977): Statistical description of rock properties and sampling, Proc.18th U. S. Symp. on Rock Mech., AIME, 5CI-I à 5CI-B.
- Baffaut, C.; Benson, V. W. (2003): A bacteria TMDL for Shoal Creek using SWAT modeling and DNA source tracking. In Proc Total Maximum Daily Load (TMDL) Environmental Regulations II. Ali Saleh, ed. ASAE Paper No. 701P1503. St Joseph, Mich.: ASAE.
- Baillargeon, S: 2005. Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitation. Mémoire de D.E.A., Université de Laval.
- Balestrat, M. (2006) : Usage de la télédétection et du SIG pour construire une méthode d'évaluation du risque potentiel de pollution diffuse des eaux de surface par les produits phytosanitaires. Mémoire de Master 2 professionnel en géomatique, Toulouse 2 le Mirail, 76 p.
- Bardsley, W. E.; Major, T. J.; Selby, M. J. (1990): Note on a Weibull property for joint spacing analysis, *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 27, n° 2, pp. 133-134.
- Baril, P. (2002): Impacts de la production porcine sur le milieu naturel, présentation à la Commission sur le développement durable de l'industrie porcine. Ministère de l'environnement du Québec. Document numéro BIO7.
- Beauduin, N.; Parnaudeau, V.; Mary, B.; Makowski, D.; Meynard, J- M. (2004): Simulation de l'impact de différents scénarios agronomiques sur les pertes de nitrate a l'échelle d'un bassin hydrologique, in *Organisation spatiale des activités agricoles et processus environnementaux*. Eds P. Monestiez, S. Lardon and B. Seguin, INRA, Paris, pp. 117-141.
- Berville, D.; Marcary, F. (2006) : Synthèse bibliographique: bilan des connaissances sur l'érosion et les phénomènes de ruissellement. Université Montaigne Bordeaux III. p28.
- Beven, K. J.; Binley, A. M. (1992): The future of distributed models: model calibration and predictive uncertainty. *Hydrology Processes*, vol. 6, pp. 279 -298.
- Bioteau, T.; Bordenave, P.; Laurent, F.; Ruelland, D. (2002): Évaluation des risques de pollution diffuse par l'azote d'origine agricole à l'échelle de bassins versants : intérêts d'une approche par modélisation avec SWAT. *Ingénieries-EAT*, vol. 32, pp. 3-12.
- Bodin, J.; Razack, M. (1999): L'analyse d'images appliquée au traitement automatique de champs de fractures. Propriétés géométriques et lois d'échelle. *Bulletin de la Société géologique de France*, vol. 170, n° 4, pp. 579-593.
- Boithias, L.; Sauvage, S.; Taghavi, L.; Merlina, G.; Probst, J-L.; Sánchez Pérez, J. M. (2011): Occurrence of metolachlor and trifluralin losses in the Save river agricultural catchment during floods. *Journal of Hazardous Materials*, vol. 196, pp. 210-219.

- Bolduc, D.G. (1998): Bilan des éclosions de maladies d'origine hydrique signalées dans les directions régionales de la santé publique du Québec en 1993, 1994 et 1995. Comité de santé environnementale du Québec, Conseil des directeurs de santé publique du Québec, 18 p.
- Bombardier, D. (2011). Une gorgée d'eau de trop. La Tribune. <http://www.lapresse.ca/latribune/sherbrooke/201103/26/01-4383377-une-gorgee-deau-de-trop.php> (Page consultée le 13 octobre 2015).
- Bonham, G. F.; Agterberg, F. P.; Wright, D. F. (1989): Weights of evidence modelling: a new approach to mapping mineral potential. Statistical applications in the Earth Sciences. *Geological Survey of Canada*; Paper 89-9, pp. 171-183.
- Bougeard, M.; Le Saux, J.C.; Gnouma, R.; Dupont, S.; Pommepuy, M. (2008) : Modélisation des flux de contamination fécale et de leur impact sur la zone littorale (conséquences sur la qualité des eaux conchylicoles). Partie 1. IFREMER, Plouzané, France.
- Bousquet, F.; Gautier, D. (1999): Comparaison de deux approches de modélisation des dynamiques spatiales par simulation multi-agents : les approches "spatiale" et "acteurs". *Cybergéo*, n°89: 13 pages. URL : <http://www.cybergegeo.presse.fr/>.
- Boyacioglu, H. (2008): Water Pollution Source Assessment by Multivariate Statistical Methods in the Tahtali Basin, Turkey. *Environmental Geology*, vol, 54, pp. 275-282.
- Boyle, D. P.; Gupta, H. V.; Sorooshian, S. (2000): Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, vol. 36, n° 12, pp. 3663-3674.
- Brown, L. C.; Barnwell, T. O. (1987): The Enhanced Stream Quality Models QUAL2E and QUAL2E-UNCAS: Documentation and User's Manual. EPA/600/3-87/007.
- Bureau d'Audiences Publiques sur l'Environnement (BAPE). (2003): L'inscription de la production porcine dans le développement durable. Consultation publique sur le développement durable de la production porcine au Québec. Rapport principal, Rapport n°179, Québec.
- Burkart, M. R.; Kolpin, D.W.; Jaquis, R. J.; Cole, K. J. (1999): Agrichemicals in ground water of the Midwestern USA: relations to soil characteristics. *Journal of Environmental Quality*, vol. 28, pp. 1908-1915.
- Camesano, T. A.; Logan, B.E. (1998): Influence of fluid velocity and cell concentration on the transport of motile and nonmotile bacteria in porous media. *Environmental Science and Technology*, vol. 32, 1699-1708.
- Campbell, E.; Fox, P. (1999): A bayesian approach to parameter estimation and pooling in nonlinear flood event models. *Water Resources Research.*, vol. 35, n°1, pp. 211-220.
- Campolongo, F.; Cariboni, J.; Saltelli, A. (2007): An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, vol. 22, n°10, pp. 1509-1518.
- Carlos, G-D.; Gozálvez-Zafrilla, J. (2012): Uncertainty and sensitive analysis of environmental model for risk assessments: An industrial case study. *Reliability Engineering & System Safety*, vol. 107, pp.16-22.

- Centre de développement du porc du Québec. (2005): Écho Porcs. Hebdomadaire d'information économique pour la filière porcine québécoise. <http://www.cdpqinc.qc.ca/document/2005%20%C3%89cho-Porc%20lexique.pdf>
- Chagnon, M.; Bolduc, D. G. (2000): Bilan des éclosions de maladies d'origine hydrique signalées dans les directions régionales de la santé publique du Québec en 1996 et 1997. Unité risques biologiques, environnementaux et occupationnels, Institut national de santé publique du Québec, 22 p.
- Chagnon, M.; Bolduc, D. G. (2001): Bilan des éclosions de maladies d'origine hydrique signalées dans les directions régionales de la santé publique du Québec en 1998 et 1999. Unité risques biologiques, environnementaux et occupationnels, Institut national de santé publique du Québec, 23 p.
- Champagne, L.; Chapui, R.P. (1993): Évaluation et cartographie de la vulnérabilité des formations aquifères de la MRC de Montcalm selon la méthode DRASTIC. *Sciences et techniques de l'eau*, vol. 23, n° 3, pp.169-176.
- Chandler, D. S.; Craven, J. A. (1980): Relationship of soil moisture to survival of Escherichia coli and Salmonella typhimurium in soils. *Australian Journal of Agricultural Research*, vol. 31, pp. 547-555.
- Chorowicz, J.; Ichoku, C.; Riazanoff, S.; Kim, Y-K.; Cervelle, B. (1992): A combined algorithm for automated drainage network extraction. *Water Resources Research*, vol. 28, pp.1293-1302.
- Christiaens, K.; Feyen, J. (2002): Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model. *Water Resources Research*, vol. 38, n° 9, pp.81-85.
- Cibin, R.; Sudheer, K. P.; Chaubey, I. (2010): Sensitivity and identifiability of stream flow generation parameters of the SWAT model. *Hydrological Processes*, vol. 24, pp.1133-1148.
- Coalition eau secours. (2003): « Avant que nous nous enlisions : pour un élevage sans danger pour l'eau ». Mémoire présenté dans le cadre de la Commission du BAPE sur la production porcine, 15 p.
- COGEBY, (2010) : Plan directeur. Portrait du bassin versant de Yamaska, 2007. Mise à jour 2010. 38 pages.
- COGESAF. (2006): Analyse du bassin versant de la rivière Saint-François. Sherbrooke, 255 p.
- Collins, R.; Rutherford, K. (2004): Modelling bacterial water quality in streams draining pastoral land. *Water Research*, 38: 700-712.
- Corgne. (2004): Hiérarchisation des facteurs de changements de l'occupation hivernale des sols Application au bassin versant du Yar (Bretagne). *Norois*, vol, 193, pp.17-29.
- Coyne, M. S.; Cornelius, P. L. (1996): Effect of sediment particle size and temperature on fecal bacteria mortality rates and the fecal coliform/fecal streptococci ratio. *Journal Environmental Quality*, vol. 25, pp. 1216-1220.
- Crane, S. R.; Moore, J. A.; Grismer, M. E.; Miner, J. R. (1983): Bacterial pollution from agricultural sources: A review. *Transactions of the ASAE*, vol. 26, pp. 858-866.

- Crosetto, M.; Tarantola, S.; Saltelli, A. (2000): Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems & Environment*, vol. 81, pp. 71-79.
- Crowther, J.; Kay, D.; Wye, M. D. (2001): Relationships between microbial water quality and environmental conditions in coastal recreational waters: the fylde coast, UK. *Water Research*, vol. 35, n° 17, pp. 4029-4038.
- Debieche, T. H.; Mania, J.; Mudry, J. (2001): Impact des rejets industriels et urbains sur la qualité des eaux superficielles (Fer, Manganèse et Ammonium), application à la basse vallée de la Seybouse (Nord-Est algérien), Colloque National d'Hydrogéologie et d'Environnement, (présentation orale), Fès – Maroc, p. 102.
- Depraetere, C.; Moniod, F. (1991): Contribution des modèles numériques de terrain à la simulation des écoulements dans un réseau hydrographique : exemple du bassin de Bras-David (Guadeloupe). *Hydrologie Continentale*, vol. 6, n° 1, pp. 29-53.
- Di Luzio, M.; Arnold, J. G. (2004): Formulation of a hybrid calibration approach for a physically based distributed model with NEXRAD data input. *Journal of Hydrology*. vol. 298, n° 1-4, pp. 136-154.
- Di Luzio, M.; Arnold, J. G.; Srinivasan, R. (2005): Effect of GIS data quality on small watershed streamflow and sediment simulations. *Hydrological Processes*, vol. 19, n° 3, pp. 629-650.
- Dörfliger, N.; Jauffert, D.; Loubier, S.; Petit, V. (2004): Cartographie de la vulnérabilité des aquifères karstiques en Franche-Comté. Rapport final BRGM/RP-53576-FR.
- Durand, P.; Tortrat, F.; Viaud, V.; Saadi, Z. (2006): Modélisation de l'effet des pratiques agricoles et de l'aménagement du paysage sur les flux d'eau et de matière dans les bassins versants. In: INRA (Editor), Qualité de l'eau en milieu rural - Savoirs et pratiques dans les bassins versants. *Update Sciences & technologies*, pp. 193-209.
- EPA, (1996): Better Assessment Science Integrating Point and Nonpoint Sources, EPA-823-R-96-001.
- Etzelmüller, B. (2000): Quantification of thermo-erosion in pro-glacial areas-examples from Spitsbergen. *Geomorphology*, vol. 44, n° 3, pp. 343-361.
- Fan, X. Y.; Cui, B. S.; Zhao, H.; Zhang, Z. M.; Zhang, H. G. (2010): Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. *Procedia Environmental Sciences*, vol. 2, pp. 1220-1234.
- Favis-Mortlock, D. T.; Smith, F. R. (1990): A sensitivity analysis of EPIC. In Sharpley, Williams (Eds). EPIC Erosion/Productivity Impact Calculator, 1. Model Documentation. U.S. Department of Agriculture. *Agriculture Technical Bulletin*, vol. 1768, pp. 178-190.
- Fenlon, D. R.; Ogden, I. D.; Vinten, A.; Svoboda, I. (2000): The fate of Escherichia coli and E. coli O157 in cattle slurry after application to land. *Society of Applied Microbiology*, vol. 88, pp. 149-156.
- Fishman, G. S.; Rubin, D. S. (1992a): Bounding the variance in Monte Carlo experiments. *Operations Research Letters*, vol. 11, pp. 243-248.

- Fishman, G. S.; Rubin, D. S. (1992b): Evaluating best-case and worst-case coefficients of variation when bounds are available. *Probability in the Engineering and Informational Sciences*, vol. 6, pp. 309-322.
- FitzHugh, T. W.; Mackay, D. S. (2000): Effects of parameter spatial aggregation on an agricultural nonpoint source pollution model. *Journal of Hydrology*, vol. 236, n° 1-2, pp. 35-53.
- Francos, A.; Elorza, F. J.; Bouraoui.; Bidoglio, F. G.; Galbiati, L. (2003): Sensitivity analysis of distributed environmental simulation models : understanding the model behaviour in hydrological studies at the catchment scale. *Reliability Engineering and System Safety*, vol. 79, n° 2, pp. 205-218.
- Frey, H. C.; Burmaster, D. E. (1999): Method for characterization of variability and uncertainty : comparison of bootstrap simulation and likelihood-based approaches. *Risk Analysis*, vol. 19, pp. 109-129.
- Freyberg, D. L.; Reeder, J. W.; Franzini, J. B.; Remson, I. (1980): Application of the Green-Ampt Model to Infiltration under Time-Dependent Surface Water Depths. *Water Resources Research*, vol. 16, pp. 517-528.
- Gabet, E. J.; Reichman, O. J.; Seabloom, E. (2003): The effects of bioturbation on soil processes and sediment transport. *Annual Review of Earth and Planetary Sciences*, vol. 31, pp. 259-273.
- Gagliardi, J. V.; Karns, J. S. (2000): Leaching of Escherichia coli in diverse soils under various agricultural management practices. *Applied and Environmental Microbiology*, vol. 66, pp. 877- 883.
- Gagliardi, J. V.; Karns, J. S. (2002): Persistence of Escherichia coli O157:H7 in Soil and on Plant Roots. *Environmental Microbiology*, vol. 4, n° 2, pp. 89-96.
- Gangbazo, G. (2000): Relations empiriques entre les utilisations du territoire et la qualité de l'eau des rivières. *Vecteur Environnement*, vol. 33, pp. 42-49.
- Gangbazo, G.; Barnett, G. M. (1997): Effets de l'épandage des engrais minéraux et de grandes quantités de lisier de porc sur l'eau, le sol et les cultures. Ministère de l'Environnement et de la Faune, Direction des écosystèmes aquatiques, 46 p.
- Gangbazo, G.; Bernard, C.; Côté, D. (1996): Effets de l'épandage du lisier de porc sur les eaux de ruissellement et de drainage. *Agrosol*, vol, 9, n°1, pp. 46-51.
- Garcia-Armisen, T.; Touron, A.; Petit, F.; Sevais, P. (2005): Sources of microbiological contamination in the Seine estuary (France). *Estuaries*, vol, 28, vol. 627-633.
- Garcia-Lara, J.; Menon, P.; Servais, P.; Billen, G. (1991): Mortality of fecal bacteria in seawater. *Applied and Environmental Microbiology*, vol. 57, pp. 885-888.
- Gardi, C. (2001): Land use, agronomic management and water quality in a small Northern Italian watershed. *Agriculture, Ecosystems and Environment*, vol. 87, pp. 1-12.
- Gaume, E.; Villeneuve, J- P.; Desbordes, M. (1998): Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. *Journal of Hydrology*, vol. 210, n°4, pp. 38-50.

- Gélinas et Direction générale des eaux (1977) : Caractéristiques physiques : Bassin versant de la Saint-François. Québec, Ministère des richesses naturelles.
- Gemperli, A.; Vounatsou, P.; Sogoba, N.; Smith, N. (2005) : Malaria Mapping Using Transmission Models: Application to Survey Data from Mali. *American Journal of Epidemiology*, vol, 163, n°3, p. 289-297.
- Ghattas. (2008) : Sélection de variables pour la classification binaire en grande dimension : Comparaisons et application aux données de biopuces. *Journal de la Société Française de Statistique*, vol. 145, n°3. pp. 104-109.
- Gillespie, P. A.; Howard, C. B.; Walsh, J. J.; Watterson, J. (1993): Measurement and characterization of spatial distributions of fractures. *Tectophys*, vol. 226, pp. 113-1141.
- Gingras, B. (1998): L'agriculture, l'eau, l'air et la santé. Agriculture et environnement. Direction de la santé publique Chaudière-Appalaches et Comité de santé environnementale du Québec. 14 p.
- Gingras, B.; Leclerc, J- M.; Bolduc, D.G.; Chevalier, P.; Laferrière, M.; Fortin, S. (2000): Les risques à la santé associés aux activités de production animale, Rapport scientifique du comité de santé environnementale, Ministère de la Santé et des Services sociaux du Québec, 38p.
- Gomez, E. (2002): Modélisation intégrée du transfert de nitrate à l'échelle régionale dans un système hydrologique. Application au bassin de la Seine. Thèse de doctorat de l'École Nationale Supérieure des Mines de Paris, 241 p.
- Goovaerts. (2002): Geostatistical modelling of spatial uncertainty using p-field simulation with conditional probability fields. *International Journal Geographical Information Science*, vol. 16, n° 2, pp. 167-178.
- Goss, M.; Richards, C. (2008): Development of a risk-based index for source water protection planning, which supports the reduction of pathogens from agricultural activity entering water resources. *Journal of Environmental Management*, vol. 87. pp. 623- 632.
- Gouldby, B.; Kingston, G. (2007): Uncertainty and sensitivity analysis method for flood risk analysis, T024-10-07, FLOODsite, HR Wallingford.
- Greatorex, J. S.; Thorne, G. M. (1994): Humoral immune response to shiga-like toxins and Escherichia coli O157 lipopolysaccharide in hemolytic-uremic syndrome patients and healthy subjects. *Journal of Clinical Microbiology*. vol. 32, pp. 1172-1178.
- Green, C. H.; Arnold, J. G.; Williams, J. R.; Haney, R.; Harne, R. D. (2007): Soil and water assessment tool hydrologic and water quality evaluation of poultry litter, Application to small-scale subwatersheds in Texas. *American Society of Agricultural and Biological Engineers*, vol. 50, n°4. Pp. 1199-1209.
- Green, C. H.; Van Griensven. A. (2008): Autocalibration in hydrologic modeling: Using 763 SWAT2005 in small-scale watersheds. *Environmental Modelling & Software*, vol. 23, pp.422- 434.
- Griffin, P. M.; Tauxe, R.V. (1991): The epidemiology of infections caused by Escherichia coli O157: H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. *Epidemiologic Reviews*. vol. 13, pp. 60-98.

- Grimaldi, C.; Dorioz, J. M.; Lefrançois, J.; Poulinard, J.; Macary, F.; Gascuel-Oudou, C. (2006): De l'intérêt de coupler les approches bassin versant et cours d'eau dans les petits bassins versants en zone d'élevage, Qualité de l'eau en milieu rural : savoirs et pratiques dans les bassins versants. *Collection Update Sciences & Technologies*, pp. 365.
- Grimont, P. (1987): Taxonomie des *Escherichia*. *Méd Mal Infect* (Numéro spécial), pp. 6-10.
- Groupe central ÉCLOSIONS (GCÉ). (2002): Rapport d'analyse des données du registre ÉCLOSIONS, Québec, 2000 et 2001, données provisoires 2003. 4p.
- Guber, A. K.; Pachepsky, Y. A.; Sadeghi, A. M. (2007): Evaluating uncertainty in E. coli retention in vegetated filter strips in locations selected with SWAT simulations. In: *ASABE* (Editor), *Watershed Management to Meet Water Quality and TMDLS*, San Antonio, Texas, pp. 286-293.
- Guber, A. K.; Shelton, D. R.; Pachepsky, Y. A. (2005): Effect of manure on *Escherichia coli* attachment to soil. *Journal of Environmental Quality*, vol. 34, pp. 2086-2090.
- Guyonnet, D.; Bourguine, B.; Dubois, D.; Fargier, H.; Côme, B.; Chilès, J. P. (2003): Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering, ASCE*, vol. 129, pp. 68-78.
- Guyonnet, D.; Côme, B.; Perrochet, P.; Parriaux, A. (1999): Comparing two methods for addressing uncertainty in risk assessments. *Journal of Environmental Engineering, ASCE*, vol. 125, n° 7, pp. 660-666.
- Guyot, G. (1989). *Signatures spectrales des surfaces naturelles*. Caen.
- Hamby, D. M. (1994): A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, vol. 32, n° 2, pp. 135-154.
- Hansen, S.; Thirup, C.; Refsgaard, J. C.; Jensen, L. S. (2001): Modeling nitrate leaching at different scales. Application of Daisy model. Shaffer-Liwang-Hansen Eds, *Lewis Publishers*, vol. 16, pp. 511-548.
- Hasmadi, M.; Taylor, J. C. (2008): Sensitivity analysis of an optimal access road location in HillyForest area: a GIS approach. *American Journal of Applied Sciences*, vol. 5, n° 12, pp. 1686-1692
- Hébert, S.; Ouellet, M. (2005): Le Réseau-rivières ou le suivi de la qualité de l'eau des rivières du Québec, Québec, ministère du Développement durable, de l'Environnement et des Parcs, Direction du suivi de l'état de l'environnement, 9 p.
- Hébert. (2001) : Modélisation de la qualité bactériologique d'un site potentiel de baignade à l'île saint-Quentin, Trois-Rivières. Québec, ministère de l'Environnement, Direction du suivi de l'état de l'environnement, 16 p.
- Hendry, M. J.; Lawrence, J. R.; Maloszewski, P. (1997): Role of sorption in the transport of *Klebsiella oxytoca* through saturated silica sand. *Ground Water*, vol. 35, pp. 574-584.
- Hendry, M. J.; Lawrence, J. R.; Maloszewski, P. (1999): Effects of velocity on the transport of two bacteria through saturated sand. *Ground Water*, vol. 37, pp. 103-112.

- Heuvelink, G. B. M.; Burgers, S. L. G. E.; Tiktak, A.; Berg, F. V. (2010b): Uncertainty and stochastic sensitivity analysis of the GeoPEARL pesticide leaching model. *Geoderma*, vol. 155, n°4, pp.186-192.
- Heuvelink, G. B.; Brus, M.; 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 Resources and Environmental Sciences, pages 85–88, Leicester, UK, July 20-23. 23, 141.
- Holvoet, K.; Seuntjens, P.; Mannaerts, R.; De Schepper, V.; Vanrolleghem, P. A. (2007): The water-sediment as a highly dynamic system: results of an intensive pesticide monitoring campaign. *Water Science and Technology*, vol. 55, n°3, pp. 177-182.
- Holvoet, K.; Van Griensven, A.; Gevaert, V.; Seuntjens, P.; Vanrolleghem, P. A. (2008): Modifications to the SWAT code for modelling direct pesticide losses. *Environmental Modelling & Software*, vol. 23, pp. 72-81.
- Holvoet, K.; Van Griensven, A.; Seuntjens, P.; Vanrolleghem, P. A. (2005): Sensitivity analysis for hydrology and pesticide supply towards the river in SWAT. *Physics and Chemistry of the Earth*, vol. 30, pp. 518-526.
- Homma, T.; Saltelli, A. (1996): Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety*, vol. 52, n°1, pp. 1-17.
- Houdart, M.; Bonin, M.; Le Page, C.; Fort, M.; Saudubray, F. (2005): SIG, chorèmes et systèmes multi-agent : évolution d'un système rural martiniquais et pression polluante. *Revue Internationale de Géomatique*, vol. 15, n 3, p. 339-356.
- Huang, Q.; Angelier, J. (1989): Fracture spacing and its relation to bed thickness. *Geological Magazine*, vol. 4, n° 126, pp. 355-362.
- Hughes, K. A. (2003): Influence of seasonal environmental variables on the distribution of presumptive fecal coliforms around an antarctic research station. *Applied and Environmental Microbiology*, vol. 69, pp. 4884-4891.
- Huijbregts, M. (1998): Application of Uncertainty and Variability in LCA Part I: A General Framework for the Analysis of Uncertainty and Variability in Life Cycle Assessment. *International Journal of LCA* 3, 273-280.
- Imhoff, J. C.; Kittle, J. L.; Gray, M. R. (2007): Using the Climate Assessment Tool (CAT) in U.S. EPA BASINS integrated modeling system to assess watershed vulnerability to climate change. *Water Science and Technology*, vol. 56, n° 8, pp. 49-5
- Iooss, B. (2011): Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, vol. 152, n° 1, pp.1-23.
- Islam, M.; Michael, P.; Doyle.; Sharad, C.; Phatak.; Millner, P.; Xiuping. J. (2005): Survival of *Escherichia coli* O157:H7 in soil and on carrots and onions grown in fields treated with contaminated manure composts or irrigation water. *Food Microbiology*, vol. 22, pp. 63-70.
- Jackson, D. R.; Aron, G. (1971): Parameter Estimation in Hydrology: The State of the Art. *Water Resources Bulletin*, vol. 7, n° 3, pp. 457-472.

- Jacques, J. (2005): Contributions à l'analyse de sensibilité et à l'analyse discriminante généralisée. Thèse de doctorat, Université Joseph Fourier - Grenoble 1, 209 p.
- Jacquet, S. (2005): Étude des virus, des eubactéries et des archées dans les sédiments du lac du Bourget. Colloque de restitution ECCO Toulouse, 5-7 décembre 2005.
- Jenson, S. K.; Domingue, J. O. (1988): Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering and Remote Sensing*, vol. 54, n° 11, pp. 1593-1600.
- Jiang, X.; Morgan, J.; Michael, P.; Doyle. (2002): Fate of Escherichia coli O157:H7 in manure-amended soil. *Applied and Environmental Microbiology*, vol. 68, pp. 2605-2609.
- Juan, M.; Chaix, B.; Yang, M.; Lynch, J.; Rastam, L. (2005): A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology. And Community Health*, vol. 60, pp. 290-297.
- Kang, M. S.; Park, S. W.; Lee, J. J.; Yoo, K. H. (2006): Applying SWAT for TMDL programs to a small watershed containing rice paddy fields. *Agricultural Water Management*, vol. 79, n° 1, pp. 72-92.
- Kanso, A. (2004): Inventaire des méthodes de calage et de validation des modèles paramétriques. Rapport interne. CEREVE- ENPC, URGC Hydrologie Urbaine, UMR 5569 Hydrosiences Montpellier. 43 p.
- Kavetski, D.; Kuczera, G.; Franks, S. (2006): Bayesian analysis of input uncertainty in hydrological modelling 2. Application. *Water Resources Research*, 42, doi:10.1029/2005WR004376, in press.
- Kavouri, K.; Gouin, J. (2009) : Cartographie de la vulnérabilité des aquifères karstiques, guide méthodologique de la méthode PaPRIKa. Rapport BRGM RP-57527-FR, 100 p.
- Kay, D.; Aitken, M.; Crowther, J.; Dickson, I.; Edwards, A. C.; Francis, C.; Hopkins, M.; Jeffrey, W.; Kay, C.; McDonald, A. T.; McDonald, D.; Stapleton, C. M.; Watkins, J.; Wilkinson, J.; Wyer, M. D. (2007): Reducing fluxes of faecal indicator compliance parameters to bathing waters from diffuse agricultural sources: The Brighthouse Bay study, Scotland, 147: 138-149.
- Kay, D.; Crowther, J.; Stapleton, C. M.; Bradford, M.; McDonald, A. T.; Greaves, J.; Francis, C.; Watkins, J. (2005a): Predicting faecal indicator fluxes using digital land use data in the UK's sentinel Water Framework Directive catchment: The Ribble study. *Water Research*, vol. 39, n°16. pp. 3967-3981.
- Kay, D.; Crowther, J.; Wilkinson, J.; Stapleton, C.; Glass, P. (2005b): Sustainable reduction in the flux of microbial compliance parameters from urban and arable land use to coastal bathing waters by a wetland ecosystem produced by a marine flood defence structure. *Water Research*, vol. 39, n° 14. pp. 3320-3332.
- Kirsch, K.; Kirsch, A.; Arnold, J. G. (2002): Predicting sediment and phosphorus loads in the Rock River basin using SWAT. *American Society of Agricultural Engineers*, vol. 45, n°6. pp. 1757-1769.

- Kleijnen, J. P. (1995): Verification and validation of simulation models. *European Journal of Operational Research*, vol. 82, n° 1, pp. 145-162.
- Kon, T.; Weir, S. C.; Howell, E. T.; Lee, H.; Trevors, J. T. (2009): Repetitive element (REP) - polymerase chain reaction (PCR) analysis of *Escherichia coli* isolates from recreational waters of southeastern Lake Huron. *Canadian Journal of Microbiology*, vol. 55, pp. 269-276.
- Kretzschmar, R. (1999): Mobile subsurface colloïds and their role in contaminant transport. *Advances in Agronomy*, vol. 66, p. 21-193.
- Kucherenko, S.; Tarantola, S.; Annoni, P. (2012): Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, vol. 183, n° 4, pp. 937-946.
- Kuczera, G. (1997): Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research*, vol. 33, n°1, pp. 177-185.
- Kuczera, G.; Parent, E. (1998): Monte Carlo assessment of parameter inference in catchments models: The Metropolis algorithm. *Journal of Hydrology*, vol. 211, pp. 69-85.
- Lahlou, M.; Shoemaker, L.; Choundhury, S.; Elmer, R.; Hu, A.; Manguerra, H.; Parker, A. (1998): Better Assessment Science Integrating Point and Nonpoint Sources-BASINS. EPA-823-B-98-006, USEPA, Washington, DC.
- Laine, P. (1991) : Bilan des épisodes de maladies d'origine hydrique rapportés par les DSC du Québec en 1989 et 1990. Comité de santé environnementale des DSC du Québec, sous-comité eau potable, Lévis, 58 p.
- Lamothe. (2009): Évaluation du potentiel minéral pour les gîtes porphyriques de Cu-Au ± Mo de la Baie-James. Document publié par géologie Québec ISBN : 978-2-550-55314, 54 p.
- Lang, M.; McDonough, O.; McCarty, G.; Oesterling, R.; Wilen, B. (2012): Enhanced detection of wetland-stream connectivity using LiDAR. *Wetlands*, vol. 32, pp. 461-473.
- Laroche, A. M.; Gallichand, J.; Lagace, R.; Pesant, A. (1996): Simulating atrazine transport with HSPF in an agricultural watershed. *Journal of Environmental Engineering*, vol. 122, pp. 622-630.
- Lavoie, Suzanne. (1994): Aptitudes physiques des terres agricoles à recevoir des engrais de ferme: classification à l'aide d'un système expert. Mémoire de fin de cycle, INRS-Eau, Québec, 134 p.
- Le Dimet, F-X.; Ngodock, H-E.; Luong, B.; Verron, J. (1997) : Sensitivity analysis in variational data assimilation. *Journal of the Meteorological Society of Japan*, vol. 75, n°1B, pp. 245-255.
- Levine, M. M. (1987): *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *Journal of Infectious Diseases*, vol. 155, pp. 377-389.
- Li, L.; Xia, J.; Xu, C.Y.; Singh, V. P. (2010): Evaluation of the Subjective Factors of the GLUE Method and Comparison with the Formal Bayesian Method in Uncertainty Assessment of Hydrological Models. *Journal of Hydrology*, vol. 390, pp. 210-221.

- López-Tarazón, J. A.; Batalla, R. J.; Vericat, D.; Balasch, J. C. (2010): Rainfall, runoff and sediment transport relations in a mesoscale mountainous catchment: the River Isábena (Ebro basin). *Catena*, vol. 82, pp. 23–34.
- Louchini, R.; Douville-Fradet, M. (2001): Surveillance des maladies infectieuses et des intoxications chimiques à déclaration obligatoire au Québec, de 1990 à 1999, MSSS, Direction générale de la santé publique, Bureau de surveillance épidémiologique, Québec, 279 p.
- Machado, D. C.; Maia, C. M.; Carvalho, I. D.; da Silva, N. F.; André, M. C.; Serafini, A. B. (2006): Microbiological quality of organic vegetables produced in soil treated with different types of manure and mineral fertilizer. *Brazilian Journal of Microbiology*, vol. 37, P. 538-544.
- Maggioni, V.; Anagnostou, E. N.; Reichle, R. H. (2012): The impact of model and rainfall forcing errors on characterizing soil moisture uncertainty in land surface modeling. *Hydrology and Earth System Sciences*, vol. 16, pp. 3499-3515.
- Magnan, J. (2001): Production porcine et environnement : harmonie ou contradiction, 22e colloque sur la production porcine, CRAAQ. pp.9 - 20.
- Makowski, D.; Wallach, D.; Tremblay, M. (2002): Using a bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie*, vol. 22, pp. 191-203.
- Maquaire, O.; Weber, C.; Thiery, Y.; Puissant, A.; Malet, J. P.; Wania. (2004): Current practices and assessment tools of landslide vulnerability in mountainous basins identification of exposed elements with a semiautomatic procedure – In Lacerda W.A., Ehrlich M., Fontoura S.A.B., Sayao A.S.F. (eds.), *Landslides : evaluation and stabilization*, Balkema, Taylor & Francis Group, London, p. 171-176.
- Mara, T.; Tarantola, S. (2012): Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering and System Safety*, vol. 107, pp. 115-121.
- Marrel, A.; Iooss, B.; DaVeiga, S.; Ribatet, M. (2012): Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, vol. 22, n°3, pp. 833-847.
- Mayer, A. S.; Carrière, P. P.; Gallo, C.; Pennell, K. D.; Taylor, T. P.; Williams, G. A.; Zhong, L. (1997): Groundwater quality. *Water Environment Research*, vol. 69, pp. 778-844.
- McKay, M. D.; Beckman, R. J.; Conover, W. J. (1979): A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, vol. 21, pp. 239-245.
- Merz, R.; Parajka, J.; Blöschl, G. (2011): Time stability of catchment model parameters - implications for climate impact analyses. *Water Resources Research*, 47, W02531, doi :10.1029/ 2010WR009505.
- Ministère de l’Agriculture, des Pêcheries et de l’Alimentation du Québec. (2003): Recensement de cas de toxi-infection alimentaire signalés au ministère de l’Agriculture, des Pêcheries et de l’Alimentation du Québec en lien avec l’eau entre le 1er janvier et le 31 décembre 2000. Direction des laboratoires d’expertises et d’analyses alimentaires, 2 p.

- Ministère de la Santé et des Services Sociaux (MSSS). (2005) : Santé publique : Santé environnementale. Gouvernement du Québec. <http://www.msss.gouv.qc.ca/sujets/santepub/environnementale.html#sante>.
- Ministère de l'Environnement du Québec, Ministère de l'Environnement du Québec (2001) : Portrait régional de l'eau, Accessible à: www.menv.gouv.qc.ca/eau/regions/index.htm.
- Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques. (2009): La diversité des poissons ; bassin versant de la rivière Yamaska.
- Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques, (2010) : Banque de données sur la qualité du milieu aquatique (BQMA). Québec, Ministère du Développement Durable, de l'Environnement et des Parcs.
- Ministère du Développement Durable, de l'Environnement et des Parcs. (2003): Synthèse des informations environnementales disponibles en matière agricole au Québec, Direction des politiques du secteur agricole, Ministère de l'Environnement.
- Mishra, S.; Deeds, N. E.; Ramarao, B. S. (2003): Application of classification trees in the sensitivity analysis of probabilistic model results. *Reliability Engineering & System Safety*, vol. 79, n°21, pp. 23-129.
- Moore, I. D.; Grayson, R. D.; Ladson, A. R. (1991): Digital terrain modelling; a review of hydrological, geomorphological and biological applications. *Hydrological Processes*, vol. 5, pp. 3-30.
- Moore, J. A.; Smyth, J.; Baker, S.; Miner, J. R. (1988): Evaluating coliform concentrations in runoff from various animal waste management systems. Special Report 817, Agricultural Experiment Station, Oregon State University, Corvallis. (unpublished).
- Moradkhani, H.; Sorooshian, S. (2008): General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. *Hydrological modeling and the water cycle, Water Science and Technology Library*, vol. 63, pp. 1-24.
- Morel-Septoux, H. J. (1974): Pour une théorie modifiée de l'infiltration 4e Partie : le drainage vu comme une infiltration d'air. Cah. ORSTOM. *Série Hydrologie*, vol. 6, n°3, pp. 181-188.
- Morris, M. D.; Moore, L. M.; McKay, M. D. (2006): Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs. *Journal of Statistical Planning and Inference*, vol. 136, n° 9, pp. 3203-3220.
- Moussa, R. (2005): MHYDAS Modélisation HYdrologique Distribuée des AgroSystèmes, Notice d'utilisation.
- Muela, A.; García-Bringas, J. M.; Arana, I.; Barcina, I. (2000): The effect of simulated solar radiation on Escherichia coli: the relative roles of UV-B, UV-A, and photosynthetically active radiation. *Microbial Ecology*, vol. 39, pp. 65-71.
- Mukherjee, A.; Cho, S.; Scheftel, J.; Jawahi, S.; Smith, K.; Diez-Gonzalez, F. (2006): Soil survival of Escherichia coli O157:H7 acquired by a child from garden soil recently fertilized with cattle manure. *Journal of Applied Microbiology*, vol. 101, pp. 429-436.

- Murphy, J. M.; Sexton, D. M.; Barnett, D. N.; Jones, G. S.; Webb, M. J.; Collins, M.; Stainforth, D. A. (2004): Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, vol. 430, pp. 768-772.
- Mutchler, C.; Greer, J. (1980): Effect of slope length on erosion from low slopes. *Transactions of the ASAE*, vol. 876, pp. 866-869.
- Nas, B.; Berkday, A. (2010): Groundwater quality mapping in urban groundwater using GIS. *Environmental Monitoring and Assessment*, vol. 160, pp. 215-227.
- Nash, J.; Sutcliffe, J. (1970): River flow forecasting through conceptual models: part I-A: discussion of principles. *Journal of Hydrology*, vol. 10, pp. 282-290.
- Neitsch, S. L.; Arnold, J. G.; Kiniry, J. R.; Williams, J. R. (2005): Soil and Water Assessment Tool Theoretical Documentation, Version 2005. Temple, Tex.: USDA-ARS Grassland, Soil and Water Research Laboratory. Available at: www.brc.tamus.edu/swat/doc.html. Accessed 1 october 2015.
- Neter, J.; Wasserman. (1974): *Applied Linear Statistical Models*, Irwin, Homewood (éds.), 842 p.
- Noble, R. T.; Lee, I.; Mand Schif, K. C. (2004): Inactivation of indicator micro-organisms from various sources of faecal contamination in seawater and freshwater. *Journal of Applied Microbiology*, vol. 96, n°3, pp. 464-72.
- Nolan, B.; Hit, K.; Ruddy, B. (2002): probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. *Environmental Science and Technology* vol. 36, n°10, pp. 2138-2145.
- Nossent, J.; Elsen, P.; Bauwens, W. (2011): Sobol sensitivity analysis of a complex environmental model. *Environmental Modelling and Software*, vol. 26, n°12, pp. 1515-1525.
- NRC. (1993): *Groundwater vulnerability assessment: Contamination potential under conditions of uncertainties*. National Academy Press, Washington, D.C., 185 p.
- O'Callaghan, J. F.; Mark, D. M. (1984): The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, vol. 28, pp. 323-344.
- O'Hagan, A. (2012): Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux. *Environmental Modelling & Software*, vol. 36, n° 26, pp. 35-48.
- Obled, C.; Wendling, J.; Beven, K. (1994). The sensitivity of hydrological models to spatial rainfall patterns : an evaluation using observed data. *Journal of Hydrology*, vol. 159, n°1, pp. 305-333.
- Ogden, I. D.; Hepburn, N. F.; MacRae, M.; Strachan, N. J. C.; Fenlon, D. R.; Rusbridge, S. M.; Pennington, T. H. (2002): Long-term survival of *Escherichia coli* O157 on pasture following an outbreak associated with sheep at a scout camp. *Letters in Applied Microbiology*, vol. 34, pp. 100-104.
- Omlin, M.; Reichert, P. (1999): A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling*, vol. 115, n°1, p.45-59.

- Organisme Bassin Versant Yamaska. (2010a). Plan d'action du bassin versant 2010-2015. [En ligne] <http://www.obvyamaska.qc.ca/plan-action>. Visité en octobre 2012. OBV Yamaska, 2010b. Portrait du bassin versant de la Yamaska. [En ligne] <http://www.obvyamaska.qc.ca/analyse-du-bassin-versant> Visité en décembre 2013.
- Parajuli, P.; Mankin, K. R.; Barnes, P. L. (2007): New methods in modeling source-specific bacteria at watershed scale using SWAT. In: Watershed Management to meet Water Quality Standards and TMDLs (Total Maximum Daily Load) Proceedings. ASABE 831 Publication No. 701P0207. ASABE, St. Joseph, MI.
- Passerat, J.; Ouattara, N. K.; Mouchel, J-M.; Rocher, V.; Servais, P. (2011): Impact of an intense combined sewer overflow event on the microbiological water quality of the Seine River. *Water Research*, vol. 45, n° 2, pp. 893-903.
- Pechlivanidis, I. G.; Jackson, B.; McMillan, H. (2010): The use of entropy as a model diagnostic in rainfall runoff modelling, in iEMSs 2010: International Congress on *Environmental Modelling and Software*, 5- 8 July, Ottawa, Canada.
- Pei, X. B.; Zhao, D. Z. (2000): A GIS-SD-based spatiotemporal modelling and regulating policies on water pollution in Dalian gulf. *Journal of remote sensing*, vol. 4, n°2, pp.118-124.
- Pételet, M.; Asserin, O.; Iooss, B.; Loredó, A. (2006): Échantillonnage LHS des propriétés matériaux des aciers pour l'analyse de sensibilité globale en simulation numérique du soudage. In Actes de MATERIAUX 2006, Dijon, France.
- Pételet-Giraud, E.; Dörfliker, N.; Crochet, P. (2000) : RISKE : méthode d'évaluation multicritère de la cartographie de la vulnérabilité des aquifères karstiques. Applications aux systèmes des Fontanilles et Cent-Fonts (Hérault). *Hydrogéologie*, vol. 4, pp. 71-88.
- Pitt, R.; Voorhees, J. (2000): The Source Loading and Management Model (SLAMM) A Water Quality Management Planning Model for Urban Stormwater Runoff. University of Alabama. Properties of water on solar altitude. *Limnology and Oceanography*, vol. 29, pp. 824-830.
- Plagnes, V.; Théry, S.; Bakalowicz, M.; Fontaine, L.; Dörfliker, N. (2006) : Cartographie de la vulnérabilité des karsts, évolution de la méthode RISKE, Colloque Rouen, 2006.
- Poesen, J. (1984): The influence of slope angle on infiltration rate and Hortonian overland flow volume. *Zeitschrift für Geomorphologie*. vol, 49, pp.117-131.
- Poesen, J. (1987): The role of slope angle in surface seal formation. In : V. Gardiner (editor), *International Geomorphology II* : p.437-448.
- Pohl, P.; Lintermans, B.; Mainil, J.; Deprez. (1989) Production de vérocytotoxine par les *Escherichia coli* du porc. *Annales de Médecine Vétérinaire*, vol. 133, pp. 31-38.
- Pommepeuy, M.; Hervio-Heath, D.; Caprais, M. P.; Gourmelon, M.; Le Saux, J. C.; Le Guyader F. (2005): Fecal contamination in coastal areas : An engineering approach. In: S.B.a.R.R. Colwell (Editor), *Oceans and Health: Pathogens in the Marine Environment*. New York, NY: Springer Science + Business Media, p. 331- 359.
- Prabin, K. (2012): Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. *Nat Hazards*, vol. 63, p. 479-498.

- Prado, R. B.; Daniel, L. A. (2000): Application of QUAL2E model and GIS to verify the contribution by nonpoint pollution sources in the degradation of the middle Pardo River - SP - Brazil. 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs., Banff, Alberta, Canada, September 2 - 8, 2000.
- Pranville, J.; Plagnes, V.; Rejiba, F.; Trémoule, J. (2008) : Cartographie de la vulnérabilité sur la partie sud du Causse de Gramat : application de la méthode RISKE 2. "Géologues", 156, 44-47.
- Qiu-Ling, Z.; Yingxu, C.; Ghulam, J.; Imran, H. S.; Qiaogang, Y. (2009): Model AVSWAT a propos of simulating non-point source pollution in Taihu lake Basin. *Journal of Hazardous Materials*, vol. 174, n° 2010, pp. 824-830.
- RAPPEL. (2006): Diagnostic environnemental global du bassin versant immédiat du lac Massawippi (Municipalité d'Ayer's Cliff et de Hatley). Réd. M. Desautels et C. Rivard-Sirois, Sherbrooke, 103p. + 5 annexes.
- Reghunath, R.; Murthy, T. R. S.; Raghavan, B. R. (2002): The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research* vol. 36, pp. 2437-2442.
- Rogel. (2007) : Développement et intégration au système de modélisation intégré GIBSI d'un modèle de transport et du devenir des coliformes fécaux d'origine agricole à l'échelle du bassin versant. Mémoire de fin de cycle, INRS-ÉTÉ, Québec, 178 p.
- Rosero, E.; Yang, Z-L.; Wagener, T.; Gulden, L. E.; Yatheendradas, S.; Niu, G.-Y. (2010): Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the noah land surface model over transition zones during the warm season. *Journal of Geophysical Research*, vol. 115, 21 pp D03106, doi:10.1029/2009JD012035. (21 pages) [pdf]
- Rouleau, A.; Gale, J. E. (1985): Statistical characterization of the fracture system in stripa granite, Sweden. *International Journal of Rock Mechanics and Mining Sciences*, vol. 22, pp. 353-367
- Rousseau, A. N.; Savary, S.; Konan, B. (2008): Implantation du modèle HYDROTEL sur le bassin de la rivière Montmorency afin de simuler les débits observés et de produire des scénarios de crues du printemps pour l'année 2008. Centre Eau Terre et Environnement, Institut national de la recherche scientifique, Québec.
- Rozen, Z.; Belkin, S. (2001): Survival of enteric bacteria. *FEMS Microbiology. Reviews*, vol. 25, pp. 513-529.
- Ruhard, J. P. (1978): Carte de vulnérabilité à la pollution de la nappe superficielle par les décharges d'ordures ménagères en Gironde, Rapport BRGM 78 SGN 384 AQI, 26 p.
- Sadeghi, A. M.; Arnold, J. G. (2002): A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales. In Total Maximum Daily Load (TMDL) Environmental Regulations. In: Proceedings of the March 11-13, 2002 Conference at Fort Worth, Texas, USA. ASAE Publication 701P0102, pp. 56-63.

- Saint-Geours, N.; Lavergne, C.; Bailly, J-S.; 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, pages 178-189, Salzburg, Austria, September 5-9. 144
- Saltelli, A.; Annoni, P. (2010): How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, vol. 25, n°12, pp. 1508-1517.
- Saltelli, A.; Annoni, P.; Azzini, I.; Campolongo, F.; Ratto, M.; Tarantola, S. (2010): Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, vol. 181, pp. 259-270.
- Sanborn, M.; Takaro, T. (2013): Maladies reliées aux loisirs aquatiques: Prise en charge et prévention en cabinet Révision clinique. *Le Médecin de famille canadien*, vol. 59, n° 2, pp. 225-230.
- Santhi, C.; Arnold, J. G.; Williams, J. R.; Dugas, W. A.; Srinivasan, R.; Hauck, L. M. (2001): Validation of the SWAT model on a large river basin with point and nonpoint sources. *Journal of the American Water Resources Association*, vol. 37, n°5, pp. 1169-1188.
- Saporta, G. (2006): Probabilités, Analyse des données et Statistiques, Paris, Éditions *Technip*, 622 p.
- Sathian, K. K.; Syamala, P. (2010): Application of GIS integrated SWAT model for basin level water balance. Department of Land and Water Resources, Kelappaji College of Agricultural Engineering and Technology, Tavanur.
- Sathyendranath, S.; Platt, T. (1997): Analytic model of ocean colour. *Applied Optics*. vol. 36, pp. 2620-2629.
- Schnebelen, N.; Platel, J. P.; Le Nindre, Y.; Baudry, D. (2001): Gestion des eaux souterraines en Aquitaine – Année 5. Opération sectorielle – Protection de la nappe de l’Oligocène en région bordelaise. Nouvelles connaissances hydrogéologiques – Cartographie de la vulnérabilité aux pollutions, Rapport BRGM/RP-51178-FR, 75 p.
- Schoen, R.; Ricard, J. (2000): Unités de gestion de la qualité des eaux continentales en régions Midi-Pyrénées. Avec la collaboration de Codvelle A. Rapport BRGM/RP-50569-FR, 2001.
- Sciuto, G.; Diekkrueger, B. (2010): Influence of soil heterogeneity and spatial discretization on catchment water balance modelling. *Vadose Zone Journal*, vol. 9, pp. 955-969.
- Sebillote, M. (1990): Les processus de décision des agriculteurs. Deuxième partie : conséquences pour les démarches d'aide à la décision. In : Brosier et al. (éd.), pp.103-117.
- Seguin, J. J.; Baudry, D. (2002): Cartographie de la vulnérabilité des systèmes aquifères de l’Aquitaine, Rapport BRGM/RP-52042-FR, 53 p.
- Semenzin, E.; Critto, A.; Rutgers, M.; Marcomini, A. (2007): Integration of bioavailability, ecology and ecotoxicology by three lines of evidence into ecological risk indexes for contaminated soil assessment. *Science of the Total Environment*, vol. 389, n°1, pp. 71-86.
- Sener, E.; Davraz, A. (2013): Assessment of groundwater vulnerability based on a modified DRASTIC model, GIS and an analytic hierarchy process (AHP) method: the case of Egirdir Lake basin (Isparta, Turkey). *Hydrogeology Journal*, vol. 21. n° 3, pp. 701 -714.

- Servais, P. (2005): Microbiological contamination in the Seine river drainage network. *Science of the Total Environment*, Submitted.
- Servais, P.; Garcia-Armisen, T.; Lepeuple, A. S.; Lebaron, P. (2005): An early warning method to detect faecal contamination of river waters. *Annals of Microbiology*, vol. 55, n°2, pp. 151-156.
- Shrestha, S.; Kazama, F. (2007): Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River basin, Japan. *Environmental Modelling & Software*, vol. 22, pp. 464-475.
- Sin, G.; Gernaey, K.V.; Neumann, M. B.; Van Loosdrecht, M.C.M.; Gujer, W. (2009): Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, vol. 43, n°11, pp. 2894-2906.
- Sjogren, R. E. (1994): Prolonged survival of an environmental *Escherichia coli* in laboratory soil microcosms. *Water, Air, & Soil Pollution*, vol. 75, pp. 389-403.
- Smith, A.; Reacher, M.; Smerdon, W.; Adak, G. K.; Nichols, G.; Chalmers, R. M. (2006): Outbreaks of waterborne infectious intestinal disease in England and Wales. *Epidemiology & Infection*, vol. 134, pp. 1141-1149.
- Smith, T. J.; Marshall, L. A. (2008): Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. *Water Resources Research*, 44, doi: 10.1029/2007WR006705.
- Sobol, I. M. (1993): Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, vol. 1, n°4, pp. 407-414.
- Soil Conservation Service (1972): National Engineering Handbook, Section 4, U.S. Department of Agriculture, Washington, D.C.
- Sorooshian, S.; Dracup, J. A. (1980): Stochastic parameter estimation procedures for hydrologic rainfall-runoff models. Correlated and Heteroscedastic error cases. *Water Resources Research*, vol. 16, n°2, pp. 430-442.
- Spiegelhalter, D. J.; Knill-Jones, R. P. (1984): Statistical and Knowledge Approaches to Clinical Decision-support Systems, with an Application in Gastroenterology. *Journal of the Royal Statistical Society*, vol. 147, n°1. pp. 35-77.
- Srinivasan, R.; Arnold, J. G. (1994): Integration of basin-scale water quality model with GIS. *Water Resources Bulletin*, vol. 30, n°3, pp. 453-462.
- Statistique Canada, (2012): « Recensement de l'agriculture de 2011. » Faits saillants et analyses. Données sur les exploitations et les exploitants agricoles. <http://www.statcan.gc.ca/pub/95-640-x/2012002-fra.htm>.
- Stein, M. (1987): Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, vol. 29, pp. 143-151.
- Steven, C.; White, H.; Golomb, B. A. (2001): Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*, vol. 54, n°10, p. 979-985.

- Storlie, C. B.; Swiler, L. P.; Helton, J. C.; Sallaberry, C. J. (2009): Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering and System Safety*, vol. 94, pp. 1735-1763.
- Sudret, B. (2008): Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, vol. 93, pp. 964-979.
- Swaen, G.; Van Amelsvoort, L. (2009): A weight of evidence approach to causal inference. *Journal of Clinical Epidemiology*, vol. 62, n°3, pp. 270-277.
- Tarboton, D. G.; Schreuders, K. A. T.; Watson, D. W.; Baker, M. E. (2009): Generalized terrain-based flow analysis of digital elevation models," 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, ed. R. S. Anderssen, R. D. Braddock and L. T. H. Newham, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and *Computers in Simulation*, July 2009, pp. 2000-2006.
- Thiery, Y.; Malet, J.-P.; Sterlacchini, S.; Puissant, A.; Maquaire, O. (2007): Landslide susceptibility assessment by bivariate methods at large scales: application to a complex mountainous environment. *Geomorphology*, vol. 92, p. 38-59.
- Thokozani, K.; Levy, J.; Xu, Yo.; Saka, J. (2012): Assessment of microbial contamination of groundwater in upper Limphasa River catchment, located in a rural area of northern Malawi. *Water SA*, vol. 38, n° 4, p. 581-595.
- Tian, Yong Q.; Peng Gong.; Radke, J. D.; Scarborough, J. (2002): Spatial and temporal modeling of microbial contaminants on grazing farmlands. *Journal of Environmental Quality*, vol. 31, n°3, p.860- 869.
- Tissot, J.-Y.; Prieur, C. (2012): Bias correction for the estimation of sensitivity indices based on random balance designs. *Reliability Engineering and System Safety*, vol. 107, pp. 205-213.
- Tolson, B. A.; Shoemaker, C. A. (2004): Watershed modeling of the Cannonsville Basin using SWAT2000: Model development, calibration and validation for the prediction of flow, sediment and phosphorus transport to the Cannonsville reservoir, version 1.0, technical report, Sch. of Civ. and Environ. Eng. Cornell Univ., Ithaca, N. Y.
- Tuffery. (2006): Data Mining et statistique décisionnelle. *L'intelligence des données*, vol. 26, pp.102-108
- Turányi, T.; Rabitz, H. (2004): Local Methods, in: Sensitivity Analysis, Saltelli, A., Chan, K., Scott, E. M., (Eds.), *John Wiley & Sons Publication*, pp. 81-99.
- U.S. EPA (Environmental Protection Agency) (2001): Better assessment science integrating point and non-point sources: BASINS version 3 user's manual, EPA-823-H-01-001, June 2001. Available online at <http://www.epa.gov/waterscience/basins/bsnsdocs.html>.
- Uhlenbrook, S.; Seibert, J.; Leibundgut, C.; Rodhe, A. (1999): Prediction uncertainty of conceptual rainfall runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences*, vol. 44, n°5, pp. 779-797.
- Union Québécoise pour la Conservation de la Nature (UQCN) (2005): La gestion du territoire et des activités agricoles dans le cadre de l'approche par bassin versant-bassin versant de la

rivière Kamouraska.

- Van Griensven, A.; Bauwens, W. (2003): Multi-objective auto-calibration for semi-distributed water quality models. *Water Resources Research*, vol. 39, n°10, pp. 1348-1357.
- Van Oijen, M.; Rougier, J.; Smith, R. (2005). Bayesian calibration of process-based forest models : bridging the gap between models and data. *Tree Physiology* vol. 25, 915-927.
- Vaze, J.; Chiew, F. H.; Perraud, J-M.; Viney, N. R.; Post, J.; Teng, B.; Wang, J.; Goswami, M. (2010): Rainfall-runoff modelling across southeast Australia: Datasets, models and results. *Australian Journal of Water Resources*, vol. 14, n°2, pp. 101-116.
- Vaze, J.; Teng, J. (2011): Future climate and runoff projections across new south wales, Australia: results and practical applications. *Hydrological Processes*, vol. 25, n°1, pp. 18-35.
- Wagner, T.; McIntyre, N.; Lees, M. J.; Wheeler, H. S. (2003): Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes*, vol.17, n°2, pp.455-476.
- Wang, G.; Doyle, M. P. (1998): Survival of enterohemorrhagic Escherichia coli O157:H7 in water. *Journal of Food Protection*, vol. 61, pp. 662-667.
- Wang, L.; Mankin, K. R.; Marchin, G. L (2004): Survival of fecal bacteria in dairy cow manure. *Trans Am Soc Agric Eng*, vol. 47, n°4, pp. 1239-1246.
- Weed. (2005): Weight of evidence: a review of concept and methods. *Risk Anal*, vol. 25, n°6, pp. 1545-1557.
- Weichel, T.; Pappenberger, F.; Schulz, K. (2007): Sensitivity and uncertainty in flood inundation modelling: concept of an analysis framework. *Advances in Geosciences*, vol. 11, n°112, pp. 31-36.
- Werner, A. D.; Gallagher, M. R.; Weeks, S. W. (2006): Regional-scale, fully coupled modelling of stream-aquifer interaction in a tropical catchment. *Journal of Hydrology*, vol. 328, pp. 497-510.
- Wilby, R. L. (2005): Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological Processes*, vol. 19, n°16, pp. 3201-3219.
- World Meteorological Organization (WMO) (1975): International Cloud Atlas: Manual on the Observation of Clouds and Other Meteors. Volume I, (WMO no 407), Geneva.
- Wriedt, G.; Rode, M. (2006): Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH. *Advances in Geosciences*, vol. 9, pp.145-150.
- Xu, C.; Gertner, G. (2011): Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). *Computational Statistics & Data Analysis*, vol. 55, n°1, pp. 184-198.
- Yang, Y. H.; Zhou, F.; Guo, H. C.; Sheng, H.; Liu, H.; Dao, X.; He, C. J. (2010): Analysis of spatial and temporal water pollution pattern in Lake Dianchi using multivariate statistical methods. *Environemtal Monitoring and Assessment*, vol. 170, pp. 407-416.
- Yaun, B. R.; Sumner, S. S.; Eifert, J. D.; Marcy, J. E. (2003): Response of Salmonella and E. coli O157:H7 to UV energy. *Journal of Food Protection*, vol. 66, n°6, pp. 1071-1073.

- Zabaleta, A.; Uriarte, J. A.; Antiguada, I. (2010): Streamflow response during rainfall events in a small forested catchment (Basque Country), in: Status and Perspectives of Hydrology in Small Basins. *International Association of Hydrological Sciences*, vol. 336, pp. 125-130.
- Zeilhofer, P.; Zeilhofer, L.V.; Haroim, E. L.; Marques de Lima, Z.; Silva, O. C. (2007): GIS applications for mapping and spatial modeling of urban-use water quality: a case study in District of Cuiabá, Mato Grosso, Brazil. *Cadernos de saúde pública*, vol. 23, nº4, p. 875 - 884.
- Zhang, X.; Wang, Q. S.; Liu, Y. F.; Wu, J.; Yu, M. (2011): Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. *Environmental Monitoring and Assessment*, vol. 173, pp. 17-27.
- Zhang, Y.; Guo, F.; Meng, W.; Wang, X.; Q. (2009). Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. *Environmental Monitoring and Assessment*, vol. 152, pp. 105-121.
- Zhanga, Q-L.; Chena, Y-Xu.; Ghulam, J. B.; Imran, H. S.; Yu, Q-G. (2010): Model AVSWAT a propos of simulating non-point source pollution in Taihu lake basin. *Journal of Hazardous Materials*, vol. 174, pp. 824-830.

ANNEXES

Annexe 1. Les sous-bassins versants de la rivière Yamaska

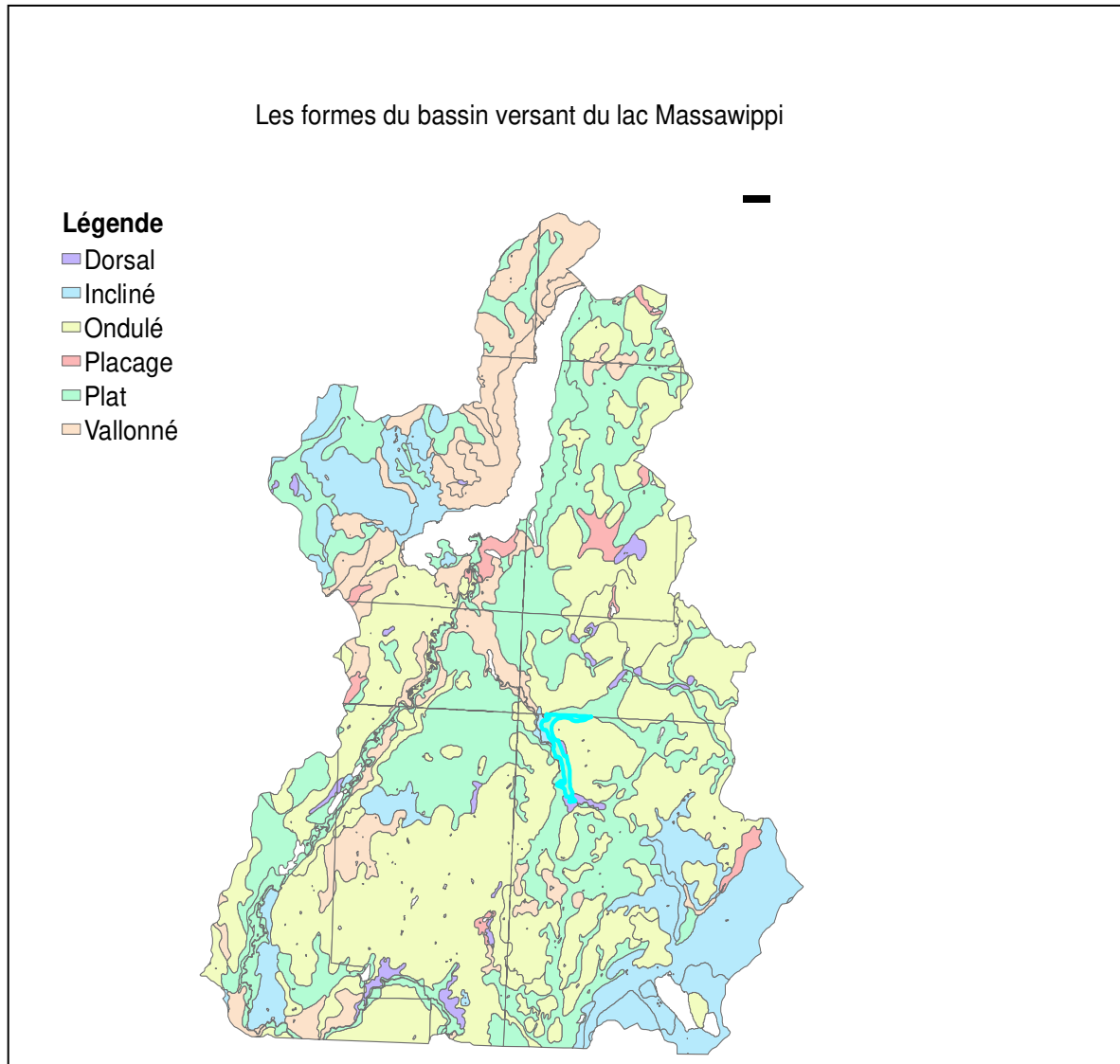


Annexe 2. Les sous-bassins du bassin versant du Lac Massawippi

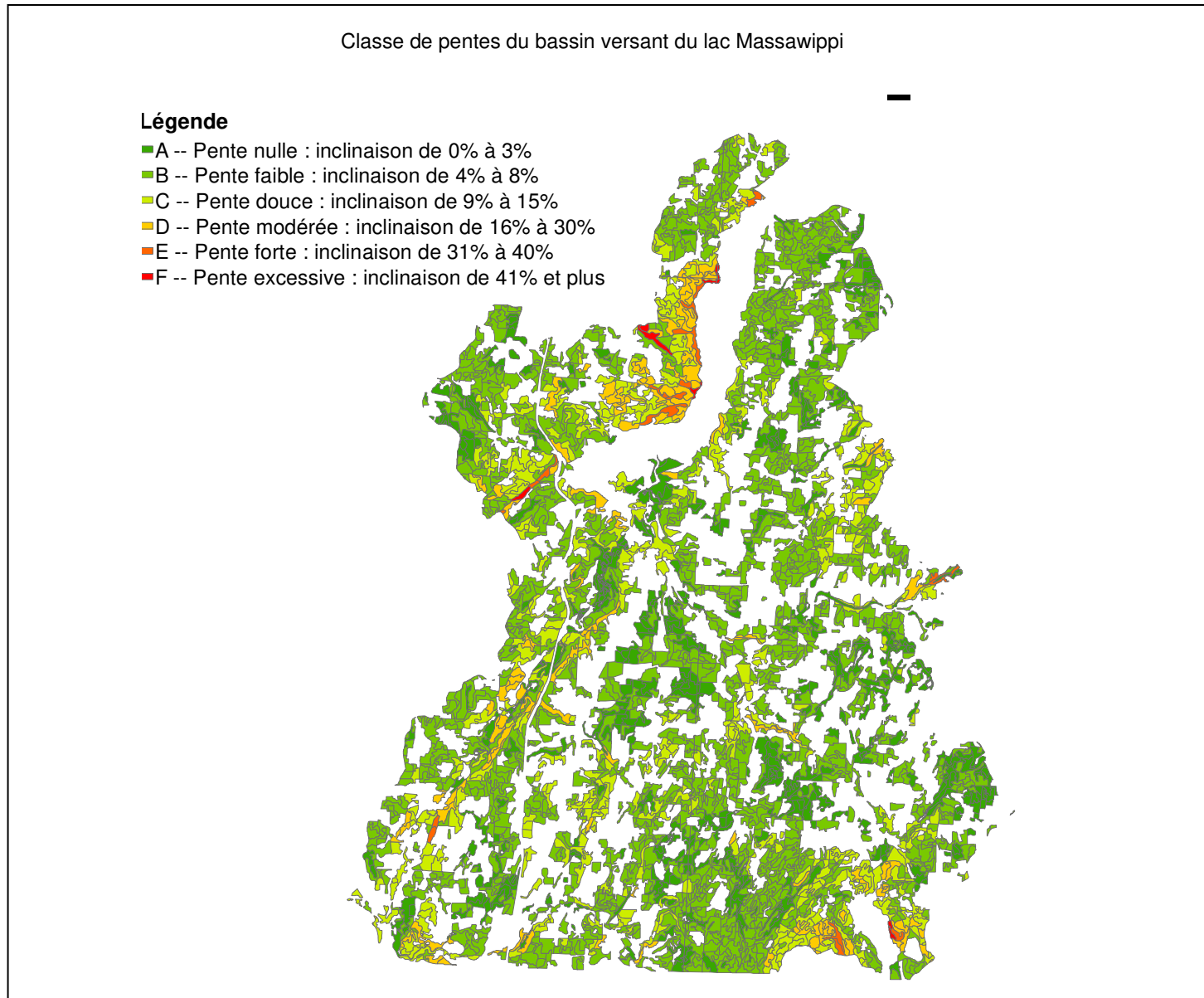


Annexe 3. Variables morphologiques

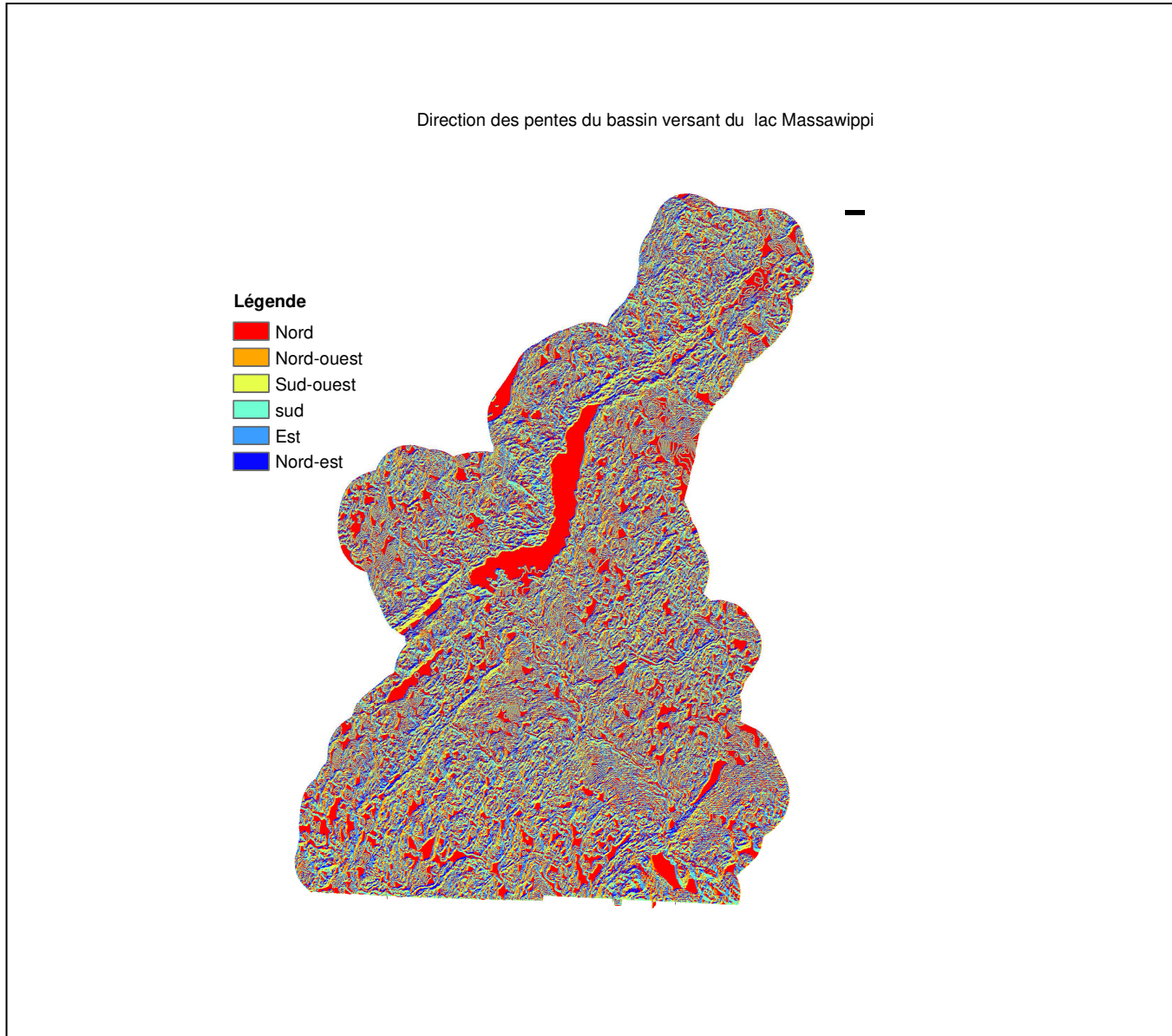
Annexe 3.1. Morphologie du bassin versant du lac Massawippi : les formes



Annexe 3.2. Morphologie du bassin versant du lac Massawippi : les pentes

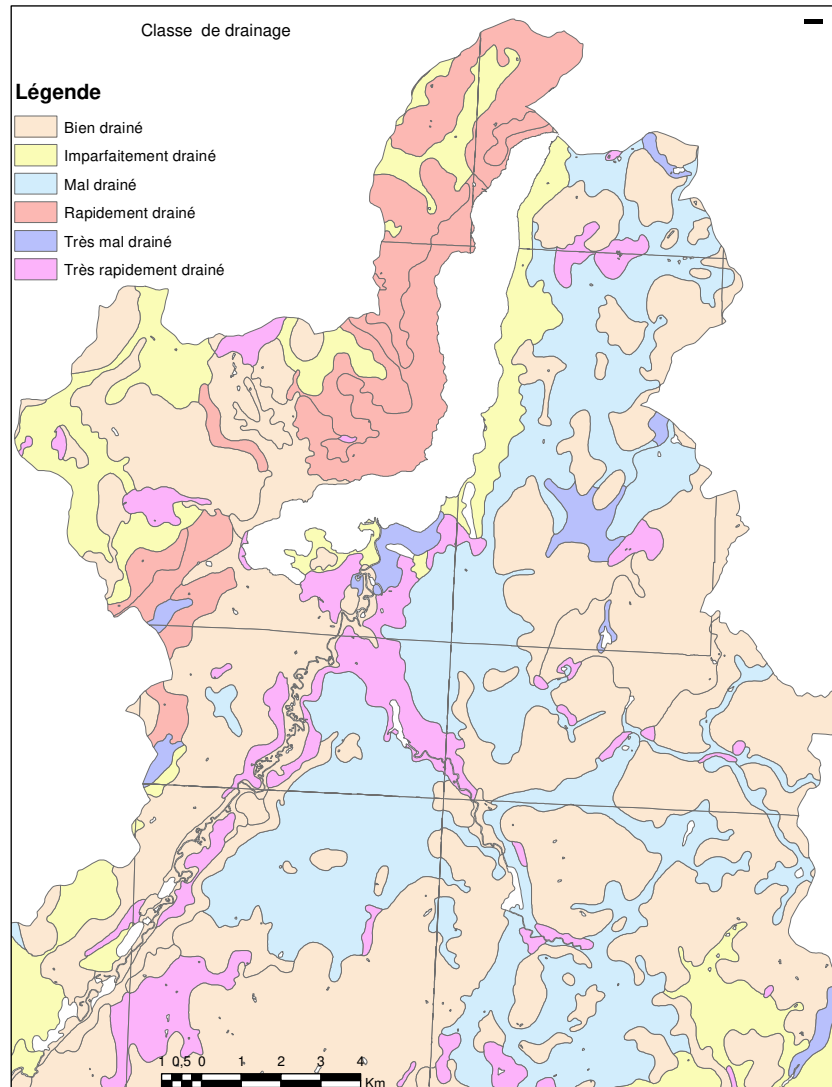


Annexe 3.3. Morphologie du bassin versant du lac Massawippi : la direction des pentes

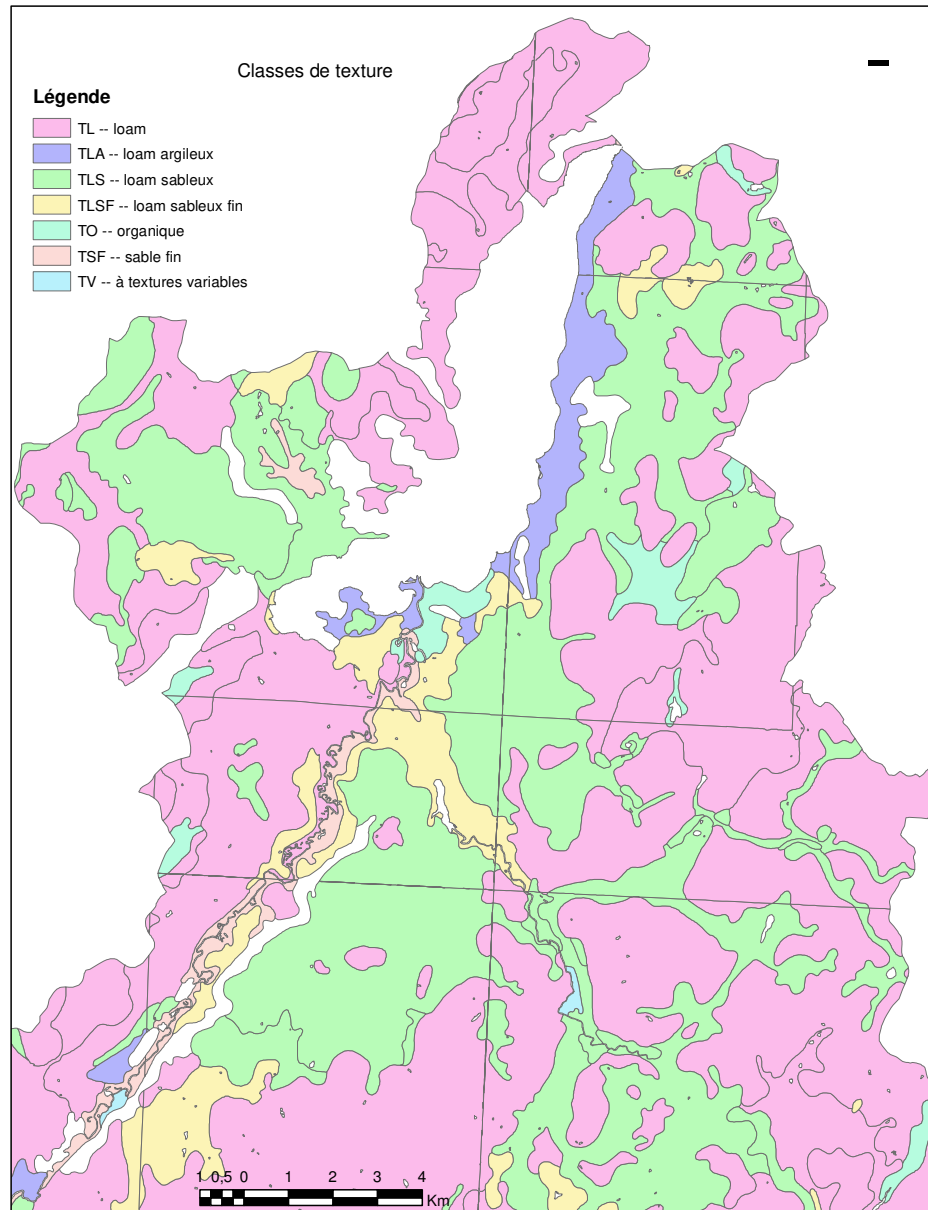


Annexe 4. Variables pédologiques

Annexe 4.1. Variables pédologiques du bassin versant du lac Massawippi : le drainage



Annexe 4.2. Variables pédologiques du bassin versant du lac Massawippi : le drainage



Annexe 5. Processus de sélection de la méthode d'interpolation des données de précipitations.

Dans le but d'obtenir des résultats les plus près possible de la réalité, plusieurs type de traitements géostatistiques furent mis à l'essai. Pour se faire nous avons établi quatre types de répartition des données journalières afin d'en évaluer les résultats avec différentes méthodes d'interpolation :

Données homogènes : Il s'agit de journées durant lesquelles les précipitations sont assez constantes d'un endroit à l'autre dans notre région d'intérêt.

Données hétérogènes : Lorsqu'on observe des fluctuations importantes d'une région à l'autre.

Données extrêmes : Lorsque les données sont plutôt constantes excepté pour quelques valeurs qui diffèrent largement des autres.

Données locales : Lorsqu'on observe un phénomène différent de la tendance générale dans une région en particulier.

Plusieurs méthodes d'interpolation peuvent être utilisées pour ce genre de données, allant de méthode assez complexe comme le co-krigeage jusqu'à des procédés plutôt simple comme la distance inverse pondérée (IDW). Puisque nous disposions d'un grand jeu de donnée, notre but était d'obtenir des résultats les plus fidèles possible tout en minimisant notre temps de traitement. Malheureusement pour automatiser le processus il faut utiliser modelbuilder dans ArcGIS 10 qui contient qu'un nombre limité de méthodes. Nous avons tout de même utilisé l'extension géostatistique pour vérifier l'éventail des possibilités. Les procédés retenus sont : krigeage ordinaire, co-krigeage ordinaire, distance inverse pondérée, spline, surface de tendance (trend) ainsi que les polynomiales locale et globale. Chaque méthode fut expérimentée sur les quatre jeux de données retenues préalablement afin d'en comparer les résultats.

Krigeage ordinaire : Cette méthode semblait donner d'excellents résultats à priori avec l'extension géostatistique, car nous pouvions personnaliser plusieurs paramètres. Toutefois puisque modelbuilder prend seulement les fonctions préétablies dans une boîte à outil, il a fallu utiliser l'outil de krigeage de l'extension d'analyse spatiale qui elle possède moins de paramètres d'ajustement. Pour les données homogènes et hétérogènes, les résultats sont satisfaisants par contre lorsqu'on observe des valeurs extrêmes et des tendances locales, la méthode ne conservait plus les minimums et maximums. De plus, dans certains cas nous obtenions des valeurs

interpolées négatives.

Co-krigeage ordinaire : la couche de donnée ajoutée à celle des précipitations est un modèle numérique d'élévation (DEM) de 5000 points. Cela améliore certes les résultats du krigeage seul, mais il aurait fallu traiter les fichiers un à un puisque cette méthode n'apparaît pas dans l'extension d'analyse spatiale ce qui aurait été une tâche longue et fastidieuse.

Distance inverse pondérée : C'est un procédé assez simple qui tient compte de la distance entre chaque station pour évaluer les précipitations aux alentours. Les résultats obtenus pour les quatre types de données journalières sont assez satisfaisant, les minimums et maximums sont conservés. De plus contrairement à d'autres méthodes, les coupures entre les diverses zones de précipitations sont assez douces, ce qui est représentatif de la réalité.

Spline : Cette méthode est plutôt adaptée pour dessiner des courbes de niveau ou encore sortir de l'information d'un modèle d'élévation. Puisque nous obtenions des valeurs négatives à chaque interpolation, nous avons exclu ce procédé.

Surface de tendance : Tout comme le spline, le problème majeur de cette méthode est qu'elle introduit des valeurs négatives dans l'interpolation. Elle n'est donc pas adaptée pour des données de précipitations.

Polynomiales locale et globale : Ces méthodes ne sont pas tout à fait appropriées, dans certains cas les résultats sont acceptables, mais souvent il n'y a pas assez de précision dans les zones de précipitations, il s'agit d'un procédé un peu trop généraliste.

Résultats : La méthode retenue est la distance inverse pondérée qui permet une automatisation des interpolations des données journalières de 2010 et 2011 à partir de modelbuilder. En plus de respecter notre contrainte de temps, ce procédé conserve nos minimums et maximums de précipitations et est celui qui semble visuellement suivre, la répartition naturelle des précipitations. Les tableaux 6 et 7 présentent respectivement la répartition des précipitations ainsi que leur minimum, maximum, moyenne et écart type. Par la suite, ces éléments statistiques sont comparés pour chaque type d'interpolation.

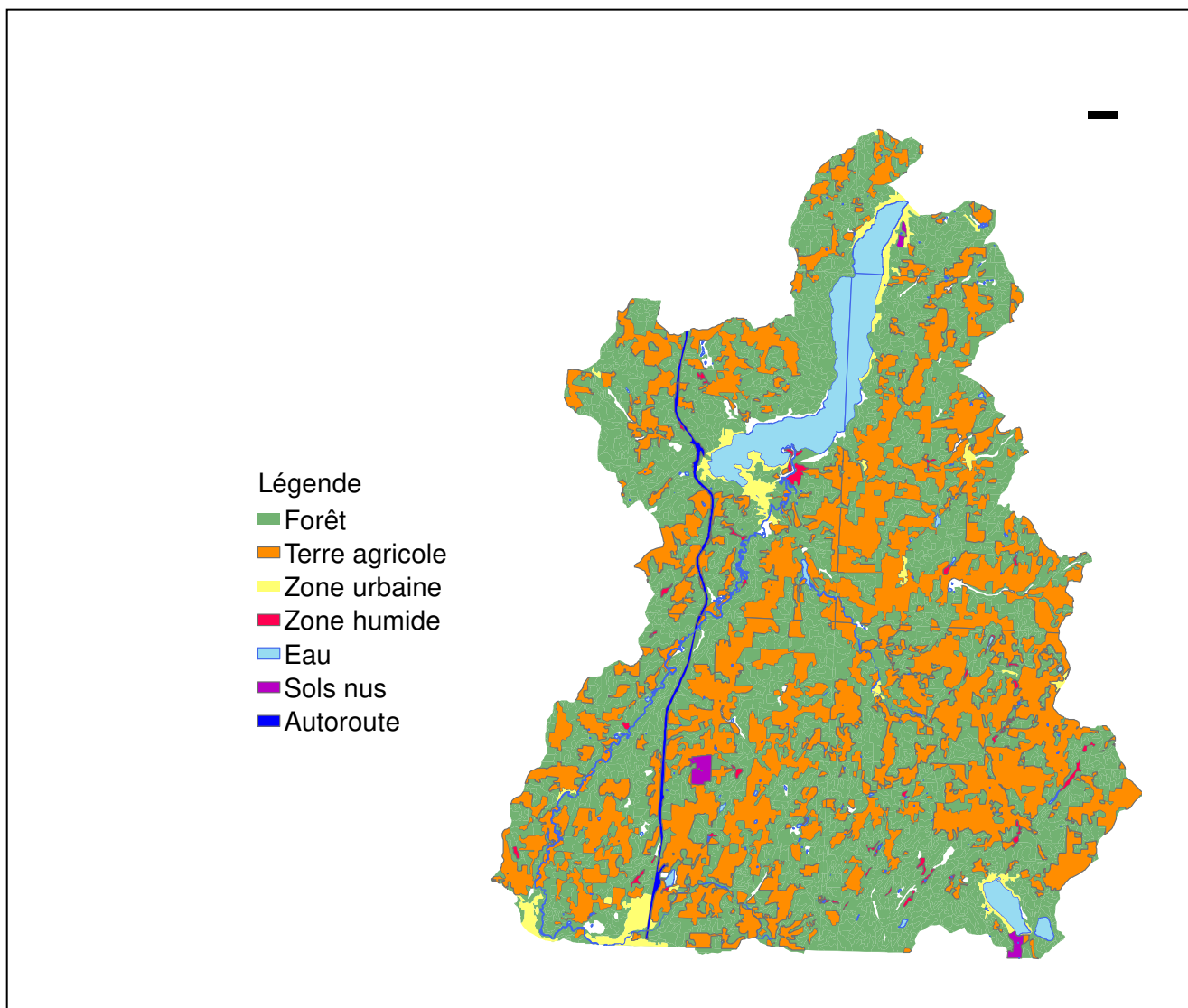
Annexe 6. Répartition des données pluviométriques

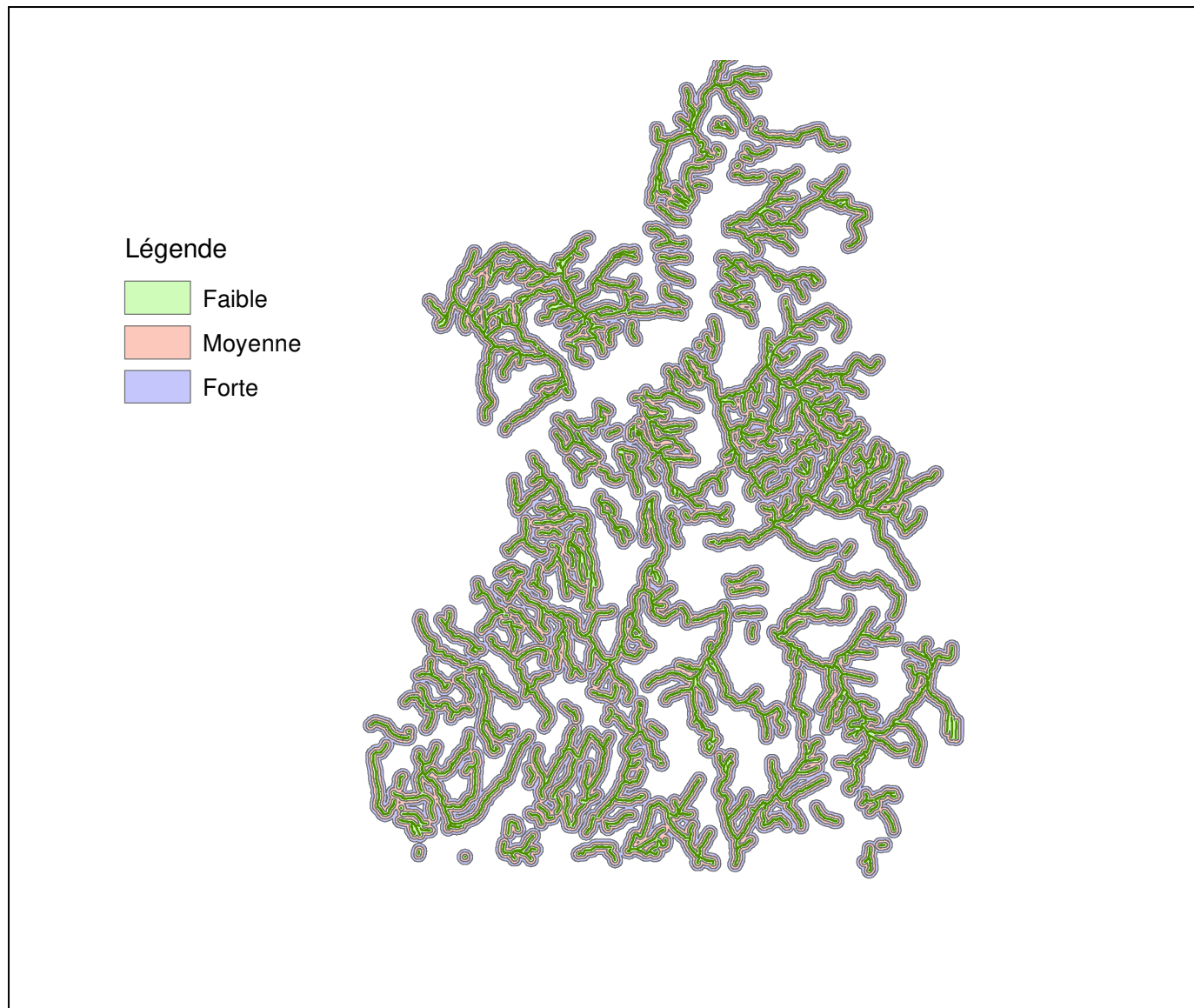
Annexe 6.1. Répartition des pluies par stations

| STATION | Homogène | Hétérogène | Locale | Extrême |
|-----------------------|--------------|--------------|-------------|-------------|
| BONSECOURS | N/A | 33 | 1,8 | 0 |
| BROME | 11 | 14,2 | 0 | 0 |
| BROMPTONVILLE | 14,5 | 23 | 0 | 0 |
| COATICOOK | 10,5 | 0 | 0 | 0,4 |
| DANVILLE | 10 | 18,2 | 1,4 | 0 |
| COOKSHIRE | 8,6 | 0 | 2 | 0,4 |
| DUNHAM | 9 | 22 | 0 | 0 |
| FARNHAM | 9,8 | 22,2 | 0 | 0 |
| FRELIGHSBURG | 11,8 | 27 | 0 | 0 |
| GEORGEVILLE | 11,2 | 0 | 0 | 0 |
| GRANBY | 9,2 | 3,2 | 2,2 | 0 |
| HENRYVILLE | 7,8 | 0 | 0 | 0 |
| IBERVILLE | 5,9 | 0 | N/A | 0 |
| LAPRAIRIE | 0 | 1,6 | 0 | 0 |
| LENNOXVILLE | 10 | 0 | 0 | 0,2 |
| LINGWICK | 13,6 | 33 | 0 | 0,4 |
| MAGOG | 12,1 | N/A | 0 | 0,6 |
| MARIEVILLE | N/A | 19 | 2 | 0 |
| MILAN | 17,2 | 1,6 | 0 | 0 |
| NOTRE-DAME-DU-BOIS | 12,6 | 5 | 0 | 0,4 |
| PHILIPSBURG | 7,5 | 21,5 | 1 | 0 |
| ROXTON | 9 | 2 | 2 | 0 |
| RICHMOND | 12,6 | 32 | 0 | 0 |
| SABREVOIS | 8,8 | 0,2 | N/A | 0 |
| ST-BERNARD-DE-LACOLLE | 9,6 | 0,6 | 6,3 | 0 |
| ST-HYACINTHE | 8,2 | 1 | N/A | 0 |
| STE-MADELEINE | 6,2 | 3,8 | 14 | 0 |
| ST-MALO-D'AUCKLAND | 13 | 1,3 | 0 | 0,6 |
| ST-NAZAIRE | 8,6 | 6,2 | 0 | 0 |
| SAWYERVILLE-NORD | 12 | 0,4 | 0 | 0 |
| SHERBROOKE | 12,1 | 21,3 | 3 | 4,1 |
| SUTTON | 9,2 | 28,6 | 0,4 | 0 |
| UPTON | N/A | 10,2 | 11 | 0 |
| STE-CAMILLE-DE-WOLFE | 12,4 | 59,6 | 0 | 0,4 |
| MIN | 0 | 0 | 0 | 0 |
| MAX | 17,2 | 59,6 | 14 | 4,1 |
| MOY | 10,13 | 12,48 | 1,52 | 0,22 |
| Ecart-type | 3,09 | 14,45 | 3,25 | 0,71 |

Annexe 6.2. Comparaisons des résultats d'interpolation

| Résultats des interpolations | | | | | | | |
|------------------------------|------------|------------|------------|-----------|------------|------------|------------|
| Homogène | | | | | | | |
| | DIP | IPL | IPG | KO | CKO | SPL | STE |
| MIN | 0,002 | 0 | 0 | 0 | 0 | -0,939 | -0,516 |
| MAX | 17,199 | 17,2 | 17,2 | 17,181 | 17,2 | 20,553 | 16,408 |
| MOY | 10,372 | 10,129 | 10,129 | 10,384 | 10,129 | 10,529 | 9,962 |
| Écart-Type | 2,413 | 3,087 | 3,087 | 2,796 | 3,087 | 3,249 | 2,99 |
| Hétérogène | | | | | | | |
| MIN | 0 | 0 | 0 | -4,294 | 0 | -21,548 | -46,305 |
| MAX | 59,599 | 59,6 | 59,6 | 59,367 | 59,6 | 60,098 | 36,727 |
| MOY | 12,65 | 12,475 | 12,475 | 13,608 | 12,475 | 14,296 | 10,059 |
| Écart-Type | 10,572 | 14,453 | 14,453 | 12,482 | 14,453 | 15,176 | 14,174 |
| Locale | | | | | | | |
| | DIP | IPL | IPG | KO | CKO | SPL | STE |
| MIN | 0 | 0 | 0 | -0,537 | 0 | -9,535 | -0,916 |
| MAX | 13,998 | 14 | 14 | 13,978 | 14 | 16,239 | 14,813 |
| MOY | 1,858 | 1,519 | 1,519 | 1,949 | 1,519 | 1,888 | 2,261 |
| Écart-Type | 2,673 | 3,252 | 3,252 | 3,294 | 3,252 | 4,384 | 2,753 |
| Extrême | | | | | | | |
| | DIP | IPL | IPG | KO | CKO | SPL | STE |
| MIN | 0 | 0 | 0 | 0 | 0 | -2,814 | -0,987 |
| MAX | 4,098 | 4,1 | 4,1 | 0,879 | 4,1 | 6,59 | 0,668 |
| MOY | 0,263 | 0,22 | 0,22 | 0,305 | 0,22 | 0,375 | 0,207 |
| Écart-Type | 0,413 | 0,712 | 0,712 | 0,328 | 0,712 | 1,159 | 0,297 |

Annexe 7. Occupation du sol du bassin versant du lac Massawippi

Annexe 8. Classes de distance

Annexe 9. Les paramètres testés dans AVSWAT

| Paramètre | Min | Max | Définition |
|-----------|-------|------|--|
| SURLAG | 0 | 4000 | Temps de réponse dû au ruissellement de surface |
| CN2 | 30 | 98 | Numéro de courbe de ruissellement |
| EPCO | 0 | 1 | Facteur de compensation d'évaporation de la plante |
| ESCO | 0 | 1 | Facteur de compensation d'évaporation du sol |
| ALPHA_BF | 0,001 | 1 | Constante de récession de l'écoulement de base souterrain |
| LAT_TTIME | 0 | 180 | Temps de parcours des écoulements latéraux |
| GWQMN | 0,001 | 5000 | Hauteur d'eau minimale pour qu'il y ait un transfert d'eau de l'aquifère peu profond vers le cours d'eau |
| RCHRG_DP | 0,001 | 1 | Fraction de l'eau qui percole de la zone racinaire et qui atteint l'aquifère profond |
| SOL_AWC | 0,001 | 1 | capacité en eau disponible de la couche de sol considérée (mm H2O/mm sol) |
| GW_DELAY | 0,001 | 365 | Délai pour que l'eau souterraine qui transite par l'aquifère peu profond atteigne le cours d'eau (jours) |
| GW_REVAP | 0,02 | 0,02 | Coefficient permettant le transfert d'eau de l'aquifère vers les horizons sus-jacents et non saturés du sol. |
| REVAPMN | 0,001 | 500 | Hauteur d'eau minimale dans l'aquifère peu profond pour qu'il y ait un mouvement d'eau vers le sol ("revap") ou l'aquifère profond |
| CANMX | 0,001 | 10 | Quantité d'eau maximale pouvant être retenue par le couvert végétal quand il est complètement développé (mm H2O) |
| DDRAIN | 0 | 2000 | Profondeur du drainage souterrain (mm) |
| TDRAIN | 0 | 72 | Temps requis pour drainer le sol à la capacité au champ (h) |

| | | | |
|------------|---------|------|--|
| GDRAIN | 0 | 100 | Temps requis pour drainer le sol à la capacité au champ (h) |
| DEPIMP_BSN | 1000 | 7000 | Profondeur de la couche imperméable (mm) |
| USLE_K | 0 | 0,65 | Facteur d'érodabilité du sol (0,013 t.m ² .h/m ³ .t.cm) |
| HRU_SLP | 0,00001 | 0,6 | Pente de l'Unité de Réponse Hydrologique |
| SLSUBBSN | 10 | 150 | Longueur de pente moyenne du sous-bassin (m) |
| USLE_P | 0,1 | 1 | Facteur de pratique de conservation |
| USLE_C | 0,001 | 0,05 | Facteur de culture (végétation) et de gestion |
| SOL_Z | 1 | 5000 | Profondeur de la couche à partir de la surface du sol (mm) |
| SPCON | 0,001 | 0,01 | Paramètre linéaire pour calculer la quantité maximale de sédiments pouvant être entraînés durant le transport dans le canal. |
| CH_S2 | 0 | 1 | Pente moyenne du canal principal le long de la longueur du canal |
| CH_EROD | -0,05 | 1 | Facteur d'érodabilité du canal |
| CH_COV | -0,001 | 1 | Facteur de couverture végétale du canal |
| CH_N2 | 0,001 | 0,5 | Coefficient de Manning "n" du canal principal |
| CH_S1 | 0,0001 | 10 | Pente moyenne du canal tributaire le long de la longueur du canal |
| CH_N1 | 0,01 | 30 | Coefficient de Manning "n" des canaux tributaires |