

## TABLE DES MATIÈRES

	Page
INTRODUCTION .....	1
CHAPITRE 1 REVUE DE LITTÉRATURE .....	3
CHAPITRE 2 BASSIN VERSANT ÉTUDIÉ .....	9
2.1 Description générale .....	9
2.2 Types de sol .....	10
2.3 Données hydrométéorologiques historiques .....	10
CHAPITRE 3 MÉTHODOLOGIE DE LA RECHERCHE .....	13
3.1 Résumé de la recherche .....	13
3.2 Données utilisées .....	14
3.3 Routage des ruissellements du MRCC .....	18
3.3.1 Résultats du schéma de routage .....	21
3.3.2 Analyse du schéma de routage .....	24
3.4 Modélisation hydrologique .....	25
3.4.1 Modèles hydrologiques utilisés .....	26
3.4.2 Calage et validation .....	28
3.5 Effets de la densité du réseau d'observations sur la variabilité spatiale des données interpolées .....	34
3.6 Effets de la densité du réseau d'observations sur la modélisation hydrologique .....	34
3.7 Détermination du nombre de stations optimal et de leur emplacement .....	36
3.7.1 Description du front de Pareto .....	36
3.7.2 Évaluation de la méthode d'optimisation du Front de Pareto .....	38
3.7.3 Application à cette étude .....	40
3.7.4 Validation de la méthode sur différentes périodes .....	41
3.8 Hypothèses .....	42
CHAPITRE 4 RÉSULTATS .....	43
4.1 Calage et validation des modèles hydrologiques .....	43
4.2 Effets de l'interpolation lors de la réduction de la densité du réseau .....	44
4.3 Effets de la densité du réseau d'observations sur la modélisation hydrologique .....	47
4.3.1 Calage avec stations sélectionnées .....	48
4.4 Optimisation multi-objective avec paramètres fixes .....	50
4.5 Optimisation multi-objective avec calage des paramètres .....	53
4.5.1 Résultats des essais aléatoires pour déterminer l'effet du hasard sur la performance du modèle .....	53
4.5.2 Front de Pareto sur périodes de 10 ans .....	55
4.6 Emplacement des stations .....	56
4.6.1 Fréquence d'utilisation des stations en optimisation multi-objective pour HSAMI .....	56

4.6.2	Utilisation des meilleures stations candidates selon la somme des distributions.....	59
4.6.3	Validation des résultats.....	61
CHAPITRE 5	ANALYSE DES RÉSULTATS ET DISCUSSION.....	69
5.1	Interpolation et densité des stations sur une grande région.....	69
5.2	Méthode manuelle.....	70
5.3	Optimisation multi-objective.....	72
5.4	Robustesse de la méthode.....	73
5.5	Différences et similitudes entre les modèles.....	75
5.6	Résultats globaux.....	75
5.7	Explications possibles de la bonne performance des modèles avec une faible densité de stations.....	77
5.7.1	Explications générales.....	77
5.7.2	Qualité du routage.....	78
5.7.3	Bruit et faibles différences entre les stations.....	79
5.8	Réponse à la problématique initiale.....	81
5.9	Vers une nouvelle approche pour la recherche en hydrologie.....	82
5.10	Estimé des retombées.....	83
CONCLUSION	.....	85
ANNEXE I	STATIONS MÉTÉOROLOGIQUES.....	87
ANNEXE II	PARAMÈTRES DES MODÈLES HYDROLOGIQUES.....	89
ANNEXE III	MESURES D'EFFICACITÉ EN CALAGE POUR DES RÉSEAUX ALÉATOIRES DE DENSITÉ VARIABLE.....	93
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....		97

## LISTE DES TABLEAUX

	Page
Tableau 3.1	Solutions Pareto-optimales pour la figure 3.2.....38
Tableau 4.1	Mesure des critères de performance des modèles HSAMI, HMETS et Hydrotel en calage et en validation.....43
Tableau 4.2	P-value des tests de Kruskal-Wallis sur la performance de HSAMI en calage avec des stations aléatoires comparativement à des stations identifiées préalablement .....64
Tableau 4.3	P-value des tests de Kruskal-Wallis sur la performance de HSAMI en validation avec des stations aléatoires comparativement à des stations identifiées préalablement .....67



## LISTE DES FIGURES

		Page
Figure 2.1	Emplacement du bassin versant de la Rivière Tournustouc.....	10
Figure 2.2	Emplacement des stations météorologiques existantes sur le bassin versant.....	11
Figure 3.1	Distribution des 1369 stations virtuelles produites par le MRCC15 et disponibles sur la région étudiée.....	15
Figure 3.2	Emplacement des 39 stations virtuelles sur le bassin versant de la Rivière Tournustouc.....	16
Figure 3.3	Moyenne journalière des variables simulées par le MRCC sur le bassin versant de la Rivière Tournustouc.....	17
Figure 3.4	Profondeur du socle rocheux .....	18
Figure 3.5	Division des 39 stations en 3 zones de routage distinctes .....	20
Figure 3.6	Comparaison entre l'hydrogramme interannuel moyen observé et l'hydrogramme reconstitué à partir du schéma de routage des écoulements simulés par le MRCC15.....	22
Figure 3.7	Écoulement total moyen (mm/jour) sur le bassin versant de la rivière Tournustouc pour les années 1961-1963.....	22
Figure 3.8	Comparaison entre l'hydrogramme reconstitué (MRCC) et l'hydrogramme simulé par HSAMI pour la période 1976-2000 et utilisant les mêmes données météorologiques .....	23
Figure 3.9	Débits observés et simulés par HSAMI avec des données station pour la période 1976-2000 .....	24
Figure 3.10	Front de Pareto pour une solution binaire vérifiable .....	37
Figure 3.11	Front de Pareto pour le problème multi-objectif ZDT-5 .....	40
Figure 4.1	Carte des précipitations annuelles moyennes .....	44
Figure 4.2	Résultats des essais de réduction de densité du réseau et erreurs produites par le schéma d'interpolation sur la précipitation moyenne annuelle.....	45

Figure 4.3	Moyenne de la précipitation pour des combinaisons aléatoirement générées contenant des nombres de stations progressivement en augmentation.....	47
Figure 4.4	Stations sélectionnées pour les essais manuels sur le bassin versant de la Rivière Toulnostouc.....	48
Figure 4.5	Critère de Nash-Sutcliffe en calage et en validation lorsque différentes combinaisons de stations sont utilisées.....	49
Figure 4.6	Optimisation sans recalage des stations utilisées dans HSAMI.....	51
Figure 4.7	Optimisation sans recalage des stations utilisées dans HMETS.....	51
Figure 4.8	Optimisation sans recalage des stations utilisées dans Hydrotel.....	52
Figure 4.9	Performance du modèle HSAMI en calage avec choix de stations aléatoires.....	54
Figure 4.10	Distribution des valeurs de Nash-Sutcliffe en fonction du nombre de stations choisies en calage pour HSAMI.....	55
Figure 4.11	Solutions finales de l'optimisateur multi-objectif pour 4 périodes de 10 années chacune avec le modèle HSAMI.....	56
Figure 4.12	Fréquence d'utilisation des stations dans les solutions finales de l'optimisateur multi-objectif pour 4 périodes de 10 ans dans le modèle HSAMI.....	57
Figure 4.13	Fréquence d'utilisation normalisée des stations dans les solutions finales de l'optimisateur multi-objectif pour 4 périodes de 10 ans dans le modèle HSAMI.....	58
Figure 4.14	Carte du bassin versant avec les couleurs associées aux régions où le potentiel de rendement hydrologique est supérieur.....	60
Figure 4.15	Précipitation moyenne journalière pour chaque décennie de 1961 à 2000.....	60
Figure 4.16	Fréquence d'occurrence des stations pour chaque décennie de 1961 à 2000.....	61
Figure 4.17	Boîte à moustaches représentant les combinaisons de 2 stations en calage.....	62
Figure 4.18	Boîte à moustaches représentant les combinaisons de 3 stations en calage.....	63

Figure 4.19	Boîte à moustaches représentant les combinaisons de 4 stations en calage .....	63
Figure 4.20	Boîte à moustaches représentant les combinaisons de 5 stations en calage .....	64
Figure 4.21	Boîte à moustaches représentant les combinaisons de 2 stations en validation.....	65
Figure 4.22	Boîte à moustaches représentant les combinaisons de 3 stations en validation.....	66
Figure 4.23	Boîte à moustaches représentant les combinaisons de 4 stations en validation.....	66
Figure 4.24	Boîte à moustaches représentant les combinaisons de 5 stations en validation.....	67
Figure 5.1	Débits reconstitués à partir des écoulements de surface et souterrains du MRCC lorsque le bassin est divisé en 3 zones et en 10 zones.....	79
Figure 5.2	Comparaison entre les stations sélectionnées avec et sans bruit ajouté aux données.....	80

Rapport-Gatit.com



## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AG	Algorithme Génétique
CGCM3	Canadian Global Climate Model, 3 <sup>ième</sup> Génération
CLASS	Canadian Land Surface Scheme
DE	Differential Evolution
GCM	Global Climate Model / General Circulation Model
MOGA	Multi-Objective Genetic Algorithm
MRC	Modèle Régional de Climat
MRCC	Modèle Régional Canadien du Climat
MRCC15	Modèle Régional Canadien du Climat à résolution de 15 kilomètres
OMM	Organisation Météorologique Mondiale
SCE-UA	Shuffled Complex Evolution – University of Arizona
SOPFEU	Société de Protection des Forêts contre le Feu
UHRH	Unité Hydrologique Relativement Homogène
WMO	World Meteorological Organisation



## INTRODUCTION

L'hydrologie est la science qui étudie le mouvement et la distribution de l'eau sur la Terre. Plus particulièrement, l'hydrologie étudie les différents aspects du cycle de l'eau. D'un point de vue d'ingénierie, elle permet de prédire le comportement de rivières et de réservoirs dans le temps. Les études hydrologiques sont utilisées pour dimensionner adéquatement la hauteur des barrages, les ponts et les ouvrages de contrôle des niveaux d'eau pour ne nommer que ces exemples. Une bonne étude du cycle de l'eau permet de déterminer la probabilité qu'une inondation ou qu'une crue exceptionnelle survienne et ainsi permettre de prendre des mesures de mitigation appropriées.

Pour parvenir à prédire le comportement d'une rivière, les hydrologues utilisent un outil indispensable : le modèle hydrologique. Il s'agit d'un logiciel permettant de prendre en compte les conditions météorologiques d'une région sur une certaine période de temps et de simuler les débits dans les rivières. Plusieurs types de modèles existent, allant de modèles très simples (conceptuels et globaux) à très complexes (physiques et distribués). D'une manière ou d'une autre, tous les modèles requièrent au strict minimum des données météorologiques et la superficie du bassin versant étudié. L'aire du bassin versant est généralement simple à trouver à l'aide d'outils d'information géographiques. Les données météorologiques sont cependant plus difficiles à obtenir, et ce, pour plusieurs raisons.

Premièrement, le coût d'une station météorologique est élevé, tant pour l'achat que pour l'entretien. Puis, la base de données historiques disponible est très limitée, ce qui veut dire que les études doivent se baser sur des informations restreintes. À titre informatif, les stations météorologiques au Canada ont enregistré en moyenne 14.5 années de données. De plus, la grande majorité d'entre elles n'est plus active aujourd'hui. Ensuite, les stations de mesure en place ne peuvent mesurer avec acuité certaines variables météorologiques; il peut y avoir des biais dans les mesures de précipitation pluvieuse ou neigeuse raison du vent, par exemple. Enfin, les stations mesurent les conditions météorologiques ponctuelles. Si un orage déverse de grandes quantités d'eau entre deux stations, ces dernières pourraient ne mesurer aucune ou

que très peu de précipitations. Les modèles hydrologiques n'auront donc pas assez d'information pour bien simuler les apports en rivière.

Il est donc clair que le réseau de stations météorologiques est un instrument indispensable à bon nombre de chercheurs, de gestionnaires de systèmes hydriques et de groupes scientifiques de toutes sortes. Il permet de mesurer et d'archiver les variables météorologiques en des lieux précis dans le temps. Cette base de données ponctuelles peut être interpolée pour obtenir des valeurs entre deux stations, mais ceci engendre nécessairement des erreurs d'approximation. La qualité des interpolations dépend donc de la distance séparant les stations météorologiques. Puisque les modèles hydrologiques dépendent de la qualité des observations, ils devraient être plus performants si le réseau d'observations contient plus de stations météorologiques.

Le but de cette recherche est de déterminer l'effet de la densité spatiale des stations météorologiques sur la réponse de 3 modèles hydrologiques différents et de déterminer s'il y a une densité critique qui serait à la fois économique et représentative de la réalité. De plus, une méthode pour déterminer l'emplacement et le nombre optimal de stations météorologiques sera évaluée.

Ce mémoire comprend 6 sections. D'abord, cette introduction est suivie par une brève revue de littérature. Puis, le bassin versant étudié, les données utilisées et les modèles hydrologiques choisis et leurs particularités seront décrits. La méthodologie utilisée sera ensuite expliquée, et cette section sera suivie de la présentation des résultats, de leur analyse et de recommandations.

## CHAPITRE 1

### REVUE DE LITTÉRATURE

Lorsqu'une zone géographique doit être bien représentée au niveau de sa météorologie, un réseau d'observation doit être mis en place. Dans le cas où un réseau existe déjà, il peut être augmenté par l'ajout de stations. L'optimisation du processus d'augmentation de la densité d'un réseau d'observation peut se faire de plusieurs manières. Selon le « WMO Guide to Hydrological Practices 2008 », la densité de stations minimale pour un bassin versant devrait être d'une station par environ 250 à 575 km<sup>2</sup> pour des régions montagneuses et/ou vallonnées. La densité actuelle des stations sur le bassin étudié, celui de la rivière Toulnostouc, est d'environ une station par 2000 à 8000 km<sup>2</sup>, ce qui est largement insuffisant selon ce standard. De plus, le WMO propose une densité moyenne, mais ne spécifie pas de méthode pour déterminer l'emplacement des stations lors de l'augmentation du réseau. Ce problème n'est pas limité au Québec. En effet, la densité moyenne des stations en opération pour le Canada en entier est d'environ 7650 km<sup>2</sup> par station, soit 1190 stations pour une superficie au sol d'environ 9 100 000 km<sup>2</sup>. En tenant compte de la distribution des stations qui favorise les zones plus peuplées, il est évident que leur densité est très faible dans les régions nordiques. Les acteurs qui nécessitent des bases de données climatiques fiables et précises doivent donc investir dans des stations météorologiques pour augmenter la fiabilité de leurs opérations. Parmi ces acteurs, Hydro Québec, le Gouvernement du Québec (via la Société de Protection contre les Feux de Forêt (SOPFEU)), le producteur d'aluminium Rio Tinto Alcan et Environnement Canada ont des réseaux d'observation. Au Québec, le programme de surveillance du climat du Centre d'Expertise Hydrique du Québec (CEHQ) est responsable du réseau de stations météorologiques.

Les méthodes d'augmentation du réseau d'observations ont été le sujet de nombreuses publications dans les dernières années. La méthode habituelle consiste à utiliser des méthodes géostatistiques pour déterminer les zones à haute variance afin de les réduire à l'aide de l'ajout de stations ou de données radar (Bastin et al., 1984; Creutin et al., 1988;

Amani et Lebel, 1997; Pardo-Iguzquiza, 1998; St-Hilaire et al., 2003). Ces approches sont basées sur l'optimisation de l'emplacement des stations pour réduire la variance spatiale globale des variables climatiques sur le bassin versant. Leur performance a été démontrée dans maintes situations, notamment pour mieux définir les orages localisés et de courte durée (Cheng et al., 2008). L'emplacement optimal des stations peut dépendre du type d'averse étudié, mais les méthodes de réduction de la variance sont également appropriées pour des averses de longue durée ou pour l'estimation des moyennes de précipitation annuelles (St-Hilaire et al., 2003). Cependant, plusieurs paramètres doivent être estimés lors des interpolations diverses, de la fabrication des variogrammes et des corrélogrammes. De plus, ces études basent leurs résultats sur des cas réels. Par exemple, l'ajout et le retrait de stations parmi un lot de stations existantes afin de voir l'effet de la densité du réseau sur la performance en modélisation hydrologique est une approche proposée. Cette approche est limitée en raison de sa dépendance à la qualité des données observées ainsi qu'à l'emplacement préétabli des stations, mais elle est la seule manière de montrer que la réduction de variance contribue à l'amélioration de la performance hydrologique du réseau. Une meilleure approche serait d'utiliser des nombres de stations aléatoires et de n'avoir aucun biais dans les données de manière à vérifier l'effet de la densité seulement. Ceci n'empêche pas les méthodes basées sur la variance de contribuer à l'amélioration de la performance des réseaux, mais il y a clairement place à amélioration (Jedermann et al., 2011).

L'approche inverse a aussi été testée par Schneebeli et Lancaster (2004). Leur approche a été de construire un atlas de toutes les précipitations de neige passées et mesurées par 107 stations manuelles sur un bassin versant en Suisse. Ils ont ensuite calculé la probabilité de précipitation par région et ont assigné un rang d'importance à chacune des stations de mesure. Cette méthode est prometteuse pour ce cas précis, mais dans la plupart des cas réels, il est impossible de procéder ainsi. La raison est simple : les données requises pour créer l'atlas de probabilités n'existent tout simplement pas à une résolution spatiale adéquate au Canada. Il est donc peu utile de considérer cette méthode dans la prise de décision sur l'emplacement des stations futures.

Une autre méthode explorée pour contourner le problème des réseaux d'observations est la construction de bassins expérimentaux (Osborn et Keppel, 1965; Osborn et al, 1972). Ces bassins, normalement de petite taille (généralement entre 30 et 1000 km<sup>2</sup>), sont parsemés de stations météorologiques et de pluviomètres pour obtenir le plus d'information possible (Amatya et Trettin, 2007). La densité de stations atteint parfois 1 station par km<sup>2</sup>. Ces bassins expérimentaux, très coûteux, sont souvent utilisés pour déterminer l'effet d'autres paramètres que la précipitation seule. Par exemple, certains bassins étudient l'effet de l'urbanisation sur le ruissellement de surface, de l'agriculture sur l'érosion des sols, etc. (Burns, 1965; Schneider-Vieira, 1993; Cafferata et Spittler, 1995). Cependant, aucune étude n'a tenté d'utiliser l'information disponible des bassins expérimentaux pour tenter de déterminer le nombre minimal de stations requises pour la modélisation hydrologique. L'approche des bassins expérimentaux est intéressante en raison de la quantité d'information qu'il est possible de retirer des bases de données, mais l'incertitude inhérente aux mesures et aux caractéristiques du bassin rend la technique peu applicable à d'autres bassins versants. Pour ce faire, il faudrait installer des stations temporaires sur les bassins à jauger, mais le coût élevé de cette approche la rend aussi peu attrayante que l'ajout aléatoire de stations dans le réseau.

La présente étude comble un vide dans la littérature au sujet de la densité optimale des stations météorologiques. Parmi toutes les sources consultées, aucune ne faisait état d'une méthode permettant de déterminer le nombre de stations requis ainsi que leur emplacement pour obtenir une performance hydrologique adéquate. Par exemple, une entreprise désirant caractériser un bassin versant non-jaugé pour des fins de production hydro-électrique devrait-elle installer 3, 5 ou 10 stations? Et à quel endroit sur le bassin ces stations devraient-elles être installées pour optimiser le réseau? Les études actuelles dépendent du fait que certaines stations existent déjà pour lancer l'analyse de la variance. Lorsqu'aucune information n'est disponible à priori ou que la distance entre le bassin et les stations les plus près est trop grande, un certain mélange de hasard, d'intuition et de considérations logistiques est le seul guide pour les hydrologues. Ceci a été très bien résumé par Husain (1989):

« It has not, so far, been possible to define the optimum level of hydrologic information required for planning, design, and development of specific projects in a region, due to difficulties in developing the cost and benefit functions of hydrologic information. It is, therefore, difficult to attain an optimum balance between, on one hand, the economic risk arising from inadequate information and, on the other hand, the cost of a hydrologic network capable of transmitting the required information. Unless techniques to evaluate such a balance are developed, the network design methods cited in the literature cannot be universally applied. »

Husain a lui-même proposé une méthode basée sur le concept de l'entropie pour tenter de déterminer comment augmenter la densité d'un réseau d'observations de la façon la plus optimale possible. Il a utilisé des stations sur un bassin versant qu'il a ensuite séparées par régions. Puis, il tente de trouver des zones sur le bassin versant qui bénéficieraient le plus d'une station météorologique afin que l'incertitude soit minimisée en modélisation. Le concept d'entropie est utilisé afin d'assurer que le transfert d'information d'une station à une autre soit maximisé (Husain, 1989).

La présente étude explore les nouvelles possibilités qu'offre l'utilisation du MRCC pour guider les hydrologues face à cette situation problématique.

Toutes les études sur le sujet sont d'accord sur un point : augmenter la quantité de stations est toujours une bonne option lorsque le réseau est densifié de manière non-optimale. Malheureusement, dans le contexte économique actuel, les ressources financières se font rares et plusieurs stations sont retirées en raison des coûts d'entretien élevés (Burn et Goulter, 1991; Ouarda, 1996). Ces stations sont retirées de manière à minimiser la perte d'information le plus possible.

Dans le cadre du problème abordé dans ce mémoire, seules deux variables météorologiques sont utilisées : la température et la précipitation. La précipitation est un phénomène local et d'intensité et de distribution spatiale variable. De plus, il s'agit de la variable qui a le plus d'impact sur la performance des modèles hydrologiques. La température joue un rôle important, mais mis-à-part dans les régions montagneuses où l'altitude peut changer brusquement, sa variabilité spatiale est faible. La densité de stations de mesure de

température ne doit donc pas nécessairement être aussi élevée que pour les points de mesure de la précipitation. De plus, les stations météorologiques sont normalement équipées à la fois de pluviomètres et de thermomètres. Donc, les stations sont installées aux endroits où la précipitation doit être mesurée étant donné qu'il s'agit de la variable dominante.

De nombreuses études sont en train de paver la voie vers une nouvelle méthode de collecte de données de précipitation, soit la télédétection par radar. Cette approche semble prometteuse, mais quelques difficultés persistent toujours. Notamment, il est difficile de déterminer le taux réel de précipitation en raison de la difficulté qu'a le radar à différencier les différents types et le diamètre des particules de précipitation (Hunter, 2009; Ramli et Tahir, 2011). L'intégration de ces données dans des modèles hydrologiques est également un aspect qui reste à améliorer. De plus, certaines approches combinées utilisent des stations météorologiques en plus de la détection radar pour tenter d'améliorer la qualité des données fournies aux modèles hydrologiques. Ces approches semblent démontrer du potentiel, mais elles sont encore au stade de développement (Krajewski et Smith, 2002).

Une autre méthode de plus en plus utilisée en hydrologie consiste à utiliser les données provenant de modèles climatiques pour effectuer des études d'impact. Cependant, jusqu'à présent, aucune étude n'a utilisé ces Modèles Régionaux de Climat (MRC) pour optimiser l'emplacement de stations météorologiques. Tandis qu'il est fréquent d'utiliser des points de grille de MRC pour les fournir directement dans des modèles hydrologiques lors d'études de changements climatiques, la résolution spatiale des MRC était, jusqu'à tout récemment, trop faible pour effectuer des études de météorologie locale. Cependant, la nouvelle version du Modèle Régional de Climat Canadien (MRCC), un MRC développé par l'Université du Québec à Montréal et le consortium Ouranos sur les changements climatiques, a une résolution de 15 km, ce qui signifie que la représentativité spatiale est beaucoup plus adaptée pour effectuer des études à l'échelle du bassin versant que les versions précédentes (Brissette et al., 2011). Cette utilisation des MRC devrait donc permettre de trouver des solutions novatrices aux problèmes de longue date auxquels font face les hydrologues.



## CHAPITRE 2

### BASSIN VERSANT ÉTUDIÉ

Cette section présente l'emplacement et les particularités du bassin versant utilisé pour cette étude ainsi que les sources de données observées disponibles. Les types de sol présents sur le bassin sont également discutés.

#### 2.1 Description générale

Le bassin versant choisi, celui de la rivière Tournestouc, est un sous-bassin de la rivière Manicouagan. Il couvre une superficie de 8109 km<sup>2</sup> et contient un réservoir de 235 km<sup>2</sup> : le lac Ste-Anne, qui alimente une centrale hydroélectrique faisant office d'exutoire pour le bassin. Situé sur la Côte-Nord du Québec, son régime d'écoulement est caractérisé en grande partie par la crue printanière.

Le bassin versant se situe entre les latitudes 49.85° et 51.71° et les longitudes -68.20° et -67.18°. Prenant sa source dans les monts Groulx, il est borné par les rivières Sainte-Marguerite à l'Est et la rivière Manicouagan à l'Ouest. Le bassin Tournestouc est un bassin pratiquement inhabité où la forêt est la principale couverture du sol. Le relief est vallonné, avec une différence d'élévation maximale d'environ 750 mètres entre une crête de montagne au Nord et l'exutoire au Sud.

La Figure 2.1 illustre l'emplacement du bassin ainsi que son relief général.

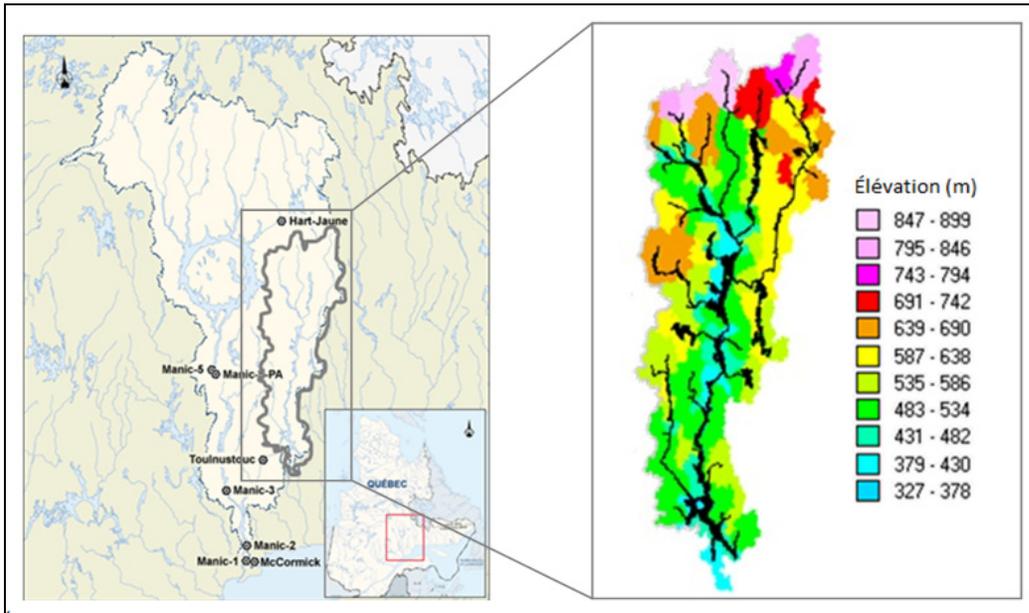


Figure 2.1 Emplacement du bassin versant de la Rivière Toulnostouc

## 2.2 Types de sol

Le sol de la région est généralement mal caractérisé en raison de la faible représentativité des forages géotechniques qui y ont été menés (SIH, service d'information hydrogéologique). Cependant, les essais effectués montrent qu'il s'agit principalement de loam limoneux à loam argileux. L'uniformité géologique apparente sur le bassin permet de généraliser ces quelques essais de terrain à la grandeur du bassin pour les besoins de l'étude.

## 2.3 Données hydrométéorologiques historiques

Il y a 5 stations météorologiques opérées par Environnement Canada sur le bassin, dont quatre sont placées sur le long de sa frontière Est. L'autre station est à l'extrémité Sud du bassin. L'emplacement de ces stations est illustré à la Figure 2.2. Il existe 3 stations à moins de 50 kilomètres de la frontière du bassin, mais celles-ci ont seulement quelques valeurs disponibles datant d'il y a 50 ans. Il y a donc une sous-représentation spatiale de la météorologie au niveau du bassin Toulnostouc.

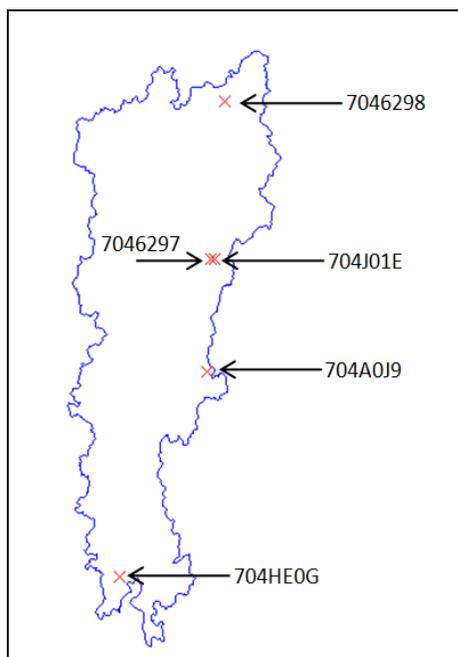


Figure 2.2 Emplacement des stations météorologiques existantes sur le bassin versant

Ensemble, les cinq stations couvrent les années 1963-1964 et 1973-1980. De plus, il y a des données manquantes dans ces séries. Par exemple, pour quatre des stations, il n'y a des données que pour les mois de juin à septembre. Pour l'autre station, il existe en plus des données estivales des données pour un mois d'hiver par année pendant deux ans. Les particularités des stations existantes sont disponibles à l'Annexe I, Tableau-A I-1. Il y a donc un manque de données météorologiques disponibles pour une utilisation en modélisation hydrologique. Pour contourner ce problème, la base de données traitée et corrigée du groupe production d'Hydro-Québec a été utilisée. Cette base de données est simplement une interpolation des stations d'Environnement Canada ainsi que des quelques stations de la SOPFEU et d'Hydro-Québec dans la région. Cette base de données journalières modifiée couvre les années 1950 à 2009 de manière complète, mais est toujours basée sur les observations des stations avoisinantes. Pour créer la base de données, une interpolation a d'abord été utilisée entre les stations les plus près du bassin (malgré leur éloignement) pour les années 1950 à 1960. Puis les polygones de Thiessen ont été utilisés pour les années 1961

à 2003 et les dernières années, de 2003 à 2009, ont été interpolées par krigeage. Les informations interpolées ne peuvent donc pas être meilleures que les données sous-jacentes.

Les débits à l'exutoire de la rivière sont disponibles pour les années 1966-2008. Ils sont disponibles dans la banque de données d'Hydro-Québec. Cette base de données a été créée par analyse des données mesurées à la station de mesure 02TC002 - Tournustouc et par interprétation des niveaux des réservoirs. Elle est utilisée dans un contexte opérationnel par Hydro-Québec. Il s'agit donc d'une base de données reconstituée comprenant des mesures réelles, des apports calculés par bilan et de modélisations.

## CHAPITRE 3

### MÉTHODOLOGIE DE LA RECHERCHE

Cette section détaille la méthodologie utilisée dans cette étude. Une brève introduction à la méthode est présentée, puis les hypothèses de travail sont énumérées. Les données utilisées et la méthodologie relative aux modèles hydrologiques seront également abordées. Chaque sous-problème de la recherche y est également détaillé.

#### 3.1 Résumé de la recherche

Ce projet de recherche nécessite une valeur de référence idéale à laquelle il serait possible de comparer les différentes simulations hydrologiques lorsque la densité du réseau d'observations est réduite. Cette référence devrait idéalement simuler exactement les différents processus hydrométéorologiques sur le bassin. Étant convenu que ce niveau de connaissances n'est pas atteignable à présent dans le monde réel, un monde virtuel a été utilisé. Il s'agit du Modèle Régional Canadien du Climat (MRCC) de quatrième génération, version 4.2.3 (Caya et Laprise, 1999; Music et Caya, 2007; de Elia et al, 2008; Music et al., 2009). Le MRCC est un modèle de tous les processus physiques qui décrivent le climat sur une région donnée. Piloté par les réanalyses ERA-40, le système de calcul du MRCC est basé sur une grille de 15 km de résolution sur laquelle nous assignerons une station virtuelle à chaque nœud. Le modèle de surface CLASS 2.7 contrôle les transferts entre le sol et l'atmosphère (Verseghy, 1991). Puisque toutes les variables météorologiques suivent des lois physiques, il n'y a pas de discordance entre les variables ni de données manquantes dans ce monde virtuel, ce qui donne un niveau de connaissances inégalé dans le monde réel. Cependant, il existe des biais entre le climat simulé par le MRC et le climat réel.

La suite de la méthodologie requiert de transformer les ruissellements de surface calculés par le MRCC en débits de rivière. Cette étape de routage est nécessaire pour produire une valeur réaliste d'apports en rivière à laquelle se compareront les simulations hydrologiques. Le débit MRCC agira donc à titre de valeur observée lors des simulations.

L'étape suivante a été de déterminer l'effet de la densité spatiale sur les processus d'interpolation de données aux endroits où il n'y a pas de station météorologique. Ceci permet de déterminer s'il y a une variation notable entre la performance des modèles d'interpolation et la réalité.

Puis, le réseau d'observation sur le bassin Tournoustouc a été graduellement dépourvu de stations, selon une méthode semi-aléatoire. Pour chaque nouveau réseau produit, les trois modèles hydrologiques ont été calibrés et leur performance a été notée sur une période de validation. Ce faisant, une courbe de performance en fonction de la densité du réseau a pu être établie. Le même type de comparaison pourrait être fait pour d'autres variables que le débit, mais cela n'a pas été traité dans ce travail.

Finalement, une calibration multi-objective du modèle HSAMI a été effectuée afin de déterminer le nombre de stations optimal ainsi que leurs positions respectives sur le bassin versant.

### **3.2 Données utilisées**

Les données utilisées pour cette étude sont toutes tirées du MRCC 15 km. Les séries temporelles du MRCC ont été générées et fournies par l'Équipe Simulations climatiques d'Ouranos. Les périodes simulées par le MRCC et qui ont été utilisées sont du 1<sup>er</sup> janvier 1961 au 31 décembre 2000. Cette période de 40 ans ne contient aucune valeur manquante sur l'ensemble des points de calcul. Sur le bassin versant de Tournoustouc, il y a 39 stations virtuelles. La base de données complète qui a servi aux calculs de performance des méthodes d'interpolation contient 1369 stations virtuelles, soit une grille carrée de 37X37 nœuds, couvrant la totalité de la région du bassin versant de la Rivière Manicouagan. Le réseau de points de calcul utilisés comme stations virtuelles est présenté à la Figure 3.1.

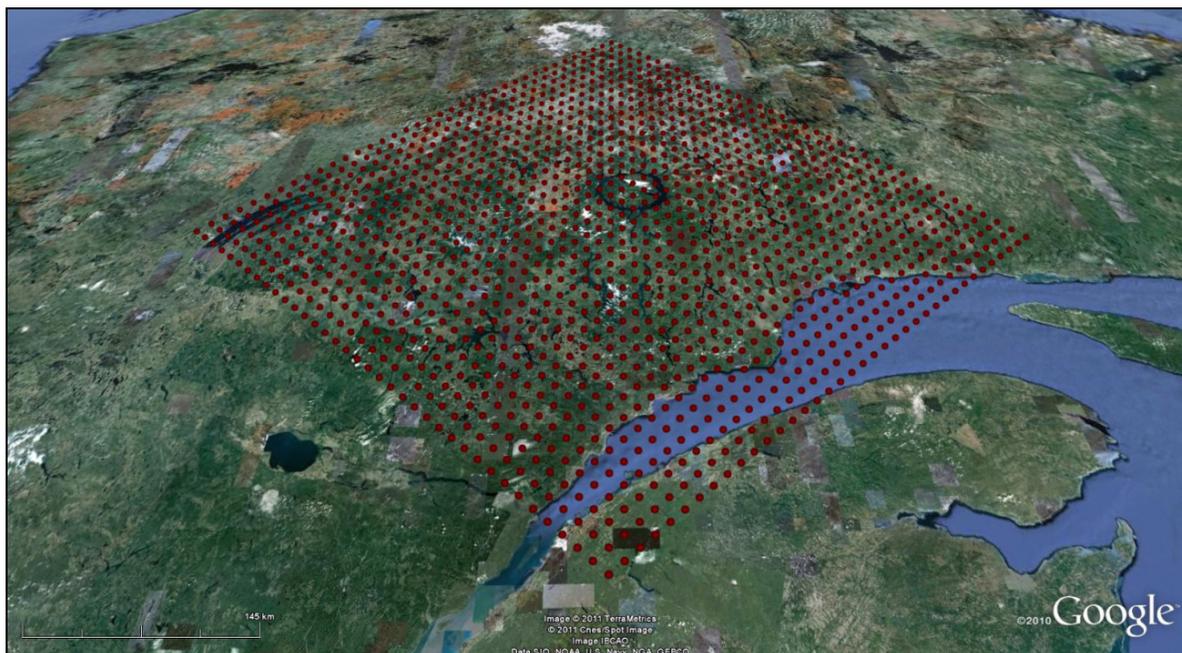


Figure 3.1 Distribution des 1369 stations virtuelles produites par le MRCC15 et disponibles sur la région étudiée

Les 1369 points de grille sont répartis uniformément sur une grille de 15 km de résolution. La grille est d'une taille assez grande pour couvrir la région du bassin de la rivière Manicouagan. Cette grille a été utilisée pour étudier l'effet de la densité de stations lors de l'interpolation des données manquantes de la section 3.5.

Pour les autres parties de l'étude, seules les stations virtuelles situées à l'intérieur du bassin versant ont été utilisées. Les 39 stations sont montrées à la Figure 3.2.

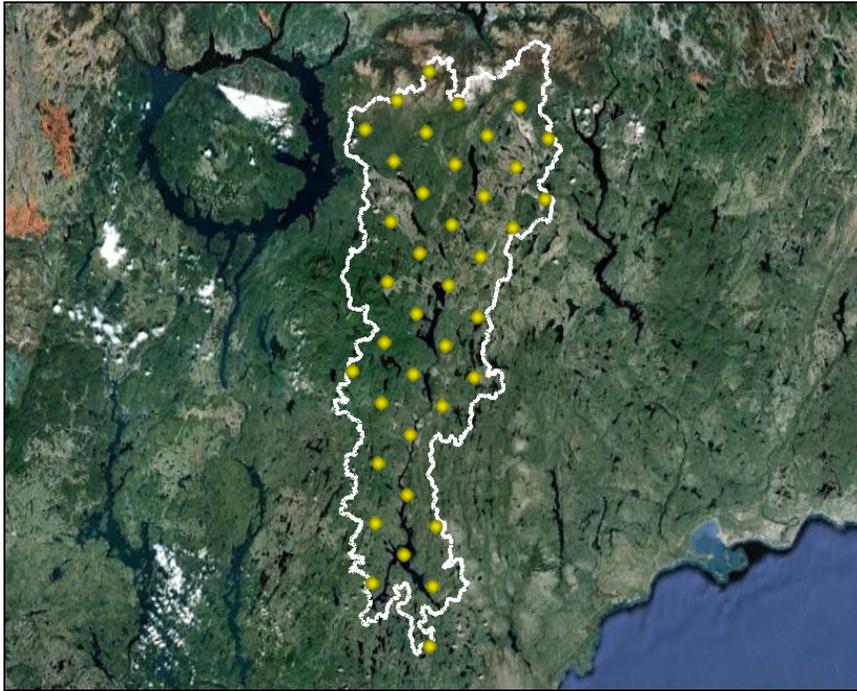


Figure 3.2 Emplacement des 39 stations virtuelles sur le bassin versant de la Rivière Toulouste

Les stations météorologiques standards mesurent généralement la température et la précipitation. Certaines d'entre elles mesurent en plus le vent, l'humidité relative et la pression de surface. Très rarement, une station peut également mesurer la radiation solaire et d'autres variables météorologiques.

La Figure 3.3 représente diverses variables simulées par le MRCC sur la période 1961 à 2000 inclusivement. Les valeurs représentent la moyenne annuelle pour chacune des variables affichées.

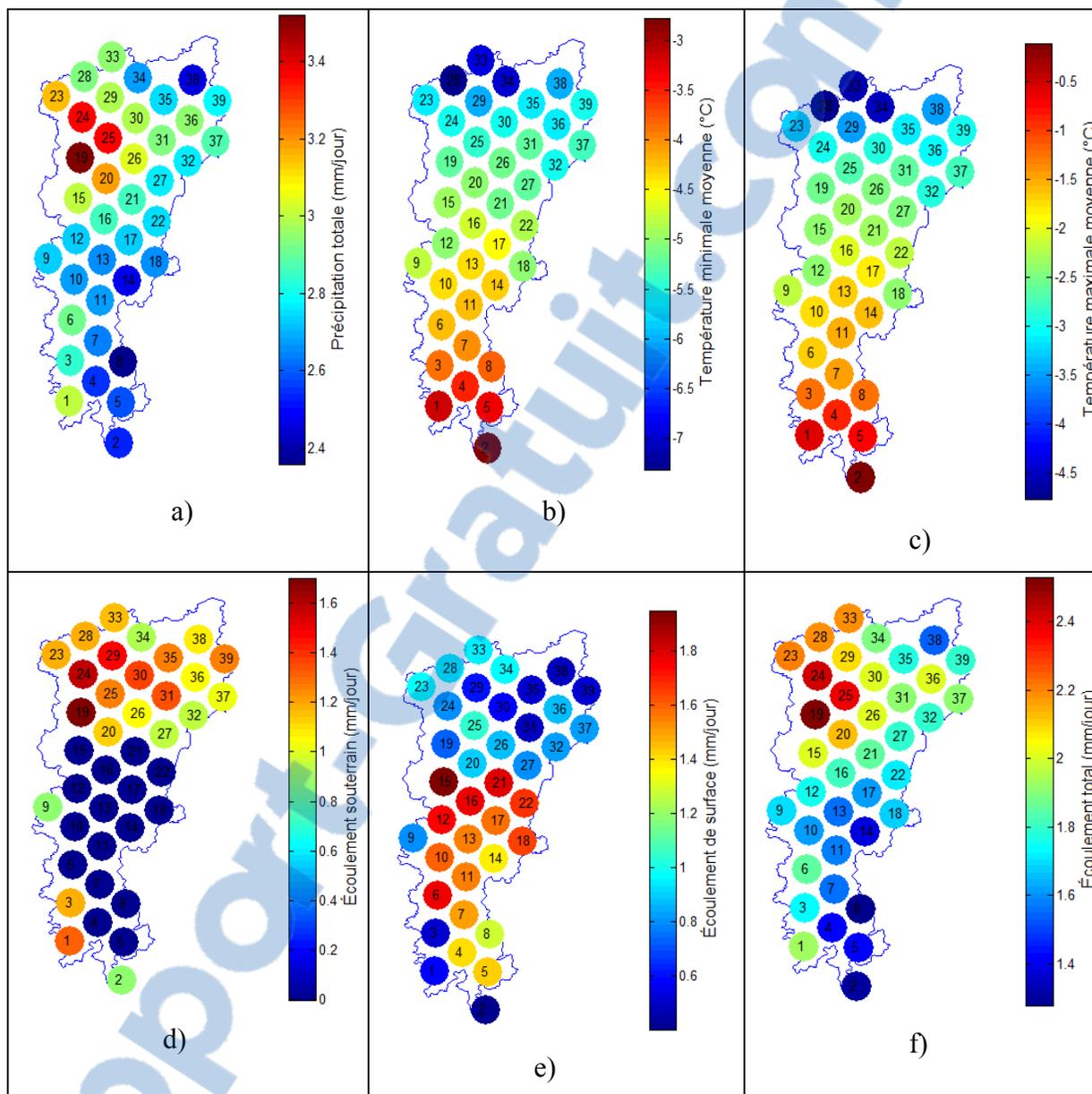


Figure 3.3 Moyenne journalière des variables simulées par le MRCC sur le bassin versant de la Rivière Toulnois. Les variables représentées sont a) la précipitation totale moyenne, b) la température minimum moyenne, c) la température maximum moyenne, d) l'écoulement souterrain, e) l'écoulement de surface et f) le ruissellement total

La Figure 3.4 représente la profondeur du socle rocheux aux différents nœuds de calcul tel qu'utilisé par le MRCC15. Cette épaisseur de sol est constante dans le temps.

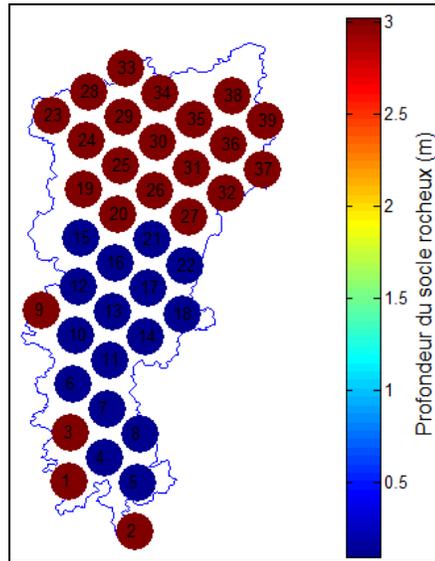


Figure 3.4 Profondeur du socle rocheux (m)

La Figure 3.4 démontre clairement qu'il y a une coupure nette entre le nord et le sud du bassin au niveau de l'épaisseur de la couche de sol. La section sud ne possède aucune épaisseur de sol, ce qui signifie que le MRCC ne simule que du roc à cet endroit. La qualité du routage peut donc être affectée par cette condition.

Les modèles hydrologiques nécessitent normalement les données les plus courantes, soit la précipitation et la température. Les données excédentaires peuvent parfois être incluses mais ne sont généralement pas obligatoires. Les données utilisées pour ce projet sont celles qui seraient normalement aisément disponibles en pratique, soient la température et la précipitation. Le potentiel d'adaptation des modèles hydrologiques sera du même coup évalué.

### 3.3 Routage des ruissellements du MRCC

Les modèles hydrologiques produisent des séries d'apports qui sont simulés en fonction des variables météorologiques, des caractéristiques du bassin et du paramétrage du modèle. Les paramètres optimaux sont obtenus en comparant les débits simulés par le modèle avec un

certain jeu de paramètres à une série d'apports observés. Les paramètres sont ensuite ajustés pour obtenir les apports simulés qui représentent le mieux les observations.

Le problème avec ce processus est qu'il est nécessaire d'avoir des apports observés liés aux données météorologiques utilisées. Puisque le MRCC simule des valeurs à chaque point de grille, il n'offre pas directement de données d'apports en rivière à un point donné. Dans la réalité, il existe des stations de mesure de débit aux endroits jugés utiles par les exploitants de systèmes hydriques.

Dans le cadre de la présente étude, des débits en rivière conséquents avec la météorologie sont nécessaires. Pour contourner cet obstacle, un schéma de routage a été mis en place. En plus des données météorologiques, le MRCC fournit des données de ruissellement de surface ainsi que de ruissellement souterrain pour chacune des tuiles.

Le bassin a été divisé en trois zones de même taille (13 points de grille par zone) en fonction de l'éloignement des points de grille par rapport à l'exutoire réel du bassin, tel que montré à la Figure 3.5. En d'autres mots, les 13 points les plus près de l'exutoire sont dans la zone 1 (rouge), les 13 points les plus éloignés dans la zone 3 (bleu) et les 13 points intermédiaires dans la zone 2 (vert).

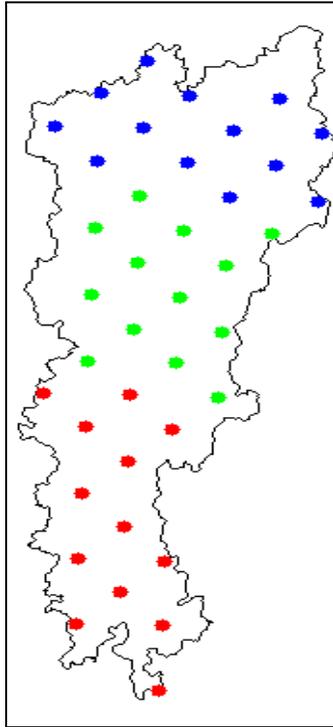


Figure 3.5 Division des 39 stations en 3 zones de routage distinctes

La méthode de l'hydrogramme unitaire a ensuite été utilisée pour faire progresser les ruissellements vers l'exutoire. Chaque zone est définie par deux hydrogrammes unitaires, soit un de surface et un souterrain. Les paramètres de ces six hydrogrammes ont ensuite été ajustés à l'aide d'un calage statistique en comparant l'apport reconstitué à l'hydrogramme observé à la station 02TC002 – Tournustouc en utilisant toutes les données disponibles (1966-2000). Ceci permet d'obtenir un schéma de routage qui produit un hydrogramme semblable à l'hydrogramme réel. Cette étape est nécessaire pour assurer que les modèles hydrologiques puissent travailler dans un environnement similaire à celui pour lequel ils ont été créés. Le routage permet de fournir un apport conséquent avec la météorologie puisque les écoulements de surface et souterrains simulés par le MRCC sont tous physiquement inter-reliés. De plus, le calage de l'hydrogramme reconstitué permet de maintenir un lien entre le monde virtuel et le monde réel. Ce faisant, il est possible d'utiliser les modèles du monde réel dans le monde virtuel.

Le calage des 7 paramètres du schéma de routage (6 pour les temps de pointe et un pour le facteur de forme unique des hydrogrammes) a été effectuée à l'aide de l'algorithme d'optimisation SCE-UA (Shuffled Complex Evolution – University of Arizona) en utilisant la valeur de Nash-Sutcliffe comme critère de performance (voir équation 3.2) (Duan et al., 1992). Dans ce cas, la valeur observée est le débit observé à la station Touloustouc tandis que la valeur modélisée est le débit reconstitué par routage des écoulements simulés par le MRCC.

Enfin, la performance de la méthode a été validée en effectuant un calage du modèle HSAMI sur les apports reconstitués en utilisant les données météorologiques du MRCC comme entrée. Une amélioration de la performance du modèle hydrologique dans le monde virtuel par rapport aux données observées dans le monde réel est attendue puisque cette méthode garantit qu'il ne devrait y avoir aucune erreur de lecture (des apports et de précipitation, par exemple), ni aucune donnée manquante.

### **3.3.1 Résultats du schéma de routage**

Les apports reconstitués à partir des écoulements souterrains et de surface du MRCC ont été calés sur les apports observés pour la durée complète disponible. Cependant, puisque la météorologie réelle n'est pas la même que celle simulée par le MRCC sur une période identique, il est impossible d'effectuer un calage de type jour par jour sur toute la période. La méthode choisie a été de caler sur la forme de l'hydrogramme annuel moyen. Ce faisant, les caractéristiques de la rivière Touloustouc et du bassin versant telles que le débit de pointe, le temps de pointe et le volume total sont bien représentées. L'hydrogramme reconstitué est donc similaire au cycle hydrologique moyen observé. Les deux hydrogrammes moyens après calage sont présentés à la Figure 3.6.

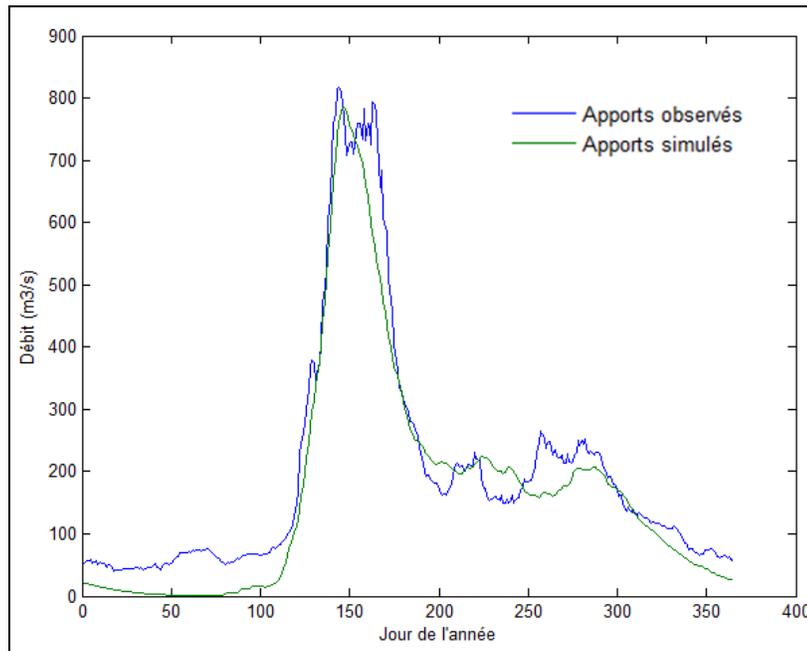


Figure 3.6 Comparaison entre l'hydrogramme interannuel moyen observé et l'hydrogramme reconstitué à partir du schéma de routage des écoulements simulés par le MRCC15

La Figure 3.6 montre que les débits hivernaux sont décroissants jusqu'à la crue, tandis que les apports observés sont relativement stables pendant l'hiver. La raison derrière cette observation est que le ruissellement simulé par le MRCC est décroissant et tombe à zéro pendant l'hiver, tel qu'illustré à la Figure 3.7.

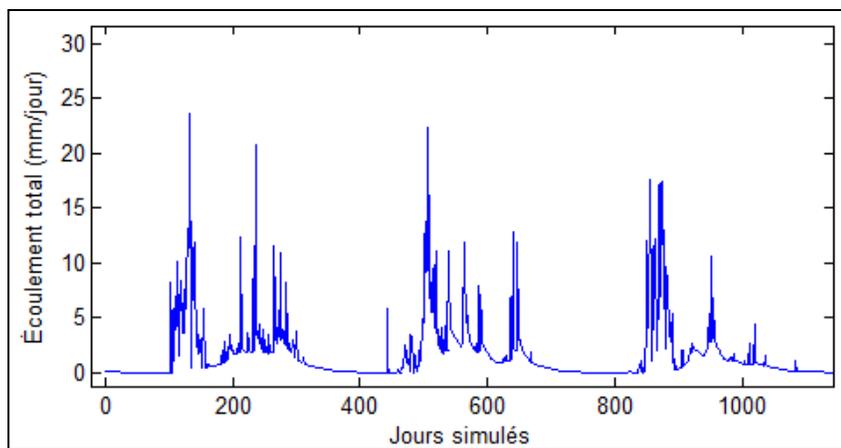


Figure 3.7 Écoulement total moyen (mm/jour) sur le bassin versant de la rivière Toulustouc pour les années 1961-1963

Le schéma de routage ne peut donc pas engendrer des débits sur la période hivernale puisque le MRCC ne fournit pas d'eau à écouler pendant cette période.

La validation du routage utilisé était nécessaire pour prouver le bien fondé de la méthode utilisée dans cette étude. L'hydrogramme simulé par HSAMI est comparé à l'hydrogramme reconstitué à la Figure 3.8.

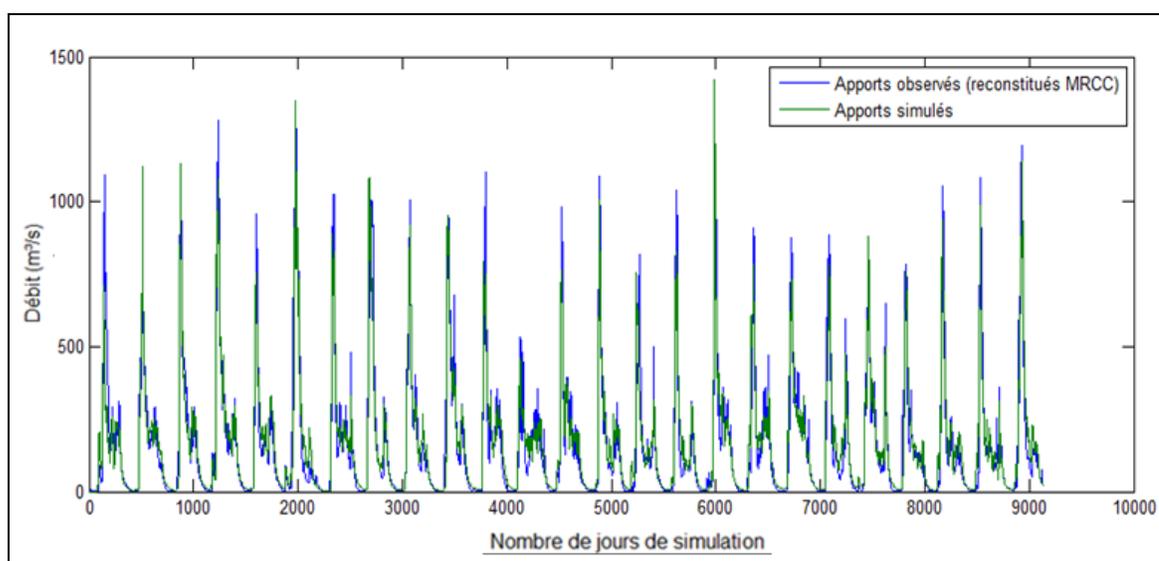


Figure 3.8 Comparaison entre l'hydrogramme reconstitué (MRCC) et l'hydrogramme simulé par HSAMI pour la période 1976-2000 et utilisant les mêmes données météorologiques

Lors du calage (période 1961-1975), le critère de Nash-Sutcliffe atteint 0.9223. Pour la période de validation (période 1976-2000), le critère tombe légèrement à 0.9122. L'écart faible entre les deux valeurs suggère que le modèle est bien calé.

Ces résultats se comparent très bien aux résultats obtenus lorsque l'exercice est mené avec les données du monde réel. La Figure 3.9 montre les hydrogrammes observé et simulé par HSAMI lors du calage du modèle.

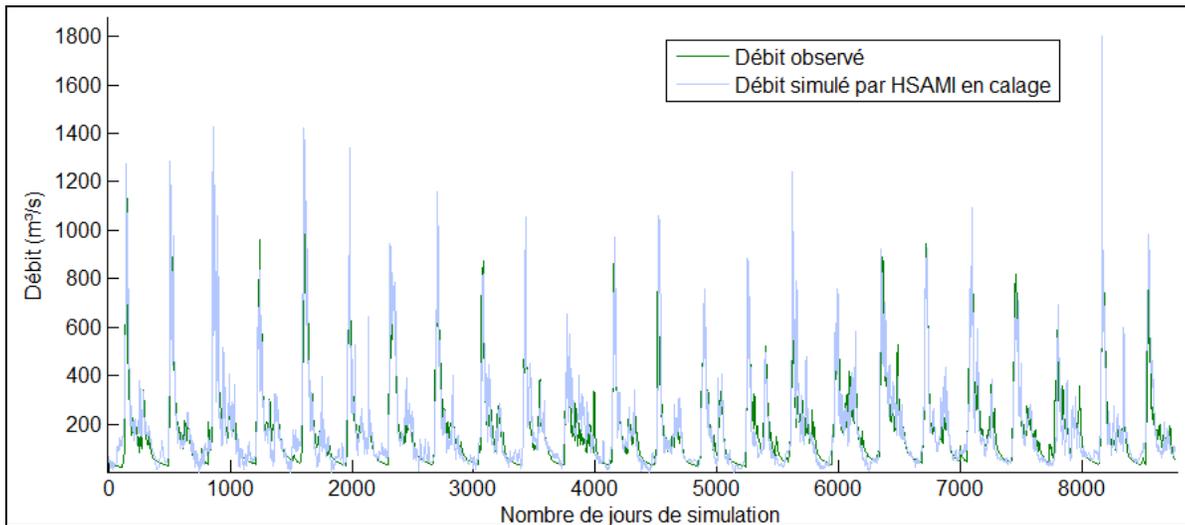


Figure 3.9 Débits observés et simulés par HSAMI avec des données station pour la période 1976-2000

La valeur de Nash-Sutcliffe de 0.847 en calage pour la période 1976-2000 est largement inférieure à la valeur de 0.9122 calculée lors de la validation pour la même période avec les données du MRCC. Les modèles peuvent donc définitivement utiliser les données du MRCC pour reproduire des apports en rivière.

### 3.3.2 Analyse du schéma de routage

La méthode utilisée pour reconstituer les apports est relativement simple. Les écoulements des trois zones ont été combinés à l'aide de la méthode de l'hydrogramme unitaire décrite à la section 3.3. Les paramètres des hydrogrammes unitaires ont été trouvés par calage, ce qui admet une incertitude dans les résultats. Il est important de rappeler que l'intention principale de cette étude n'est pas de déterminer quelles stations utiliser dans le monde réel, mais bien dans un monde virtuel. Le bassin de la Rivière Tounustouc a été sélectionné puisque les données étaient disponibles sur ce dernier et qu'il est semblable aux autres bassins de la région. La méthode de routage a donc peu d'importance dans la mesure où tout ce qui est requis pour l'étude est une base de données météorologiques complète avec un hydrogramme en rivière qui correspond aux fluctuations météorologiques sous-jacentes. Le fait que les débits soient reconstitués à partir de la même source qui a produit les données

météorologiques assure que cette cohérence existe. La méthode de routage est donc peu influente, tant qu'elle permet d'obtenir des données similaires à ce qui serait observé dans le monde réel. Ce raisonnement est derrière l'idée de caler les apports reconstitués sur les apports observés. Une autre méthode de routage aurait pu être utilisée, mais le fait qu'il y ait eu un calage sur les données observées implique nécessairement un certain nivellement quant aux différences des différentes approches possibles.

La figure 3.6 démontre clairement les similitudes entre les hydrogrammes reconstitué et observé. L'hydrogramme interannuel moyen présente des temps de pointe similaires et l'allure de la courbe est respectée. Le débit moyen, quoi qu'il soit semblable, n'a pas nécessairement besoin d'être identique puisque les volumes de précipitation ne sont pas les mêmes en entrée. Le ruissellement doit donc suivre les mêmes tendances mais pas nécessairement les mêmes quantités que les données observées.

Le routage est également fonction du type de sol en place. Dans le bassin versant virtuel, la moitié du territoire ne possède pas d'épaisseur de sol. Il va sans dire que les implications d'avoir le roc à la surface ont un impact sur les hydrogrammes simulés puisque le sol a un effet direct sur les écoulements souterrains et de surface. S'il n'y a pas de sol, le ruissellement sera entièrement de surface. Cependant, le fait que le schéma de routage soit basé sur plusieurs zones et que ses paramètres ont été calés fait en sorte que les débits reconstitués sont acceptables pour cette étude.

### **3.4 Modélisation hydrologique**

La modélisation hydrologique est l'utilisation d'un système (souvent mathématique) simulant la réponse d'un bassin versant à une série de facteurs météorologiques et physiques. Les modèles qui sont à l'étude font partie de deux catégories : global et distribué.

Un modèle global prend une moyenne des données météorologiques sur le bassin versant et, par une série de calculs encadrés par un certain paramétrage, retourne des débits à l'exutoire du bassin traité.

À l'opposé, un modèle distribué prend en compte la distribution spatiale des données météorologiques, des types de sol et des caractéristiques physiques du bassin pour simuler les processus hydrologiques qui se produisent naturellement sur le bassin versant. Les sorties du modèle distribué sont également spatialement distribuées, ce qui signifie que les apports et autres variables peuvent être simulés à n'importe quel endroit sur le bassin.

### **3.4.1 Modèles hydrologiques utilisés**

Le choix des modèles hydrologiques demande réflexion pour la présente étude car plusieurs contraintes doivent être satisfaites. Premièrement, le temps de calcul doit être raisonnable. Il est impensable d'effectuer des milliers de calages d'un modèle hydrologique distribué à forte résolution puisque cela prendrait des années. Les modèles doivent également tenir compte des caractéristiques particulières des bassins nordiques, telles que le gel de l'eau souterraine et la fonte de la neige. De plus, les modèles devraient être de nature différente l'un de l'autre. Cette section présente les trois modèles hydrologiques qui ont été retenus pour cette étude, soient HSAMI, HMETs et Hydrotel. Ces modèles ont été sélectionnés en raison de leur disponibilité et de leur diversité relative recherchée.

#### **3.4.1.1 HSAMI**

HSAMI (Bisson et Roberge, 1983; Fortin, 2000; Minville et al., 2008) est un modèle hydrologique global conceptuel utilisé par Hydro-Québec. Codé à la base en C++ et adapté au langage MATLAB, il se prête bien à l'étude en cours. Il ne nécessite que peu de données météorologiques d'entrée (moyenne spatiale de la température et de précipitations liquides et solides) et peut s'adapter à un large éventail de bassins versants à l'aide de ses 23 paramètres de calage. La liste des paramètres ainsi que leurs bornes sont détaillées dans le tableau-A II-3, à l'annexe II. HSAMI est un modèle hydrologique basé sur des réservoirs qui simule les

principaux processus hydrologiques : Fonte de neige, évapotranspiration, infiltration et ruissellement. Les concepts de gel et de hauteur de nappe phréatique sont aussi simulés. Si des relevés de neige sont effectués sur le bassin versant, il est possible d'ajouter les valeurs des lectures dans le modèle. Il s'agit du modèle qui requiert le moins de temps à exécuter.

#### **3.4.1.2 HMETS**

HMETS (Chen et al., 2011) est un modèle global et conceptuel conçu par M. François Brissette, professeur au département de génie de la construction à l'École de technologie supérieure. Il s'agit d'un modèle toujours en développement, mais qui a démontré à maintes reprises une bonne capacité à modéliser les apports en rivière sur les bassins du Québec. À l'instar de HSAMI, il possède plusieurs paramètres à optimiser. La version utilisée en compte 19. La liste de ces paramètres ainsi que leurs bornes sont détaillées dans le tableau-A II-2, à l'Annexe II. Comme HSAMI, il est basé sur des réservoirs et simule les processus de base du cycle de l'eau. Ses intrants sont la précipitation et la température moyennes sur le bassin. La principale différence entre HSAMI et HMETS est le schéma de routage utilisé, qui est basé sur l'hydrogramme unitaire direct pour le modèle HMETS et sur des hydrogrammes unitaires directs et retardés pour les écoulements souterrains et de surface pour HSAMI. De plus, les processus de neige et de gel sont simulés très différemment dans les deux modèles.

#### **3.4.1.3 Hydrotel**

Hydrotel (Fortin et al., 2001) est un modèle distribué à base plus physique que les deux précédents modèles. Les propriétés physiques du bassin versant, telles que la pente, l'occupation et le type de sol ainsi que le réseau hydrographique sont nécessaires pour que le modèle puisse effectuer une simulation. Le bassin versant est construit dans PHYSITEL, un logiciel complémentaire à Hydrotel. Lorsque le bassin est construit, il est divisé en plusieurs unités de calcul nommées UHRH (Unité Hydrologique Relativement Homogène). Hydrotel interpole et extrapole ensuite les données météorologiques sur le bassin versant selon la méthode demandée par l'utilisateur, et des données sont fournies à chaque UHRH. Ensuite, Hydrotel simule les processus physiques pendant la période utilisée sur chacune des UHRH.

Un algorithme de routage interne fait transiter les apports jusqu'à la sortie du bassin versant, puis les résultats sont sauvegardés dans des fichiers textes. Les paramètres d'entrée du modèle qui ont été optimisés sont détaillées dans le tableau-A II-1, à l'Annexe II.

Le modèle est un fichier exécutable qui doit être appelé par l'une des deux façons suivantes :

- 1) Par l'interface graphique principale du logiciel. Cette méthode ne permet que la simulation manuelle du modèle. Les paramètres peuvent être modifiés de manière très simple et visuelle;
- 2) Par MATLAB via la commande DOS. Le logiciel peut être lancé de manière automatique, mais les paramètres doivent être modifiés dans des fichiers d'entrée du modèle. Il n'y a aucun affichage graphique ici.

Pour la présente étude, l'option 2 a été retenue en raison du nombre important de simulations effectuées. Un code MATLAB a été utilisé pour modifier les paramètres au besoin dans les fichiers d'entrée de Hydrotel. De plus, les propriétés physiques du bassin de la rivière Manicouagan (dans lequel se trouve le sous-bassin Toulnostouc) avaient déjà été modélisées dans Hydrotel. La seule modification apportée pour cette étude a été d'éliminer les sous-bassins versants autres que Toulnostouc dans ce modèle.

### **3.4.2 Calage et validation**

Tous les modèles hydrologiques ont des paramètres ajustables pour améliorer leur représentation des particularités du bassin étudié. Ces paramètres sont des facteurs dans les diverses équations des modèles. Par exemple, l'évapotranspiration réelle est calculée comme étant un pourcentage de l'évapotranspiration potentielle pour la durée de la période simulée. Le calage est le processus par lequel ces paramètres sont ajustés pour que les apports simulés par le modèle soient identiques aux apports observés. Cependant, aucun modèle hydrologique ne peut représenter exactement la réalité. Le calage est donc effectué pour que

le modèle représente le mieux possible les observations historiques. Il s'agit donc d'une étape cruciale dans le processus de modélisation hydrologique.

La validation d'un modèle sert à vérifier la performance du calage effectué préalablement. Une fois le calage terminé, une autre série de données météorologiques (sur une période inutilisée lors du calage) est fournie au modèle. Les apports simulés sont ensuite comparés aux apports observés pour la période de validation. La performance du modèle en validation dépend fortement de la qualité du calage. Cependant, plusieurs jeux de paramètres différents peuvent retourner des valeurs semblables lorsqu'ils sont utilisés pour valider le modèle alimenté avec les mêmes intrants. Le choix d'un jeu de paramètres optimal et unique est donc souvent impossible. Ce problème est celui de l'équifinalité (Beven, 2006) et n'est pas discuté dans ce mémoire. Pour cette étude, le calage retournant les meilleures valeurs en calage et en validation a été utilisé.

### 3.4.2.1 Méthodes de calage

Le calage a été effectué avec des optimisateurs globaux automatiques. Ces optimisateurs utilisent des méthodes itératives et « intelligentes » pour trouver le minimum global d'une fonction multidimensionnelle. Dans le cas de la modélisation hydrologique, le nombre de dimensions équivaut au nombre de paramètres à optimiser. La fonction à optimiser, aussi appelée *fonction objectif*, est une fonction qui calcule une métrique qui définit la différence entre les apports simulés et les apports observés lors de la modélisation avec le jeu de paramètres en question.

En optimisation globale, une évaluation de modèle équivaut à choisir un jeu de paramètres, rouler le modèle avec ces paramètres et calculer la valeur de la métrique de comparaison. Plusieurs évaluations de modèle sont effectuées par les optimisateurs. Ces derniers essaient des jeux de paramètres au hasard, puis, dépendamment des résultats retournés par la fonction objectif, améliorent les paramètres de manière itérative jusqu'à ce que les métriques

calculées soient minimisées. Plusieurs métriques différentes existent et celles utilisées dans cette étude sont présentées à la section 3.4.2.3.

Trois optimisateurs globaux ont été utilisés pour identifier les paramètres à utiliser dans les modèles, tous écrits en programmation MATLAB et tous de type évolutif. Les algorithmes évolutifs optimisent des fonctions mathématiques en simulant les évolutions successives apportées aux êtres vivants selon des processus évolutifs génétiques. L'opération débute par une sélection aléatoire de points dans l'espace à optimiser. Puis, cette population initiale est évaluée dans le modèle. Les meilleurs individus (les meilleures solutions) survivront au prochain tour, tandis que les autres seront rejetées. Pour la prochaine vague, une nouvelle population est créée en modifiant les paramètres de la vague précédente. Une certaine portion de la nouvelle population subira également des modifications plus profondes qui la distingueront des vagues précédentes, qui pourraient améliorer ou dégrader sa performance. L'opération est ensuite recommencée avec la nouvelle population jusqu'à ce que la performance ne puisse être améliorée.

Le premier algorithme, le SCE-UA utilisé pour caler les paramètres du schéma de routage, est bien détaillé dans la littérature et a été créé spécialement pour le calage de modèles hydrologiques. Cet algorithme a été utilisé pour optimiser les paramètres du modèle HSAMI puisque ce dernier est couplé à une routine d'optimisation très rapide utilisant SCE-UA.

Le second optimisateur est l'Algorithme Génétique (AG) (Holland 1975; Goldberg 1989; Schmitt 2011). Il s'agit d'une autre méthode de type évolutive de plus en plus utilisée en modélisation. Il s'agit d'une méthode rapide qui permet à l'utilisateur de facilement paralléliser le processus de calage. En effet, l'algorithme génétique rendu disponible par MATLAB propose une option pour paralléliser les évaluations de modèles. Ceci est très pratique lors du calage de modèles qui demandent beaucoup de ressources informatiques pour simuler des apports. Pour cette raison, l'AG a été utilisé pour caler le modèle Hydrotel. Avec les ordinateurs multiprocesseurs utilisés (8 nœuds de calcul), le temps requis pour un calage passe de 16 jours à 2 jours avec cet algorithme.

Le dernier optimisateur utilisé est le Differential Evolution (DE) (Storn et Price 1997; Pedersen 2010). Cette méthode est semblable aux deux méthodes précédentes et fonctionne en introduisant des petites variations dans la série d'évaluations précédentes pour créer une nouvelle série de solutions candidates. L'évolution se fait jusqu'à ce que les solutions candidates aient toutes convergé vers un minimum, idéalement global. Cette méthode a été utilisée pour caler le modèle HMETS, d'une part en raison de l'efficacité de la méthode (Arsenault et al., 2011), et d'autre part parce que la mise en application de la version MATLAB de l'optimisateur était triviale sur ce modèle. Finalement, la diversité des optimisateurs utilisés permet de créer des cas-types assez différents pour assurer que les résultats ne sont pas le fruit d'un optimisateur particulier.

Les trois méthodes pourraient théoriquement chercher de nouvelles solutions pour un temps infini, mais pratiquement, un nombre maximal d'évaluations est souvent utilisé pour limiter le temps requis. Dans cette étude, le critère d'arrêt est la première des 2 éventualités suivantes :

- 1) Le nombre d'évaluations de modèle atteint 25 000;
- 2) L'écart entre la meilleure valeur de cette itération et celles des 3 précédentes itérations est nulle (la tolérance de la fonction = 0 pour les 3 dernières itérations).

Il est à noter que les optimisateurs automatiques sont performants si des bornes sont fournies pour assurer que les paramètres trouvés sont toujours physiquement possibles. Cependant, certains modélisateurs préconisent l'utilisation d'une méthode manuelle pour l'optimisation des paramètres lors du calage. Quoiqu'il soit vrai qu'un calage manuel permet d'obtenir des résultats satisfaisants et permet de mieux comprendre l'effet de chacun des paramètres, le temps requis pour arriver à des résultats optimaux est beaucoup plus grand. Par exemple, le calage automatique de HSAMI (avec 25 000 évaluations) requiert environ 60 secondes. Il serait difficile de faire 2 évaluations dans le même temps avec la méthode manuelle. Puisque

plus d'un million de calages seront effectués à 25 000 évaluations chacune en moyenne, il est indispensable d'utiliser des méthodes automatiques. De plus, ni la méthode manuelle, ni la méthode automatique ne garantissent que la solution proposée soit la solution optimale. Ceci est une caractéristique prouvée des problèmes dont la fonction objectif est non-convexe, tel que c'est le cas en modélisation hydrologique (Fortnow, 2009). Toutefois, plusieurs optimums locaux de la fonction objectif peuvent avoir la même valeur avec des jeux de paramètres différents. Ses solutions sont considérées comme étant équivalentes jusqu'à preuve du contraire.

#### **3.4.2.2 Périodes de calage**

Les périodes de calibration et de validation utilisées sont les mêmes pour tous les modèles hydrologiques afin de garder une certaine similitude entre les résultats des différents modèles. Le calage a été effectué sur la période 1961-1975 inclusivement, et la période entre 1976 et 2000 a été retenue pour la validation.

Pour la partie de l'étude qui utilise une optimisation multicritère, différentes périodes de calage ont été utilisées et seront expliquées au moment opportun.

#### **3.4.2.3 Métrique d'efficacité**

Le calage d'un modèle se fait en minimisant une fonction objectif qui mesure l'écart entre la série de données simulées et les observations. Plusieurs fonctions objectif sont utilisés pour arriver à cette fin : Nash-Sutcliffe (NS), l'erreur quadratique moyenne (RMSE), le coefficient de détermination, ( $R^2$ ), le Biais Relatif (BR), etc. Chacune a des forces et des faiblesses (Moriassi et al., 2007). Par exemple, le critère de Nash-Sutcliffe tend à favoriser des meilleures simulations de crues, tandis que le biais tend à favoriser la précision du volume global d'eau en rivière pendant la période simulée. Pour les besoins de l'étude, deux critères ont été utilisés pour le calage des modèles. Le premier est celui de Nash-Sutcliffe (voir section 4.4) et le second est un critère modifié calculé comme suit :

$$CM = \frac{1}{3} \times (1 - NS) + \frac{1}{3} \times |BR| + \frac{1}{3} \times CV(RMSE) \quad (3.1)$$

Où CM est le critère multiple, NS est la valeur de Nash-Sutcliffe, BR est le biais relatif et CV(RMSE) est le coefficient de variation de l'erreur quadratique:

$$NS = \left( \left[ 1 - \frac{\sum_{t=1}^T (Q_t^o - Q_m^t)^2}{\sum_{t=1}^T (Q_t^o - \bar{Q}_o)^2} \right] \right) \quad (3.2)$$

$$BR = \frac{\sum_{t=1}^T (Q_m^t - Q_t^o)}{\sum_{t=1}^T (Q_t^o)} \quad (3.3)$$

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{t=1}^T (Q_m^t - Q_t^o)^2}{T}}}{\bar{Q}_o} \quad (3.4)$$

Où  $Q_t^o$  est la valeur observée au temps  $t$ ,  $Q_m^t$  est la valeur modélisée au temps  $t$ ,  $\bar{Q}_o$  est la moyenne des valeurs observées et  $T$  est le nombre de valeurs d'apports quotidiens dans la série.

Ces critères composés sont monnaie courante en modélisation. La pondération utilisée permettra d'obtenir un compromis entre une bonne simulation de la crue (NS), une bonne simulation pour les plus faibles débits (CV(RMSE)) et un volume d'eau total similaire (Biais relatif).

Dans la partie de l'étude traitant de l'optimisation multi-objective en calibration avec HSAMI, le critère utilisé est celui de Nash-Sutcliffe puisque l'optimisateur automatique intégré est basé sur ce paramètre.

Lors de la construction de la fonction objectif, il faut tenir compte du fait que l'optimisateur ne peut que minimiser la fonction. Par défaut, des valeurs 0 pour la valeur absolue du biais relatif et pour le CV(RMSE) sont idéaux, donc ces deux critères ne posent pas problème. Le

critère de Nash-Sutcliffe, cependant, est compris entre  $-\infty$  (très mauvais) et 1 (parfait). Pour forcer l'optimisateur à « maximiser » le NS, la fonction objectif a été définie par  $(1 - NS)$ . Les résultats des divers calages sont présentés à la section 4.1.

### **3.5 Effets de la densité du réseau d'observations sur la variabilité spatiale des données interpolées**

Un des aspects du problème à considérer est l'effet de la densité du réseau d'observations sur les interpolations spatiales des données météorologiques fournies aux modèles hydrologiques. Pour déterminer l'impact des méthodes d'interpolation, un essai simple a été mené. La grille complète de 1369 points du MRCC a été utilisée pour cet essai.

La première étape consiste à retirer systématiquement une station sur deux de la grille complète. Le patron des stations restantes dans la nouvelle grille ressemble donc à un échiquier. Un schéma d'interpolation est ensuite utilisé pour générer des résultats aux points de grille où les stations ont été retirées. Les valeurs interpolées peuvent ensuite être comparées aux valeurs originales. Cette étape a été répétée pour plusieurs autres ratios d'enlèvement de stations pour la précipitation annuelle moyenne.

La seconde étape a été de sélectionner des points de manière aléatoire et de prendre la moyenne annuelle des valeurs de précipitation à ces points. En comparant ces valeurs à la moyenne de tous les points, il a été possible de déterminer s'il y a un nombre optimal de stations qui conserve la valeur moyenne du système. Les résultats sont présentés à la section 4.2.

### **3.6 Effets de la densité du réseau d'observations sur la modélisation hydrologique**

Les effets de la densité du réseau d'observations ont été examinés de manière pratique. Pour y arriver, trois grandes étapes ont été suivies :

- 1) Les trois modèles ont été calés avec toutes les stations virtuelles disponibles (39). Ceci permet d'obtenir la valeur de la fonction objectif de référence ( $CM_{réf}$ ), qui sera considérée comme la valeur optimale. De plus, le jeu de paramètres qui permet d'obtenir la valeur  $CM_{réf}$  sera enregistré ( $X_{optim}$ ).
- 2) Les modèles ont ensuite été roulés avec les paramètres optimaux  $X_{optim}$  tout en retirant des stations de manière semi-aléatoire entre les simulations. afin de simuler différentes densités du réseau météorologique. Le retrait semi-aléatoire signifie que les stations ont été retirées aléatoirement mais de manière à préserver une certaine homogénéité dans la distribution spatiale des stations. La fonction objectif calculée après chaque simulation a été enregistrée.
- 3) Les combinaisons de stations produites en 2) ont été réutilisées dans les modèles hydrologiques, cette fois en mode calage. Les modèles ont donc été calés en utilisant ces combinaisons de stations afin de mesurer la performance des modèles hydrologiques à compenser pour le manque de stations. Les nouvelles valeurs des fonctions objectifs ont été comparées à la valeur de référence  $CM_{réf}$ . Cette étape a été effectuée pour simuler les résultats qui seraient obtenus dans un contexte réel, où toutes les données disponibles sont utilisées lors d'un calage.
- 4) L'étape définie en 3) a été répétée sans recalage sur une autre période afin de déterminer la performance des modèles utilisant un réseau moins dense en mode validation.

Cette méthode, utilisant une approche manuelle, a ensuite été comparée à d'autres méthodes automatiques et semi-intelligentes, tel que décrit à la section 3.7.. Les résultats de cette partie de l'étude sont montrés à la section 4.3.

### 3.7 Détermination du nombre de stations optimal et de leur emplacement

Cette section traite de la détermination du nombre de stations minimal nécessaire pour obtenir des résultats adéquats en termes de performance des modèles hydrologiques. De plus, l'emplacement optimal des stations est examiné. Pour y parvenir, une méthode d'optimisation multi-objective a été utilisée.

#### 3.7.1 Description du front de Pareto

Le front de Pareto est le résultat d'une optimisation multi-objective. Chacune de ses solutions est optimale pour la combinaison de conditions imposées. Dans le cadre de ce projet, l'algorithme d'optimisation utilisé fût l'AG.

Dans une optimisation multi-objective, le but est de déterminer l'optimum de deux variables conflictuelles (Deb, 2001). Il ne peut donc pas y avoir de valeur optimale unique. L'exemple qui suit démontre le problème :

Supposons qu'on veuille générer un code binaire de 8 chiffres. Il ne peut être composé que de 1 ou de 0. En binaire, le dernier chiffre représente le coefficient multipliant  $2^0$ , l'avant dernier représente le coefficient multipliant  $2^1$ , et ainsi de suite. Dans notre nombre binaire, le premier chiffre vaut donc  $2^7$ . Pour représenter un nombre décimal, il suffit d'additionner les valeurs numériques dont le bit est à 1, et d'ignorer les bits nuls. Par exemple, le nombre binaire 1001 est égal à  $2^3 + 2^0 = 9$ .

Maintenant, ajoutons des contraintes. Premièrement, nous voulons que la somme des bits soit la plus petite possible. Ceci ne pose pas de problème en soi dans un problème d'optimisation uni-critère : Le code binaire 00000000 (0 en décimal) est la seule et unique réponse possible. Cependant, imaginons que nous voulons également que ce nombre binaire représente la valeur décimale la plus grande possible. Cette nouvelle condition est conflictuelle avec la première, car la réponse à cette condition, évaluée indépendamment, est 11111111 (255 en décimal).

Le front de Pareto permet d'optimiser les contraintes ensemble pour réduire l'espace de sélection pour l'utilisateur. Puisque cette méthode nécessite que les deux conditions soient minimisées, la deuxième condition sera présentée comme étant la minimisation de 255 moins la valeur décimale du code binaire. La figure 4.2 représente la solution du front de Pareto pour cette mise en situation.

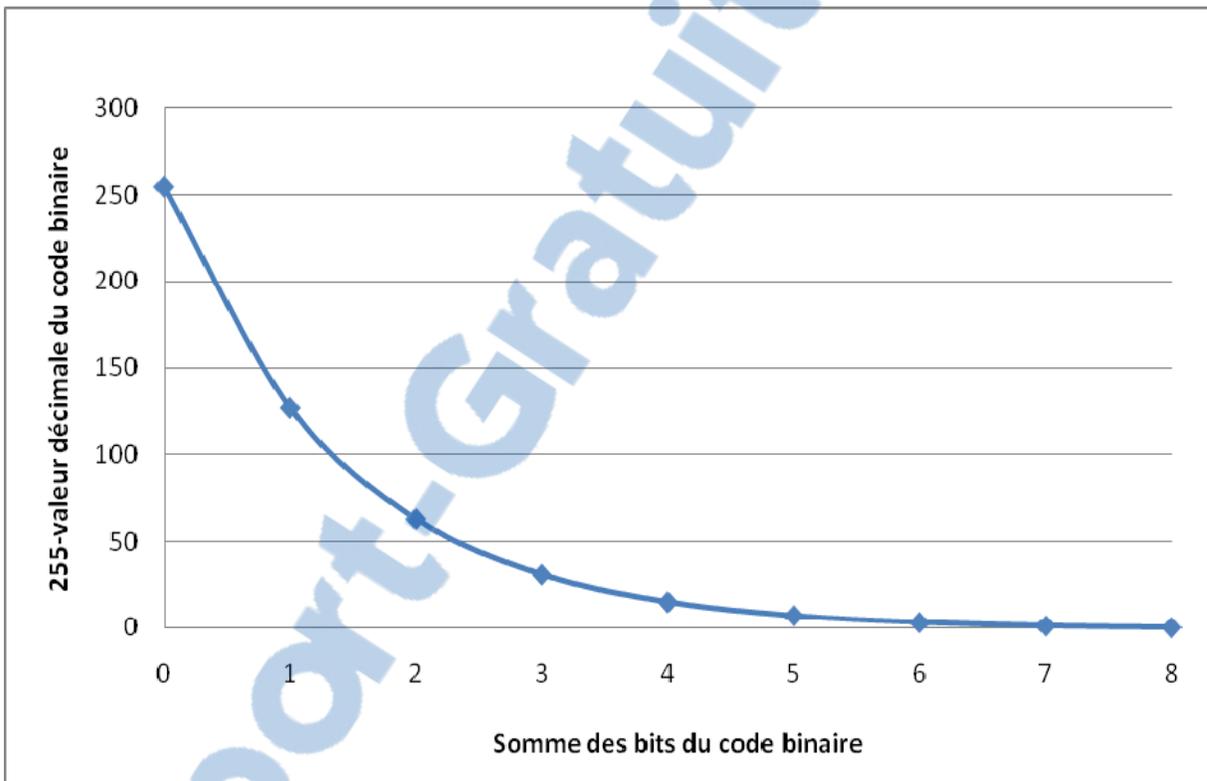


Figure 3.10 Front de Pareto pour une solution binaire vérifiable

Les solutions Pareto-optimales ainsi que les valeurs des contraintes sont présentées dans le tableau 3.1.

Tableau 3.1 Solutions Pareto-optimales pour la figure 3.2

Code binaire	Somme des bits	255-valeur décimale du code binaire
00000000	0	255
10000000	1	127
11000000	2	63
11100000	3	31
11110000	4	15
11111000	5	7
11111100	6	3
11111110	7	1
11111111	8	0

Il est à noter que le code binaire 10000001 ne pourrait faire partie du front de Pareto car il est dominé par les codes 10000000 (minimiser la somme des bits) et 11000000 (maximiser la valeur décimale avec 2 bits). Seules les solutions sur le front de Pareto sont optimales pour une condition d'optimisation ou une autre. Cela signifie que le front de Pareto fournit des solutions optimisées pour les deux conditions. Toutes les solutions se valent; il n'y en a pas une meilleure que l'autre. Pour avoir une solution unique, il faut donner une pondération aux solutions Pareto-optimales et les évaluer.

### 3.7.2 Évaluation de la méthode d'optimisation du Front de Pareto

Le front de Pareto est largement dépendant de l'algorithme d'optimisation qui est utilisé pour les problèmes complexes, surtout si ceux-ci ne peuvent pas être solutionnés de manière analytique. Pour s'assurer que l'algorithme est assez performant pour les besoins de cette étude, une série de problèmes-étalons a été testée avec l'AG. Un des problèmes multicritères Zitzler-Deb-Thiele (ZDT), qui sont parmi les plus difficiles du type à résoudre, a été résolu avec l'AG. Il s'agit du problème ZDT-5, qui est un problème de type binaire. Les fonctions à minimiser sont :

$$\text{ZDT - 5: } \left\{ \begin{array}{l} f_1(x) = 1 + u(x_1) \\ g(x) = \sum_{i=2}^{11} v(u(x_i)) \\ v(u(x_i)) = \begin{cases} 2 + u(x_i), & u(x_i) < 5, \\ 1, & u(x_i) = 5, \end{cases} \\ h(f_1, g) = \frac{1}{f_1(x)} \end{array} \right. \quad (3.5)$$

Le problème est composé de 11 variables distinctes. La première,  $x_1$ , comporte 30 bits, tandis que les 10 autres,  $x_2, x_3, \dots, x_{11}$  sont composées de 5 bits chacune. La fonction  $u(x_i)$  dénote le nombre de bits qui ont la valeur de 1 dans la variable. Les deux contraintes données à l'optimiseur multi-objectif sont :

- minimiser  $f_1(x)$ ,
- minimiser  $f_2(x) = g(x)h(f_1(x), g(x))$ .

Les fronts de Pareto réel et simulé par l'algorithme génétique multi-objectif pour le problème ZDT-5 sont montrés à la Figure 3.11. La courbe rouge représente les solutions simulées par l'algorithme génétique multi-objectif et la courbe bleue représente le front de Pareto réel du problème. La simulation a été effectuée en utilisant une population de 800 individus et 30 générations évolutives.

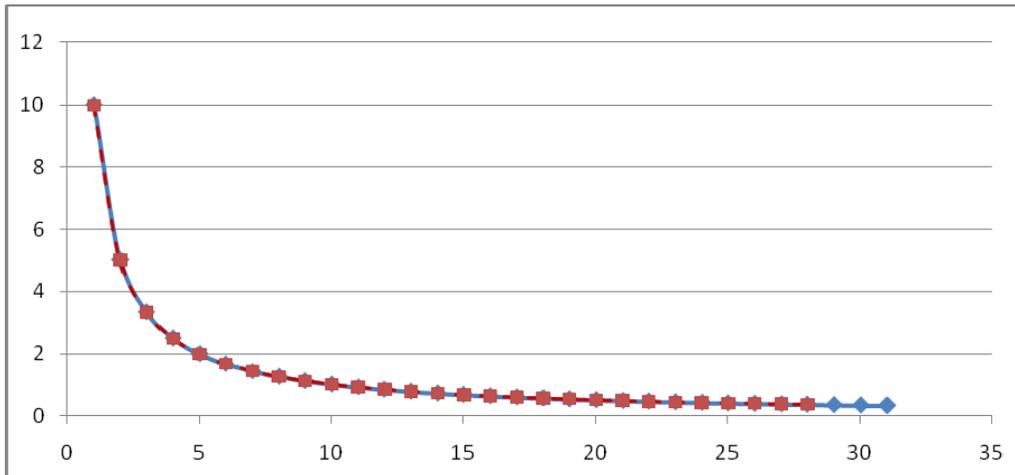


Figure 3.11 Front de Pareto pour le problème multi-objectif ZDT-5

Ce problème a été utilisé car les mêmes difficultés devraient être rencontrées lors des optimisations multicritères dans cette étude. En effet, les problèmes seront du type déceptif et binaire, tout comme ce problème d'essai. On entend par déceptif un problème dont une des fonctions à minimiser comprend plusieurs solutions qui tendent vers le même minimum local alors qu'en réalité un minimum global isolé peut exister. L'algorithme tend donc à viser les minimums locaux et ignorer la solution optimale globale.

### 3.7.3 Application à cette étude

Dans le cadre de ce projet, le code binaire utilisé est un code à 39 bits, où la position de chaque bit correspond à une station (bit 1 = station #1, bit 2 = station #2, etc.). L'algorithme tentera de déterminer le front de Pareto pour :

- 1) Minimiser le nombre de stations météorologiques;
- 2) Minimiser les fonctions d'erreur entre les débits simulés et observés.

En théorie, ces deux conditions sont conflictuelles puisque plus il y a de stations météorologiques, meilleur devrait être le résultat de la simulation (Cheng et al., 2008).

Le front de Pareto a été mis en œuvre pour deux types d'analyses dans cette étude. Dans le premier cas, les modèles ont été préalablement calés avec les 39 stations disponibles, puis le front de Pareto a été utilisé pour voir la dégradation des métriques d'erreur en fonction de la diminution du nombre de stations utilisées. Dans cette portion, aucun calibrage n'est effectué pendant l'optimisation des stations utilisées.

Dans le second cas, le modèle est calibré pour chacune des évaluations du front de Pareto. Ceci permet de donner un aspect réel au problème. En effet, s'il existe en réalité seulement 5 stations météorologiques sur un bassin versant, le modèle serait vraisemblablement calé en utilisant les 5 stations. Cette étape a été effectuée avec HSAMI seulement en raison des contraintes de temps. Un calage automatique de HSAMI requiert environ 60 secondes, tandis que le même calage pour HMETTS requiert environ 2 heures. Pour Hydrotel, il faudrait compter au moins 2 jours avec un ordinateur à 8 processeurs de haute performance de 2010, et ce, pour chaque évaluation. Lorsque le nombre de calages à effectuer est pris en compte (minimum 10 000), il devient évident que seul le modèle HSAMI se prête à une telle pratique.

Les solutions Pareto-optimales ainsi que la dernière population des optimisations multi-objectives ont été recueillies et comptabilisées pour quatre périodes de 10 ans (1961-1970, 1971-1980, 1981-1990 et 1991-2000) ainsi que pour la période complète. La fréquence où chacune des stations était présente dans les ensembles-solutions pour chaque période a été calculée. Puis, la moyenne de la fréquence d'occurrence des stations dans les ensemble optimaux a été prise pour les quatre périodes, donnant ainsi un aperçu de l'importance de chacune des stations lors du calage, et ce, sur toute la période. Les résultats de cette partie de l'étude sont présentés aux sections 4.4 et 4.5.

#### **3.7.4 Validation de la méthode sur différentes périodes**

L'identification des stations les plus souvent présentes dans les ensembles optimaux permet de réduire le nombre de stations considérées lorsqu'un nombre restreint d'entre elles doit être

choisi pour représenter le bassin versant. Cependant, il se pourrait qu'une combinaison de quelques stations supposément idéales donne des résultats moins performants que si toutes les stations avaient été utilisées. Une méthode de validation a donc été mise en place.

Cette validation utilise une approche simple visant à montrer que les combinaisons utilisant les stations optimales retournent des résultats acceptables en calibration, et que leur performance est meilleure que si les stations avaient été choisies au hasard. Le modèle HSAMI a donc été utilisé une fois de plus pour cette étape. Des combinaisons de 2, 3, 4 et 5 stations ont été testées. Le modèle a été calibré automatiquement avec l'algorithme SCE-UA avec les données météo des stations utilisées sur la période entière de 40 ans. Le critère d'efficacité de Nash-Sutcliffe retourné a été utilisé comme barème de comparaison. La performance du modèle HSAMI utilisant des stations aléatoires a été comparée à sa performance en utilisant les stations démontrant le meilleur potentiel. Cet exercice a également été mené dans un contexte de validation, où les stations sélectionnées ont été comparées aux stations choisies de manière aléatoire (section 4.6).

### **3.8 Hypothèses**

Cette recherche est basée sur quelques hypothèses de base.

- 1) Le monde virtuel qui fait office de réalité absolue est en fait une approximation physique la plus proche de la réalité qui nous est accessible. Cette contrainte n'est pas cruciale puisque la performance des modèles sera évaluée à l'aide des données - et comparée aux sorties - de ce monde virtuel. Il n'est donc pas nécessaire de valider les données du monde virtuel par rapport aux observations.
- 2) Le schéma de routage proposé simule des débits qui se comportent sensiblement de la même façon que dans la réalité, mais il est fort probable que les paramètres qui le régissent soient au moins légèrement différents que ce qui se produit dans la réalité.
- 3) Les deux conditions à minimiser (la fonction objectif et le nombre de stations) sont contradictoires dans le processus d'optimisation multi-objective.

## CHAPITRE 4

### RÉSULTATS

Cette section expose les résultats obtenus dans cette étude. Les résultats des divers calages, de l'effet des interpolations spatiales ainsi que les effets de la densité des réseaux d'observation selon l'approche manuelle et le processus multi-objectif seront présentés.

#### 4.1 Calage et validation des modèles hydrologiques

Le calage des modèles avec les apports reconstitués a été mené sur la période 1961-1976 et la validation a été effectuée sur la période 1976-2000. Les valeurs des critères de performance pour chacun des modèles sont présentées dans le tableau 4.1. Les calages et validations ont été effectués en utilisant les 39 stations afin d'établir la valeur de référence.

Tableau 4.1 Mesure des critères de performance des modèles HSAMI, HMETS et Hydrotel en calage et en validation

Période	Critère	HSAMI	HMETS	Hydrotel
Calage	Nash-Sutcliffe	0.9328	0.8976	0.7732
	Biais relatif	+0.00005	-0.0136	+0.1437
	CV(RMSE)	0.3132	0.3836	0.612
Validation	Nash-Sutcliffe	0.9207	0.8764	0.7325
	Biais Relatif	+0.0041	-0.0079	-0.0596
	CV(RMSE)	0.3722	0.4646	0.6298

Il est à noter que la performance du modèle Hydrotel est significativement inférieure à celle de HSAMI et de HMETS. Une cause possible est le fait que Hydrotel tient compte des bases de données de sol et de couvert végétal différentes et de résolution plus fine que celles utilisées par le MRCC. Ceci peut engendrer des écarts lors de la simulation puisqu'aucun paramètre de calage ne contrôle cet écart.

## 4.2 Effets de l'interpolation lors de la réduction de la densité du réseau

Les modèles hydrologiques distribués utilisent souvent des méthodes d'interpolation de données pour pallier au manque de données sur les zones non-représentées du bassin versant. La figure 4.1 représente la précipitation annuelle moyenne simulée par le MRCC sur l'ensemble des 1369 (37x37) points de la grille. Chacun des pixels représente un point de grille du MRCC, et chaque tuile mesure approximativement 15 km de côté.

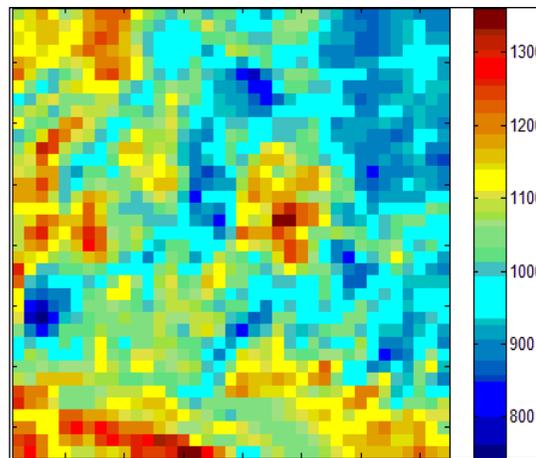


Figure 4.1 Carte des précipitations annuelles moyennes (mm) simulées par le MRCC15 sur le bassin versant de la Rivière Manicouagan. La haut de la carte pointe vers le Nord-Ouest.

Les résultats des essais de réduction uniforme de la densité du réseau sont présentés à la figure 4.2. Elle représente les valeurs interpolées ainsi que la différence entre les valeurs interpolées et les valeurs initiales lorsque 75% des stations (une sur quatre) et 86% des stations (une sur sept) sont retirées. Le choix d'un sur sept est basé sur le fait qu'il s'agit de la densité maximale des stations météorologiques qu'il y a eu historiquement sur le bassin versant. L'interpolation a été effectuée sur la moyenne climatique de la précipitation.

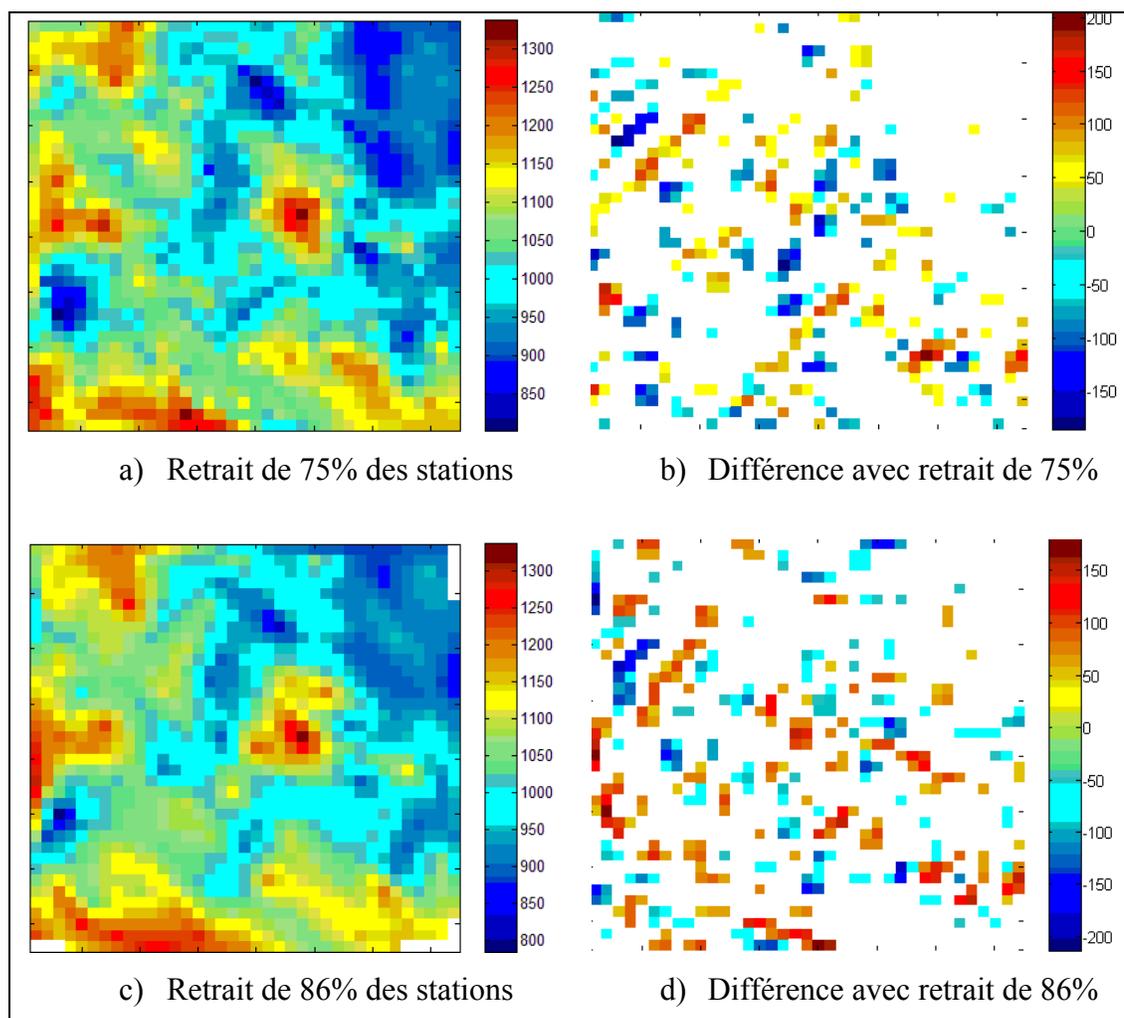


Figure 4.2 Résultats des essais de réduction de densité du réseau et erreurs produites par le schéma d'interpolation sur la précipitation moyenne annuelle (mm) Les valeurs interpolées (a) et la différence avec les valeurs initiales (b) lorsque 75% des stations sont retirées et les valeurs interpolées (c) et la différence avec les valeurs initiales (d) lorsque 86% des stations sont retirées.

L'effet de la réduction de la densité se fait rapidement sentir lorsque trois stations sur quatre sont retirées. L'interpolation triangulaire utilisée produit les grilles montrées aux Figures 4.2a et 4.2c. Ces images permettent de constater que le niveau de détail a diminué quelque peu. Les différences entre la grille initiale et reconstituée peuvent être mieux aperçues dans les Figures 4.2b et 4.2d. Il est à noter que les différences de 50mm et moins entre les valeurs initiales et interpolées ne sont pas affichées sur les figures illustrant les écarts d'interpolation, ce qui explique les zones qui n'ont aucune couleur (qui sont blanches). De plus, la figure 4.2c

a des données manquantes dans les coins. Ceci est une conséquence de la faible densité des stations qui ne permet pas au schéma d'interpolation de calculer des données à ces endroits.

À l'échelle réelle, le bassin de la rivière Toulouste comprend actuellement 5 stations météorologiques pour une superficie de 8109 km<sup>2</sup>, ce qui équivaut à une densité d'un point de grille sur sept. Les résultats d'une interpolation à cette échelle sont encore plus marquants, tel que démontré à la figure 4.2c. Il y a dans ce cas une perte de résolution de la moyenne climatique de la précipitation. Les différences entre cette carte interpolée et la réalité sont montrées à la figure 4.2d. Lorsqu'on tient compte de la résolution temporelle des stations actuellement en fonction au Québec, il devient évident que la quantité de données disponibles pour réaliser de études hydrologiques complètes est insuffisante. Dans cet exemple simple, les écarts dus à l'interpolation font varier la précipitation de  $\pm 200$  mm par année sur une portion non-négligeable du domaine étudié. Ceci peut engendrer des erreurs assez importantes dans les modèles lors du calage et peut mener à des simulations trompeuses. Il serait alors probable qu'une surestimation ou une sous-estimation des apports survienne. Dans cette optique, les gestionnaires de systèmes hydriques doivent composer avec cette possibilité pour minimiser les pertes qui y sont associées.

Pour tenter de déterminer une approximation de la densité optimale pour perdre le moins d'information possible, une méthode brute a été utilisée. Une station a été choisie aléatoirement, et sa moyenne de précipitations annuelle a été calculée. Ensuite, le même calcul a été refait sur deux stations choisies de manière aléatoire, et ainsi de suite jusqu'à ce que le nombre de stations choisies « aléatoirement » soit de 1369. Plus le nombre de stations augmente, moins grand devrait être l'écart entre la moyenne des précipitations aux stations choisies et la moyenne observée lorsque 1369 stations sont utilisées. Ceci a été vérifié à la figure 4.3. La droite rouge correspond à la moyenne réelle, qui est atteinte lorsque les 1369 points sont utilisés.

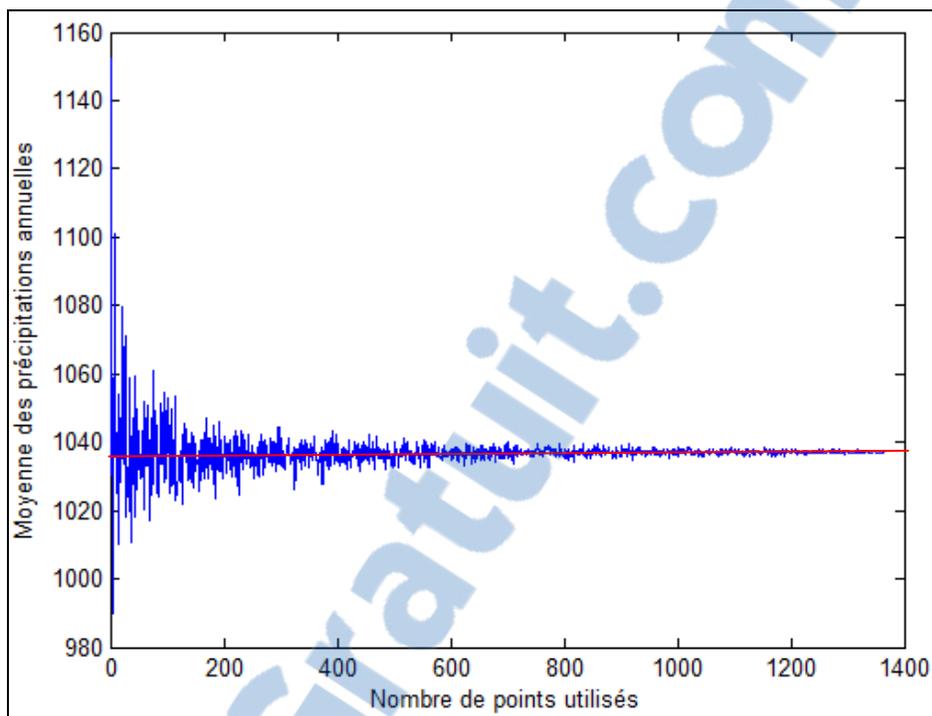


Figure 4.3 Moyenne de la précipitation pour des combinaisons aléatoirement générées contenant des nombres de stations progressivement en augmentation.

La figure 4.3 semble montrer que la moyenne n'est que très peu affectée lorsque le nombre de stations dépasse 400. Ceci correspond à une station par 770 km<sup>2</sup> en considérant le maillage de 15 km par 15 km du MRCC ayant produit la grille sous-jacente.

#### 4.3 Effets de la densité du réseau d'observations sur la modélisation hydrologique

Les stations initialement utilisées pour déterminer l'effet de la densité du réseau sont illustrées à la Figure 4.4. Les stations ont été sélectionnées de manière aléatoire tout en préservant une certaine homogénéité dans l'espace.

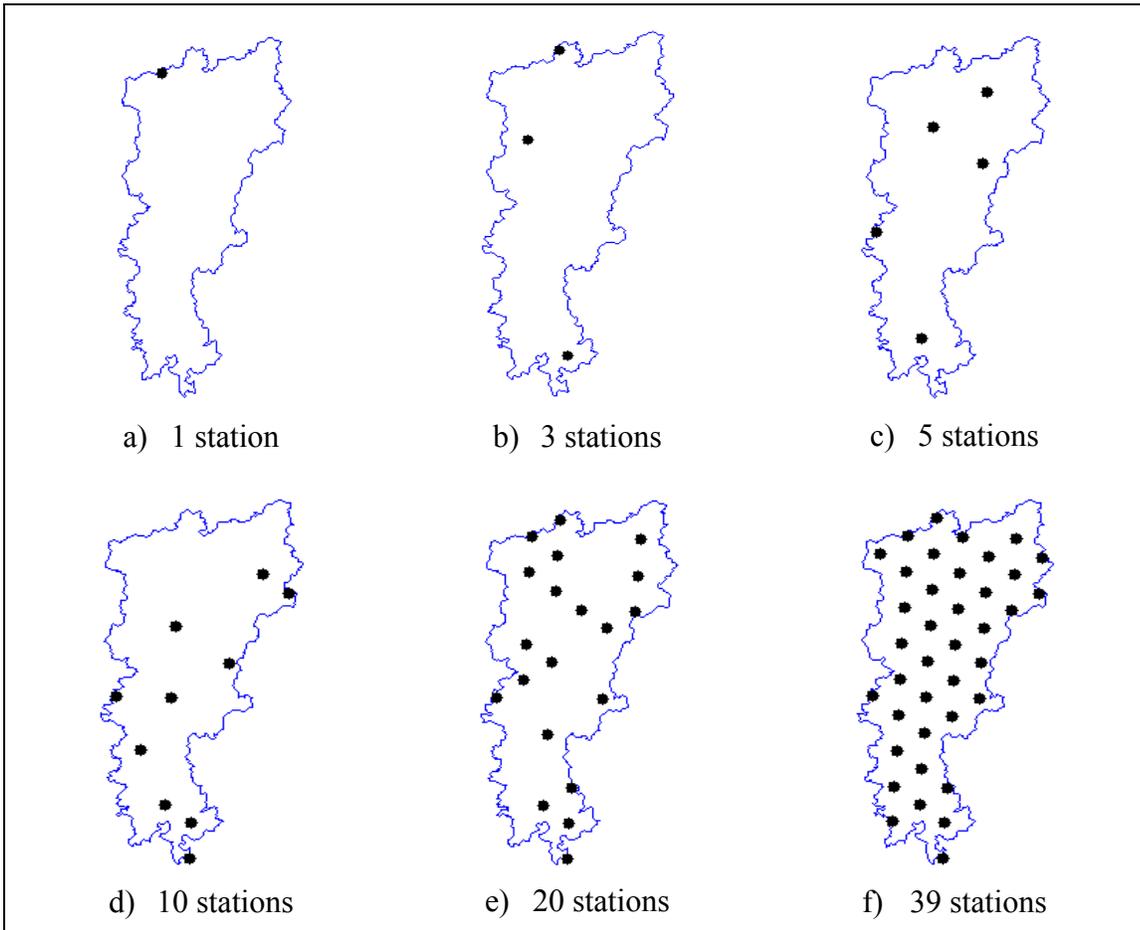


Figure 4.4 Stations sélectionnées pour les essais manuels sur le bassin versant de la Rivière Touloustouc pour a) 1 station, b) 3 stations, c) 5 stations, d) 10 stations, e) 20 stations et f) toutes les stations (39 stations).

### 4.3.1 Calage avec stations sélectionnées

Les stations sélectionnées ont été utilisées pour déterminer l'effet de la densité du réseau d'observations. Les critères de Nash-Sutcliffe pour chacune des combinaisons ainsi que pour les trois modèles sont présentés à la Figure 4.5.

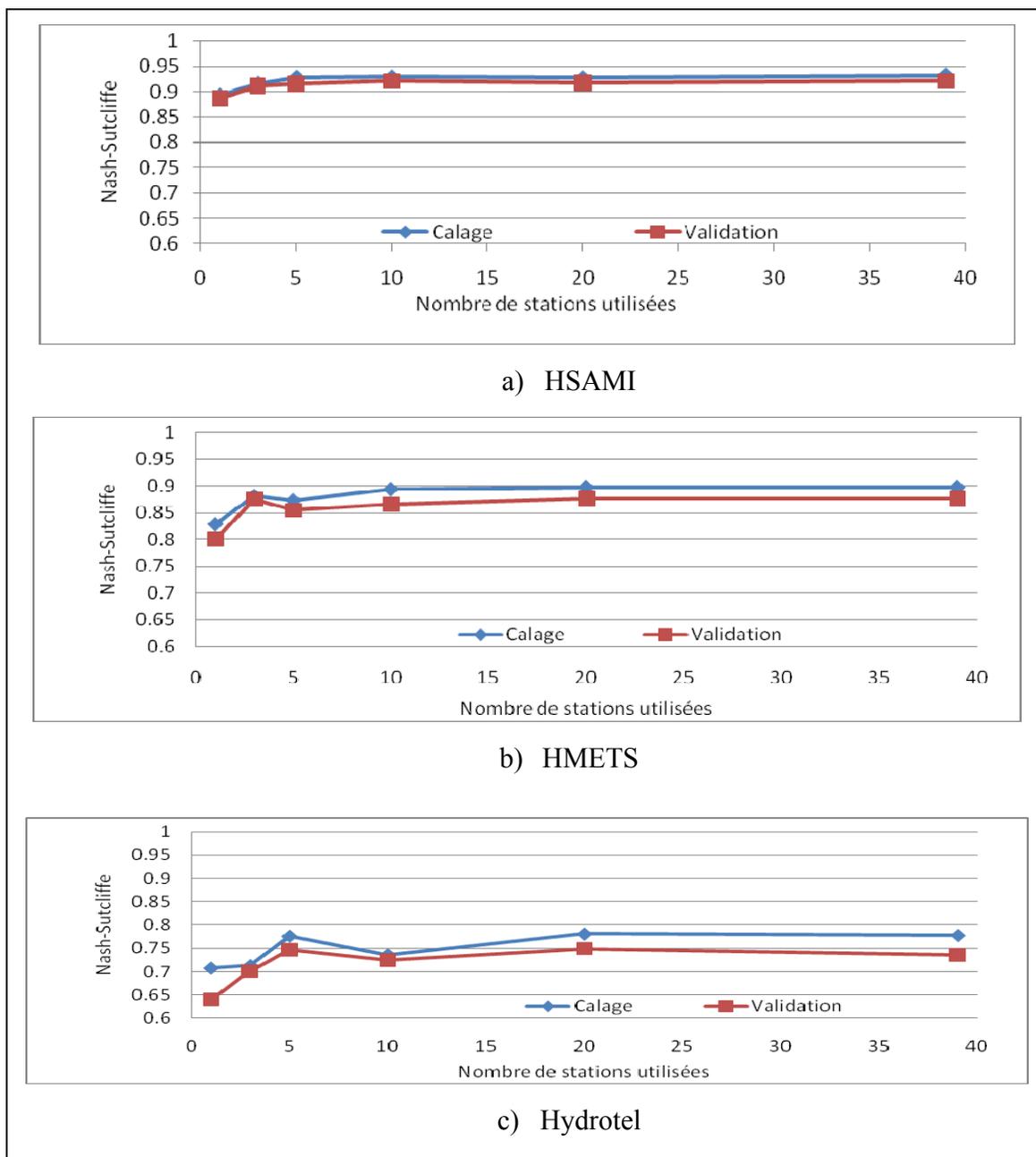


Figure 4.5 Critère de Nash-Sutcliffe en calage et en validation lorsque différentes combinaisons de stations sont utilisées pour les modèles a) HSAMI, b) HMETs et c) Hydrotel

Ces résultats semblent démontrer qu'il y a possiblement moyen de réduire le nombre de stations tout en conservant la même performance en modélisation hydrologique pour les trois

modèles. Le nombre de stations pourrait être réduit considérablement sans affecter la performance des modèles.

Les résultats de calage et de validation pour les autres composantes du critère pondéré décrit à la section 3.4.2.3 sont présentés à l'Annexe III.

#### **4.4 Optimisation multi-objective avec paramètres fixes**

Cette section expose les résultats obtenus lorsque la méthode d'optimisation multi-objective a été utilisée avec les paramètres trouvés lors du calage avec 39 stations (jeu de paramètres conservés pour tout nombre de stations). Seules les stations utilisées lors de la modélisation ont été modifiées dans le processus d'optimisation. Les figures 4.6 à 4.8 présentent respectivement les résultats obtenus pour les modèles HSAMI, HMETTS et Hydrotel.

Dans cette section, l'algorithme génétique multi-objectif optimise la sélection des stations afin de réduire le nombre de stations par combinaison tout en maximisant la valeur de Nash-Sutcliffe. L'algorithme sélectionne des stations selon le processus évolutif décrit en 3.4.2.1 et les utilise comme intrants dans les modèles hydrologiques. Ces derniers produisent un débit simulé qui est comparé au débit reconstitué  $Q_{mrcc}$ . Puis, une valeur de Nash-Sutcliffe est calculée. L'algorithme génétique utilise cette information afin de sélectionner par la suite un meilleur jeu de stations (préférentiellement avec moins de stations dans la combinaison) et qui retourne un Nash-Sutcliffe égal ou meilleur.

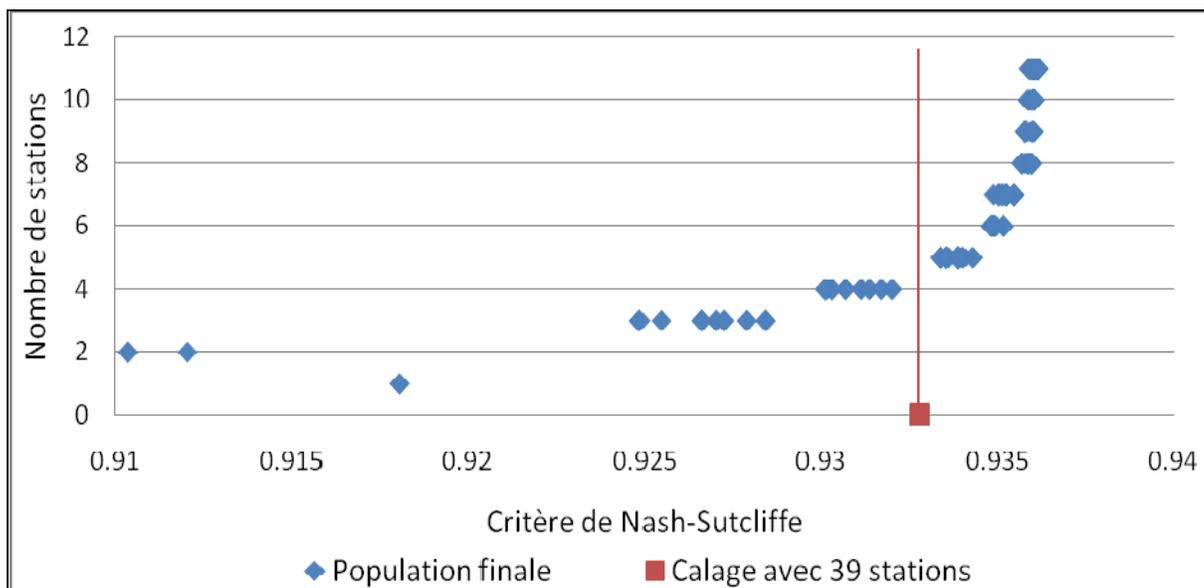


Figure 4.6 Optimisation sans recalage des stations utilisées dans HSAMI

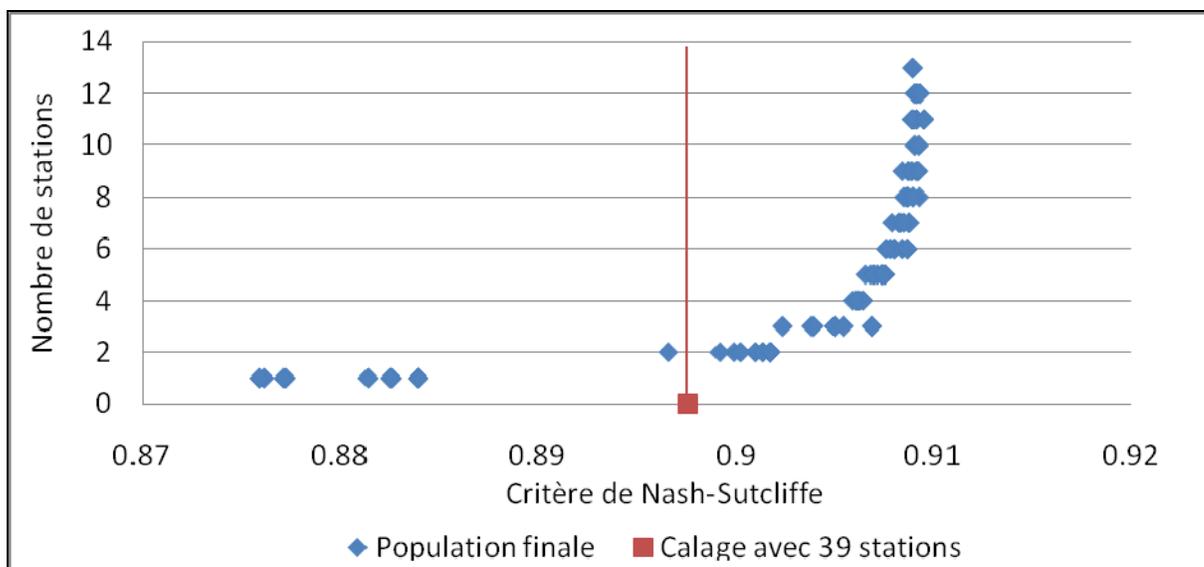


Figure 4.7 Optimisation sans recalage des stations utilisées dans HMETS

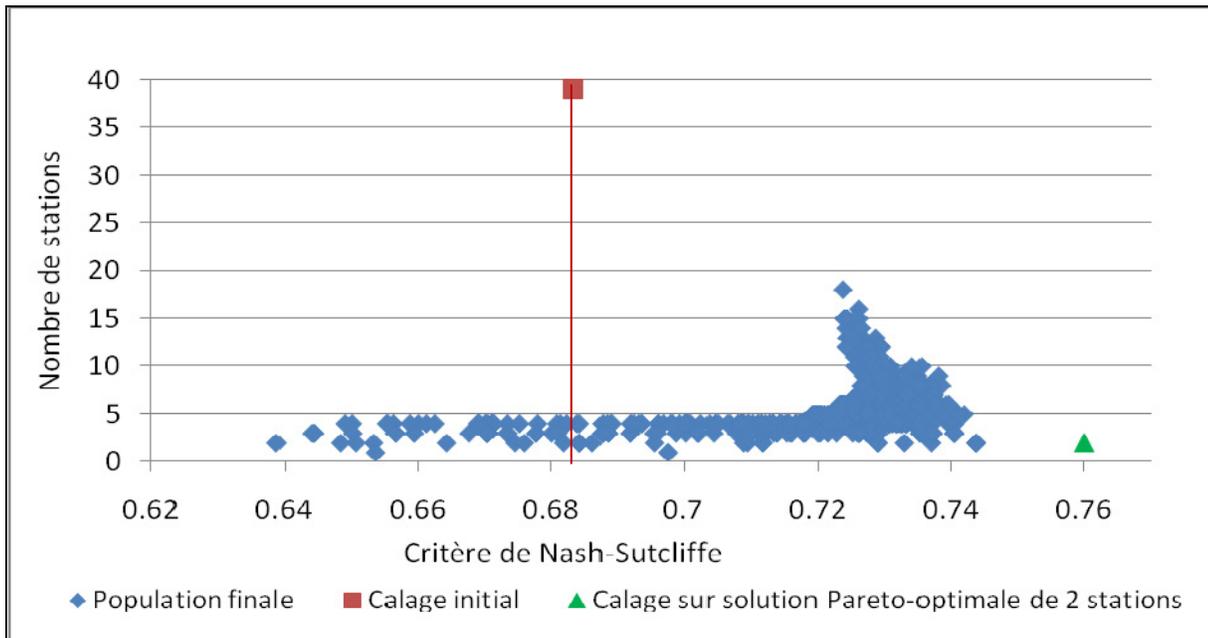


Figure 4.8 Optimisation sans recalage des stations utilisées dans Hydrotel

Les figures 4.6 à 4.8 démontrent qu'il y a place à amélioration dans le choix des stations utilisées lors du calage. Tous les points de la population finale positionnés à droite de la droite verticale définie à la valeur du calage initial avec 39 stations sont de meilleures solutions pour la modélisation hydrologique en calage, et ce, en utilisant les paramètres optimaux trouvés lors du calage à 39 stations. Il est à noter que pour cette partie de l'étude, Hydrotel a été calé sur seulement 5 années, soit la période 1961 à 1965 pour réduire le temps de calcul, ce qui explique sa performance moindre. Pour les fins de l'étude, la meilleure solution à 2 stations pour Hydrotel a été analysée davantage. Les deux stations en question ont été utilisées pour effectuer un calage du modèle dans les mêmes conditions. La performance du modèle augmente sous ses conditions (NS passe de 0.744 à 0.76), alors qu'il était de 0.685 en utilisant l'ensemble des stations. Il est donc possible d'obtenir de meilleurs résultats en utilisant moins de stations en calage. La même expérience avec HSAMI et HMETS n'a pas permis d'améliorer notablement la valeur des solutions Pareto-optimales, probablement en raison des valeurs déjà très élevées du critère de Nash-Sutcliffe.

## 4.5 Optimisation multi-objective avec calage des paramètres

Cette section présente les résultats obtenus lorsque l'approche multi-objective avec recalage a été utilisée. Le modèle a donc été recalé à chaque fois que l'algorithme génétique essayait une combinaison de stations différentes. Pour cette étape, seul le modèle HSAMI a pu être utilisé en raison du temps de calcul prohibitif des autres modèles.

### 4.5.1 Résultats des essais aléatoires pour déterminer l'effet du hasard sur la performance du modèle.

La figure 4.9 illustre les critères de Nash-Sutcliffe en calage lorsque différentes combinaisons de stations sont utilisées. La courbe la plus importante est celle composée de croix rouges (« Population finale de l'AGMO »), qui représente la solution finale de l'optimisation multi-objective pour minimiser le nombre de stations et maximiser le critère de Nash-Sutcliffe. Le losange vert à 39 stations représente la valeur de référence lorsque toutes les stations sont utilisées.

La boîte à moustaches au même niveau illustre la distribution des valeurs de Nash-Sutcliffe pour 60 calages avec les 39 stations. Tous les autres points représentent des essais avec différentes combinaisons de stations. Un test statistique de Mann-Whitney a permis de certifier que l'ensemble de solutions Pareto-optimales est significativement supérieur à l'ensemble des calages à 39 stations avec un seuil de 99%

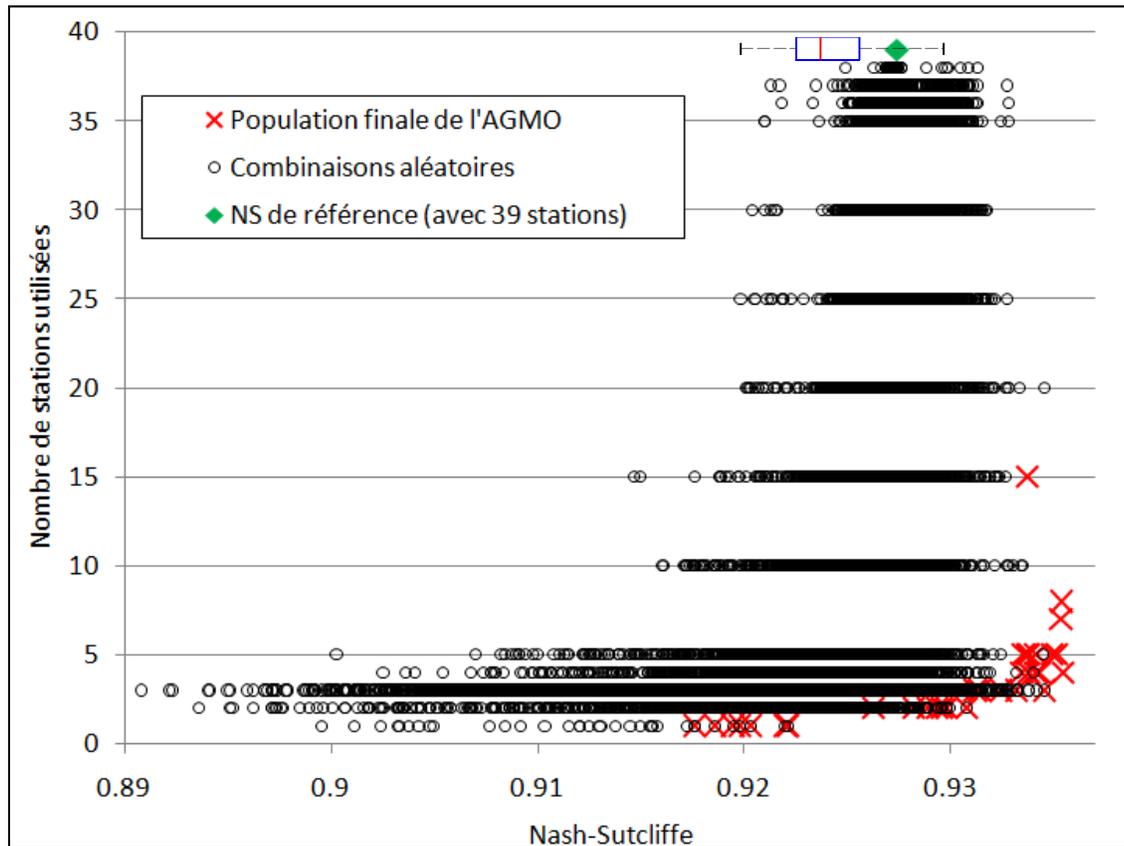


Figure 4.9 Performance du modèle HSAMI en calage avec choix de stations aléatoires

Pour les combinaisons à 1, 2, 3, 37, 38 et 39 stations, tous les cas possibles ont été évalués. Ceci permet de déterminer avec acuité les extrêmes du problème. Pour les autres cas, il y a simplement trop de combinaisons possibles. Par exemple, pour exactement 20 stations, il existe 68 923 264 410 combinaisons différentes. Il a donc été préféré d'utiliser un échantillon significatif à 5% d'erreur pour réaliser l'essai. La taille de l'échantillon a été majorée à 1000 cas par essai. Ces résultats sont présentés à la figure 4.9, et pour plus de clarté, ils sont également présentés à la figure 4.10 par une distribution d'Epanechnikov. Ce noyau permet de définir de manière non-paramétrique la densité de probabilité des données (Epanechnikov, 1969).

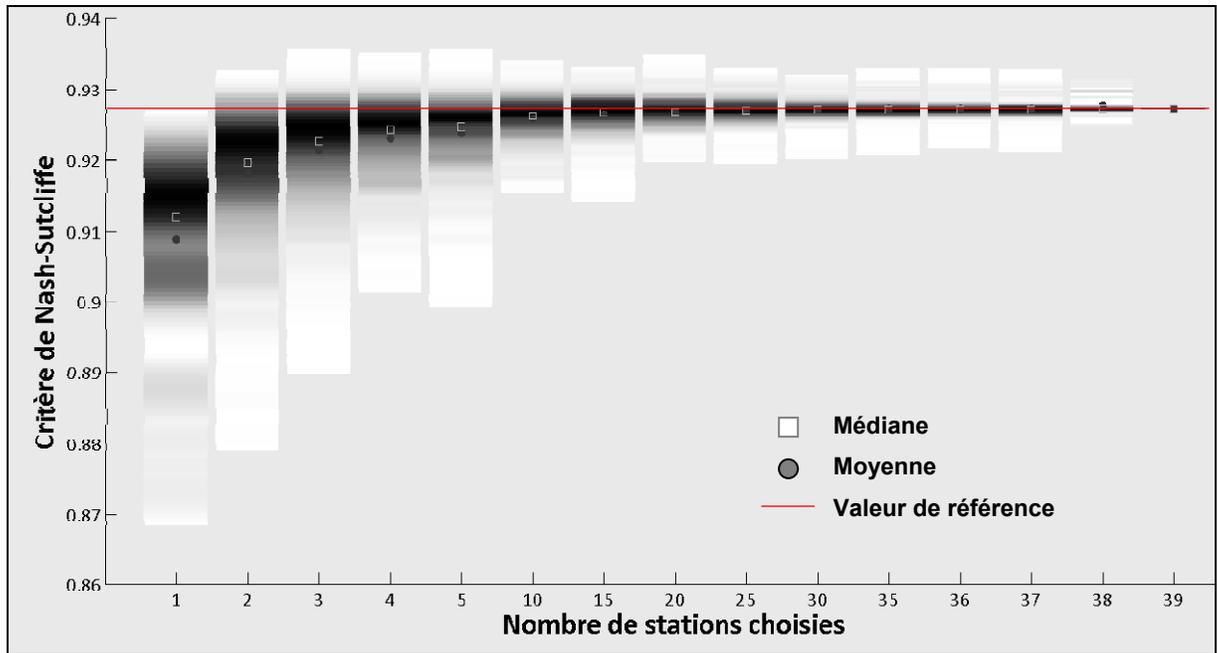


Figure 4.10 Distribution des valeurs de Nash-Sutcliffe en fonction du nombre de stations choisies en calage pour HSAMI

La Figure 4.10 démontre clairement la distribution des points pour les différents nombres de stations utilisés. Le marqueur carré représente la valeur de la médiane tandis que le point représente la moyenne de l'échantillon. La droite est placée à la valeur obtenue lors du calage à 39 stations, ce qui est la valeur de référence. Tous les points au dessus de cette ligne sont des combinaisons supérieures à la valeur de référence tout en utilisant moins de stations en mode calage. Il est également clair qu'en utilisant des stations purement au hasard, il vaut mieux utiliser plus de stations afin de réduire la variance et converger sur la valeur de référence.

#### 4.5.2 Front de Pareto sur périodes de 10 ans

Cette section regroupe les résultats obtenus lors de l'application de l'optimisateur multi-objectif sur 4 périodes distinctes de 10 années chacune. Le modèle utilisé est encore une fois HSAMI en raison des contraintes de temps.

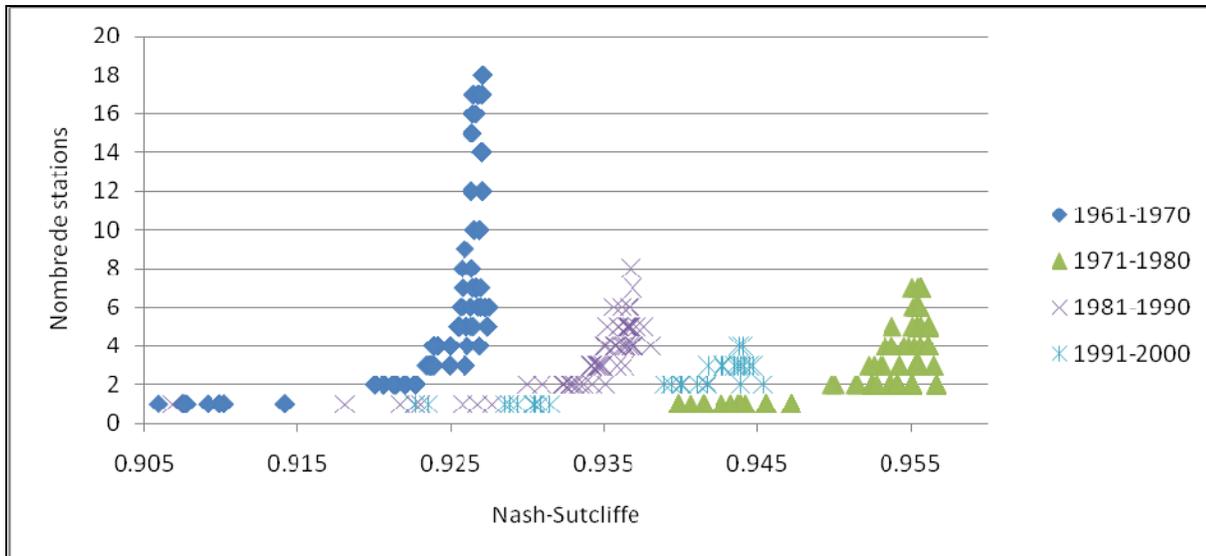


Figure 4.11 Solutions finales de l'optimisateur multi-objectif pour 4 périodes de 10 années chacune avec le modèle HSAMI

Pour chacune des périodes à la Figure 4.11, la série de points est la solution finale de l'optimisateur multi-objectif lorsque l'objectif est de minimiser le nombre de stations et de maximiser le critère de Nash-Sutcliffe. Les quatre courbes de la figure 4.11 sont indépendantes et ne devraient pas être comparées entre-elles. Chacune représente la population finale de l'optimisateur multi-objectif pour une période indépendante.

## 4.6 Emplacement des stations

L'emplacement optimal des stations a été déterminé en utilisant les populations finales des optimisations multi-objectives de la section 4.5.2. Cette section présente les résultats de chaque étape menant à la détermination des emplacements.

### 4.6.1 Fréquence d'utilisation des stations en optimisation multi-objective pour HSAMI

L'utilisation du modèle HSAMI pour l'optimisation multi-objective a permis d'effectuer des calages à grande échelle, augmentant du coup la représentativité des résultats. Pour cette

partie de l'étude, 4 périodes de 10 ans ont été utilisées indépendamment afin de vérifier le comportement de l'optimisateur selon la période choisie. La figure 4.12 démontre le nombre d'occurrences de chacune des stations dans la population finale de l'optimisateur multi-objectif. Une population de 1000 individus a été utilisée, donc le maximum possible pour une station est 1000 pour une période si elle a été utilisée dans toutes les solutions finales. Les populations finales ont été utilisées au lieu des solutions Pareto-optimales en raison du faible nombre de solutions sur le front de Pareto. En utilisant la population finale, une grande diversité de solutions meilleures que la valeur de référence (lorsque 39 stations sont utilisées) permet de mieux cerner les stations performantes.

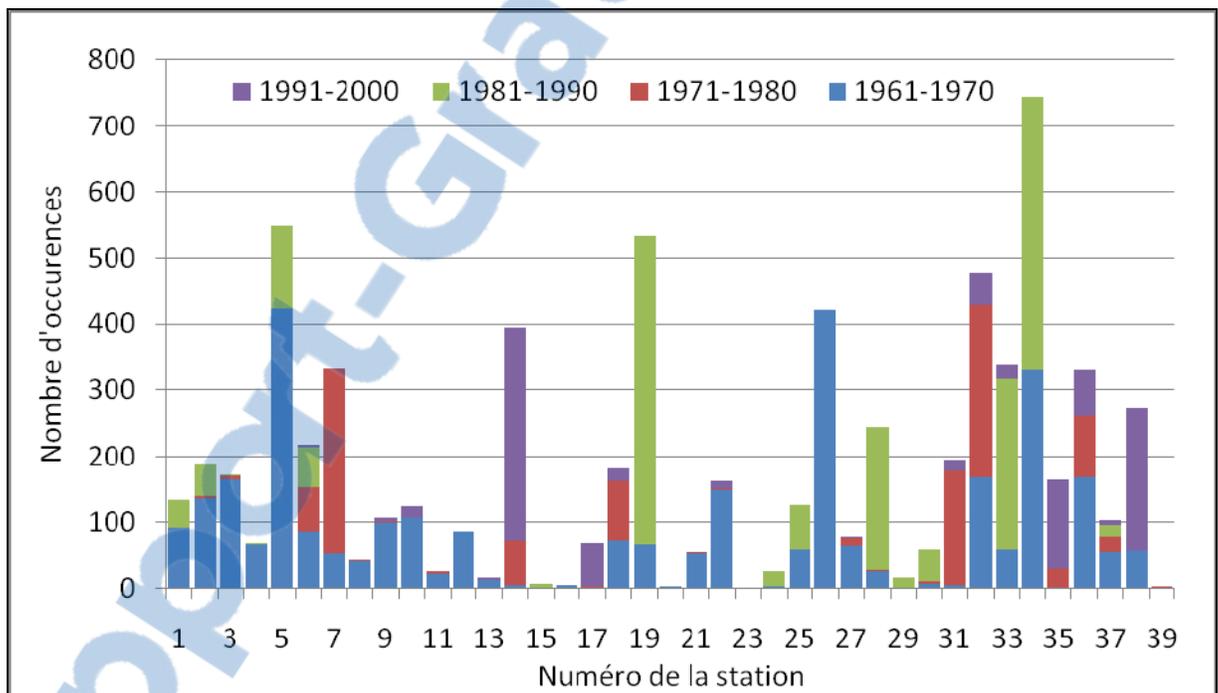


Figure 4.12 Fréquence d'utilisation des stations dans les solutions finales de l'optimisateur multi-objectif pour 4 périodes de 10 ans dans le modèle HSAMI

Pour chaque période et pour chaque station, le nombre de fois que la station fait partie de l'une des populations finales a été comptabilisé. Cet exercice démontre que les solutions optimales dépendent de la période sélectionnée et que toute méthode d'optimisation du nombre de stations à utiliser doit en tenir compte. Afin de normaliser le nombre

d'occurrences entre les périodes, la somme des occurrences totales a été normalisé pour pouvoir comparer l'utilisation relative de chacune des stations. Le nouvel histogramme pondéré est montré à la figure 4.13. Cet histogramme est utilisé pour montrer l'utilisation relative de chacune des stations, et ce, à titre indicatif seulement.

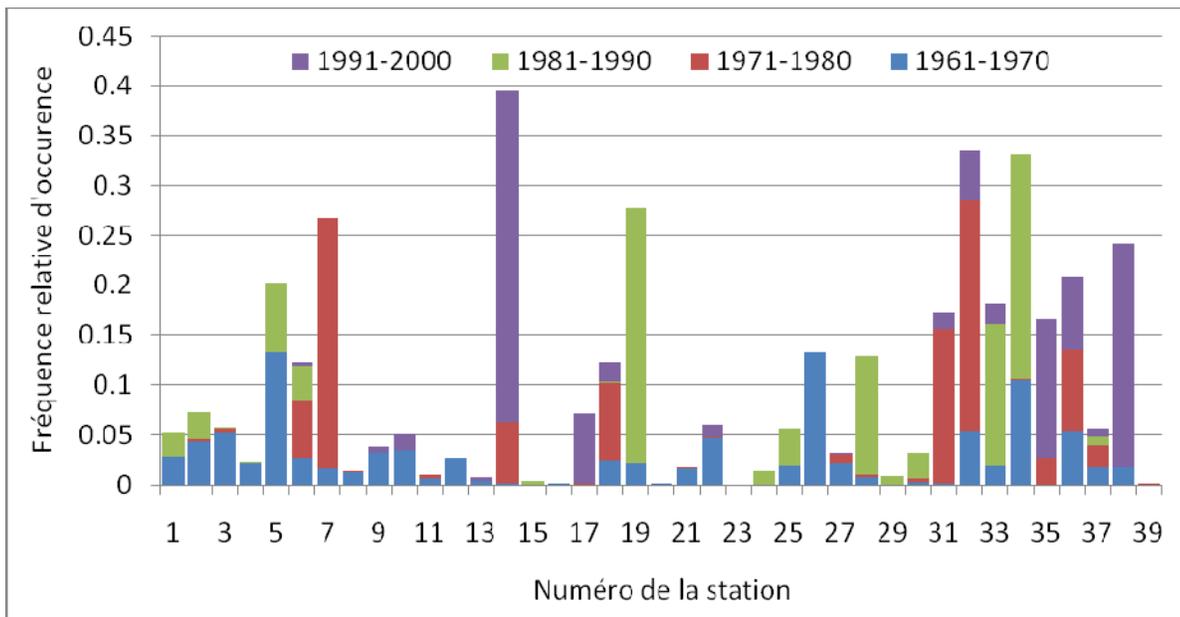


Figure 4.13 Fréquence d'utilisation normalisée des stations dans les solutions finales de l'optimisateur multi-objectif pour 4 périodes de 10 ans dans le modèle HSAMI

Les figures 4.12 et 4.13 montrent que les stations qui performant le mieux pour une certaine période ne font généralement pas partie des combinaisons les plus performantes pour d'autres périodes. Ceci peut être problématique lorsque le meilleur jeu de stations à long terme est requis et illustre le fait que la solution optimale dépend de la période. Un jeu de stations optimal à long terme devra être capable de bien performer dans toutes les périodes.

#### 4.6.2 Utilisation des meilleures stations candidates selon la somme des distributions

Afin de cibler les stations qui donnent les meilleurs résultats lorsqu'elles sont utilisées en calage, la somme de la récurrence des stations pour les quatre périodes de 10 ans mentionnées à la section 4.6.1 a été utilisée.

Cependant, la somme a été effectuée avec les données non-normalisées de la figure 4.12. Ceci assure de ne pas omettre de stations qui ont une fréquence moyenne ou faible dans chacune des quatre périodes mais qui peuvent quand même contribuer à la robustesse de la méthode étant donné leur présence dans plusieurs tranches de 10 ans. Par exemple, les stations numéros 6, 32 et 36 sont faiblement représentées dans chacune des périodes, mais lorsqu'on considère la somme des périodes, leur poids devient plus important.

Cet histogramme a ensuite été converti en carte bidimensionnelle avec contours de couleurs identifiant le poids relatif de chacune des stations, tel qu'affiché à la figure 4.14. Les couleurs chaudes signifient une plus grande importance pour les stations dans ces régions.

La nouvelle carte peut donc aider à montrer les zones qui ont offert de bonnes performances en calage sur ce bassin virtuel. Cet outil pourrait servir pour isoler des emplacements préférables lors de l'installation de stations météorologiques en se basant sur la performance du modèle dans le passé.

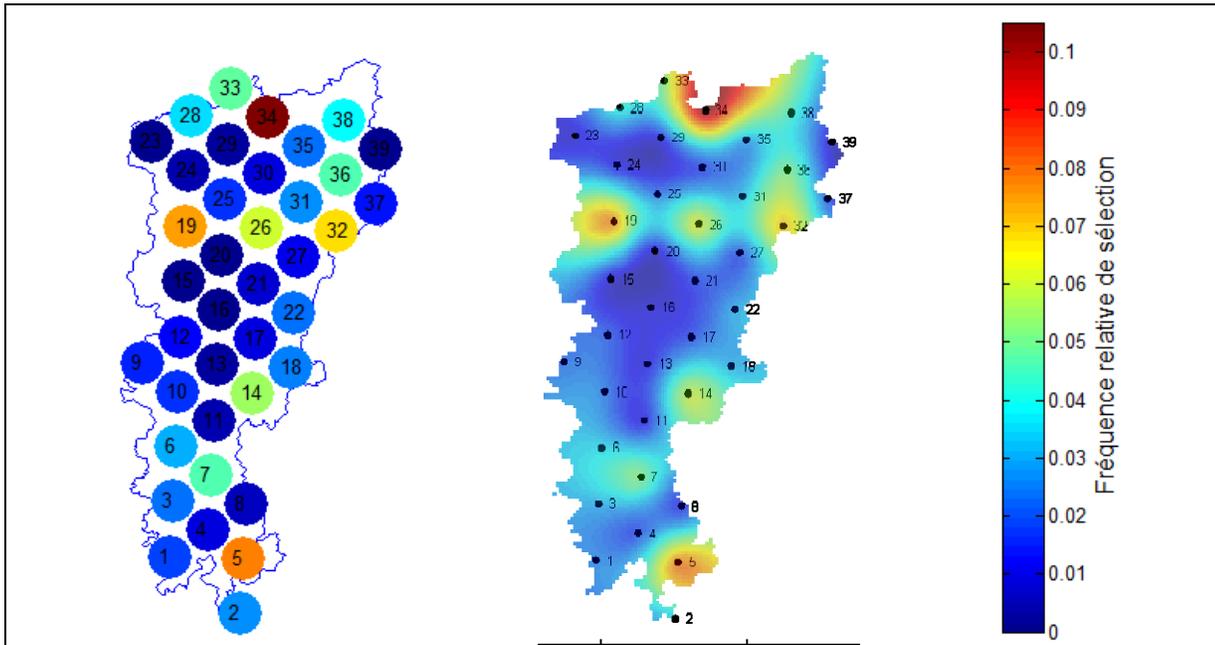


Figure 4.14 Carte du bassin versant avec les couleurs associées aux régions où le potentiel de rendement hydrologique est supérieur. La première carte montre les valeurs aux points de grille tandis que la seconde montre les valeurs interpolées entre les stations

La figure 4.14 indique quelles régions sont plus performantes pour le modèle en mode calibration. Les précipitations annuelles moyennes ainsi que les stations sélectionnées pour chaque période de 10 ans sont montrées aux figures 4.15 et 4.16.

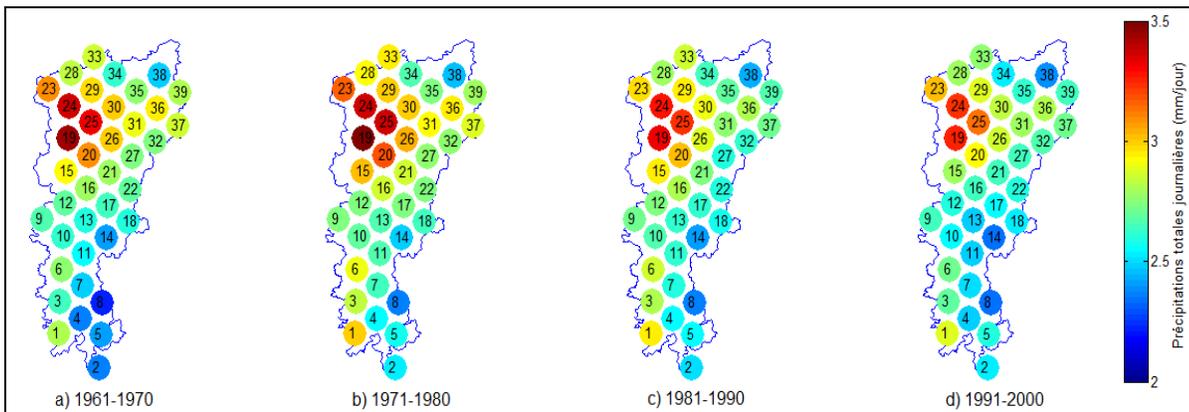


Figure 4.15 Précipitation moyenne journalière (mm/jour) pour chaque décennie de 1961 à 2000

La figure 4.15 montre la distribution de la précipitation pour les quatre décennies étudiées. Pour chacune des périodes, la section Nord-Ouest est celle qui reçoit le plus de précipitations, tandis que le sud du bassin versant en reçoit le moins.

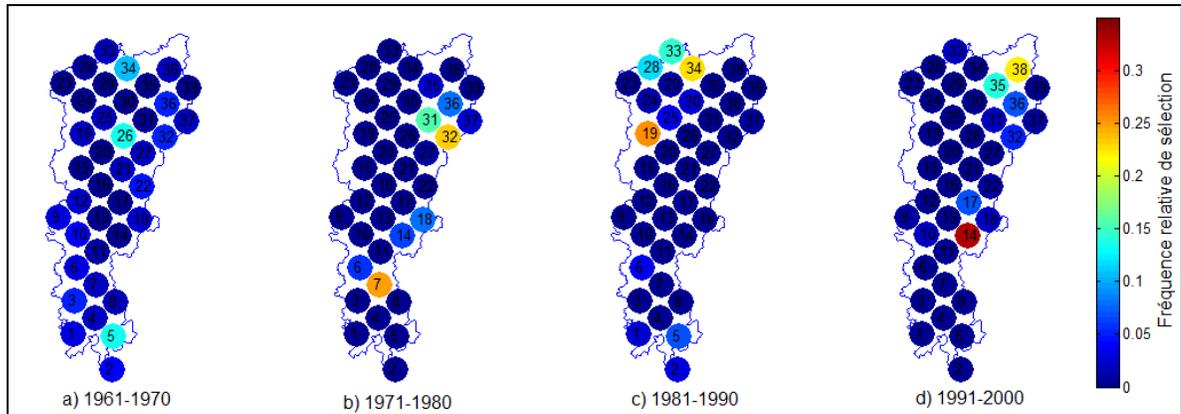


Figure 4.16 Fréquence d'occurrence des stations pour chaque décennie de 1961 à 2000

La figure 4.16 montre la fréquence relative de sélection des stations par l'algorithme d'optimisation multi-objectif. Les stations avec une fréquence de sélection non-nulle font donc partie des ensembles performants.

#### 4.6.3 Validation des résultats

Une validation de la méthode a été mise en place pour assurer que les résultats obtenus ne soient pas le fruit du hasard. Cette validation a été effectuée en 2 étapes. Premièrement, des combinaisons de stations parmi les 6 meilleures selon l'histogramme à la figure 4.12 ont été comparées à des combinaisons de stations composées aléatoirement. Cette première étape vise à comparer la performance des combinaisons de stations soigneusement sélectionnées aux combinaisons aléatoires pour déterminer laquelle des deux approches procure la meilleure performance en calage. La significativité statistique a également été testée pour quantifier les différences entre les deux groupes. La comparaison a été effectuée pour les quatre périodes de 10 ans prises indépendamment. La suite de l'essai a été de vérifier le

comportement des deux groupes lorsque les périodes utilisées le sont en contexte de validation. Les mêmes tests statistiques ont été réalisés sur ces valeurs en validation.

Les stations sélectionnées ont été choisies parmi les 6 plus importantes en termes de fréquence, soient les stations 5, 14, 19, 26, 32 et 34. Les valeurs de Nash-Sutcliffe en calage ont été calculées pour différentes combinaisons de 2, 3, 4 et 5 stations. Toutes les combinaisons possibles avec les stations sélectionnées ont été utilisées. De plus, 25 combinaisons de stations aléatoires parmi les 39 disponibles ont été créées pour chaque essai. Il y a donc 25 combinaisons de 2 stations, 25 combinaisons de 3 stations et ainsi de suite.

L'étape suivante a été de comparer la performance des différentes combinaisons une à une. Le critère de Nash-Sutcliffe pour chaque période de 10 ans avec les stations sélectionnées a été comparé au même essai avec les stations aléatoires. Les résultats pour les différentes combinaisons sont présentés aux figures 4.17 à 4.20. Les figures comparent les combinaisons sélectionnées « S » aux combinaisons aléatoires « A ». De plus, les périodes sont comparées deux à deux pour les périodes 1961-1970 « 1 », 1971-1980 « 2 », 1981-1990 « 3 » et 1991-2000 « 4 ». Ces figures comparent la performance du modèle HSAMI en calage.

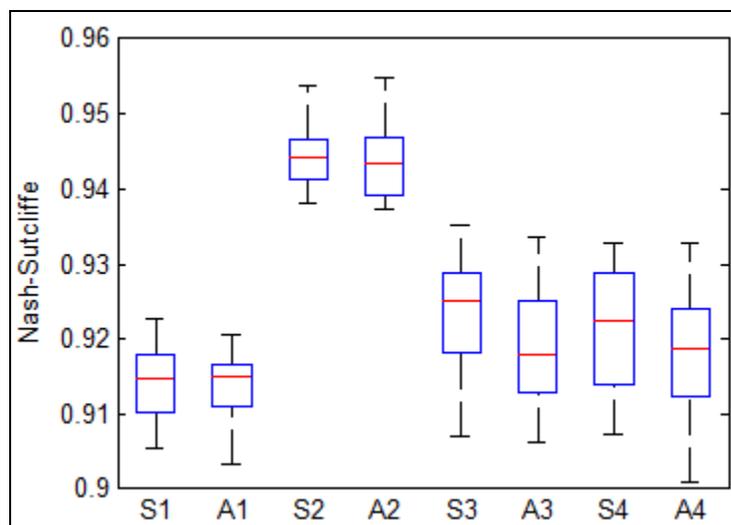


Figure 4.17 Boîte à moustaches représentant les combinaisons de 2 stations en calage

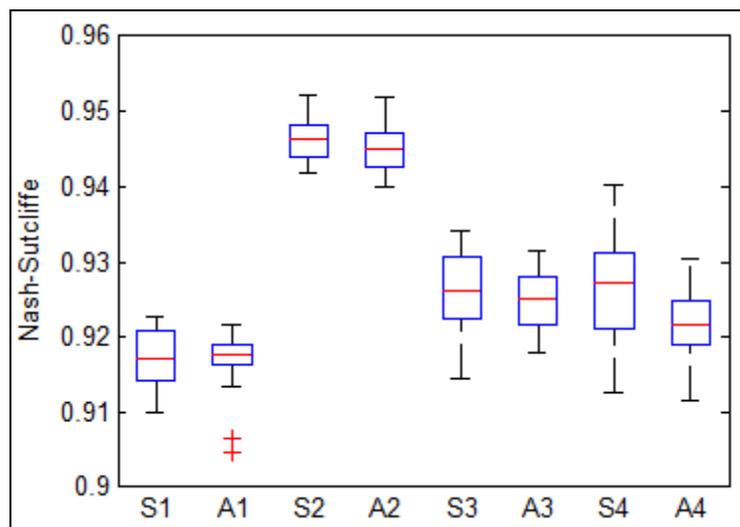


Figure 4.18 Boîte à moustaches représentant les combinaisons de 3 stations en calage

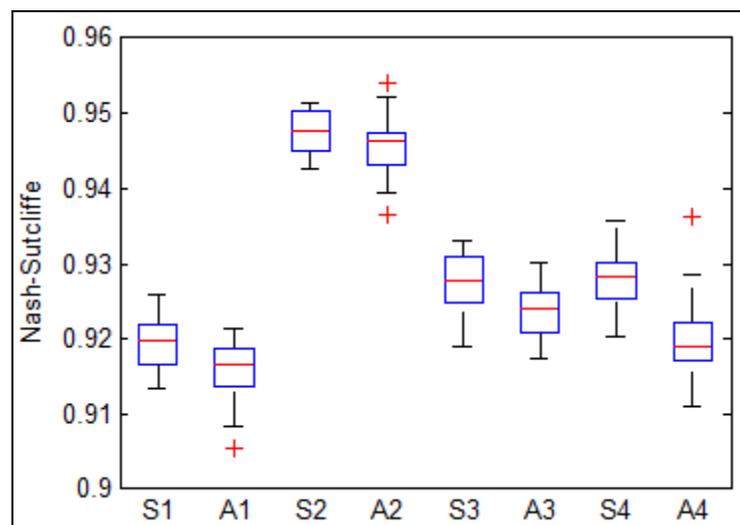


Figure 4.19 Boîte à moustaches représentant les combinaisons de 4 stations en calage

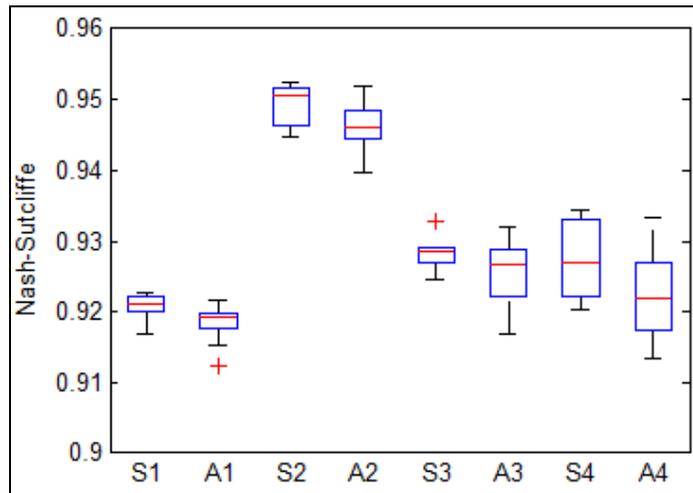


Figure 4.20 Boîte à moustaches représentant les combinaisons de 5 stations en calage

Les figures 4.17 à 4.20 semblent démontrer qu'en calage, cette méthode proposerait que d'utiliser quatre stations serait l'idéal. Elles démontrent également que la deuxième période performe mieux en général pour tous les essais. Pour les combinaisons de 4 et 5 stations, les valeurs sont généralement meilleures que pour 2 ou 3 stations, qui elles-mêmes sont meilleures que les valeurs calculées avec les stations aléatoires.

Afin de concrétiser cette tendance, le test de Kruskal-Wallis (Kruskal et Wallis, 1953) a été utilisé pour vérifier si un groupe était significativement meilleur qu'un autre dans ces comparaisons. Le test de Kruskal-Wallis est une analyse de la variance non-paramétrique basé sur les rangs des valeurs de deux ou plusieurs groupes afin de vérifier si les groupes proviennent de la même distribution. Pour ce faire, la comparaison a été établie sur la moyenne des quatre périodes. Les résultats sont consignés dans le tableau 4.2.

Tableau 4.2 P-value des tests de Kruskal-Wallis sur la performance de HSAMI en calage avec des stations aléatoires comparativement à des stations identifiées préalablement

Nombre de stations par combinaison	P-value de Kruskal-Wallis
2 stations	0.03493
3 stations	0.01121
4 stations	0.00002
5 stations	0.00807

Pour cette étude, la valeur de  $\alpha = 0.05$  a été choisie comme seuil de significativité. Un P-value inférieur à 0.05 signifie donc que l'hypothèse nulle, celle qui considère que les deux groupes ont la même médiane, est fautive. Dans chacun des cas, les stations sélectionnées avec la carte de la figure 4.14 performant mieux que les stations choisies aléatoirement.

La validation a ensuite été effectuée utilisant deux périodes différentes. Lorsque le calage était effectué sur les périodes 1961-1970 et 1971-1980, la validation a été faite sur la période 1981-2000. Lorsque le calage avait lieu entre 1981-1990 ou 1991-2000, la période de validation sélectionnée était 1961-1980. La démarche est la même que pour le calage. Les résultats sont présentés aux figures 4.21 à 4.24.

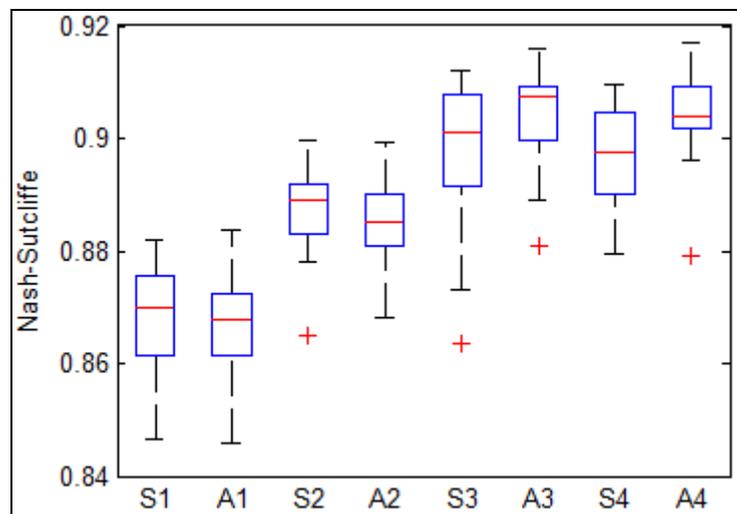


Figure 4.21 Boîte à moustaches représentant les combinaisons de 2 stations en validation

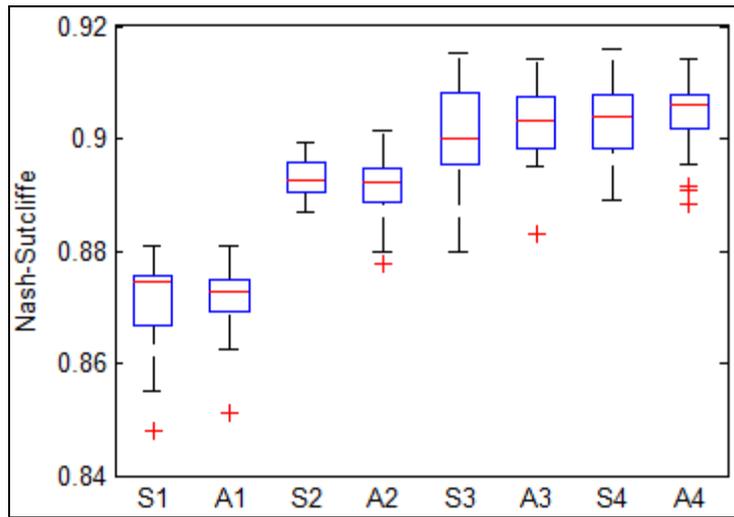


Figure 4.22 Boîte à moustaches représentant les combinaisons de 3 stations en validation

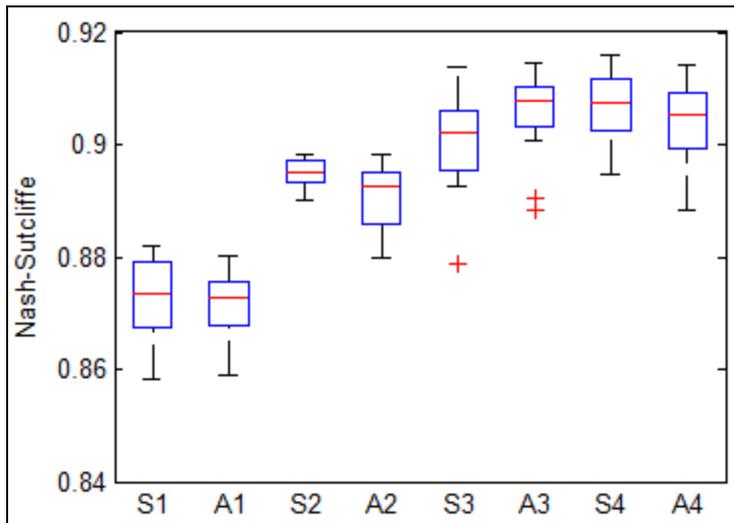


Figure 4.23 Boîte à moustaches représentant les combinaisons de 4 stations en validation

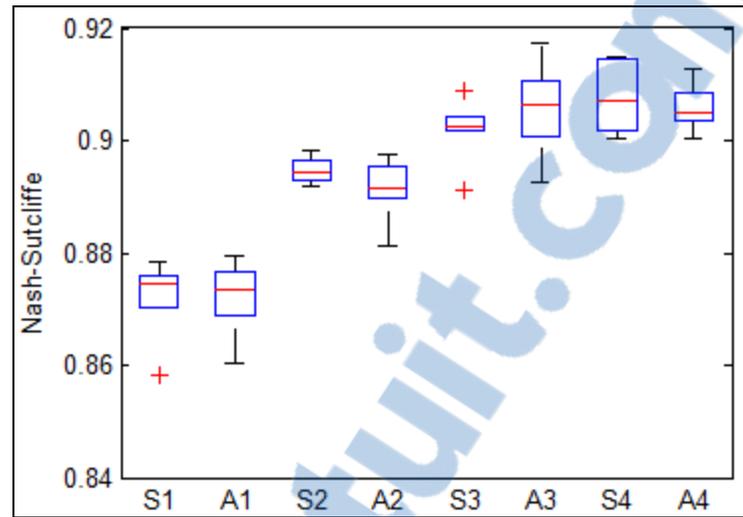


Figure 4.24 Boîte à moustaches représentant les combinaisons de 5 stations en validation

Les figures 4.21 à 4.24 semblent démontrer qu'en validation, les résultats sont plutôt mitigés. Pour certaines périodes, les stations aléatoires sont meilleures que les stations sélectionnées et vice-versa. Le test de Kruskal-Wallis a été utilisé une fois de plus pour déterminer la significativité des résultats obtenus. Les résultats sont présentés dans le tableau 4.3.

Tableau 4.3 P-value des tests de Kruskal-Wallis sur la performance de HSAMI en validation avec des stations aléatoires comparativement à des stations identifiées préalablement

Nombre de stations par combinaison	P-value de Kruskal-Wallis
2 stations	0.3493
3 stations	0.8016
4 stations	0.5115
5 stations	0.7642

Les P-value, tous supérieurs à 0.05, signifient qu'il n'y a pas lieu de rejeter l'hypothèse nulle. Les deux groupes sont donc considérés identiques ou du moins ayant la même médiane. Il est donc possible d'affirmer que malgré le fait que les moyennes soient légèrement supérieures avec les stations sélectionnées, cette différence n'est pas appréciable d'un point de vue statistique.



## CHAPITRE 5

### ANALYSE DES RÉSULTATS ET DISCUSSION

La pensée traditionnelle en hydrologie veut que tout ajout de stations météorologiques se traduise nécessairement par des gains en performance lors de la modélisation hydrologique. Les résultats de cette étude montrent non seulement que l'ajout de stations n'apporte pas nécessairement d'information supplémentaire, mais qu'en plus, il se peut qu'une station positionnée à un mauvais endroit creuse davantage l'écart entre les débits mesurés et les débits simulés par les modèles hydrologiques lors d'un calage.

Plusieurs facteurs importants sont à considérer, tant dans cette étude que dans une éventuelle application au monde réel. Cette section traite de ces facteurs et des différentes causes possibles de l'obtention de ces résultats contre-intuitifs.

#### 5.1 Interpolation et densité des stations sur une grande région

La première partie de l'étude visait à montrer les écarts importants que peuvent produire les schémas d'interpolation lorsque le nombre de stations météorologiques est restreint. Il a été prouvé que plus la densité de stations est grande, plus les données interpolées sont similaires aux données réelles. L'incertitude liée aux schémas d'interpolation est donc réduite si la densité est plus élevée. Cela ne signifie pas que le modèle hydrologique sera plus performant, mais qu'il sera plus robuste. Les résultats de la modélisation devraient donc être raisonnablement similaires aux valeurs observées en rivière.

La figure 4.3 démontre clairement la situation. La courbe montre l'évolution de la moyenne des précipitations lorsque le nombre de stations utilisées dans le calcul augmente. Plus le nombre de stations est élevé, plus la moyenne se stabilise autour de la moyenne globale. Il est à noter que l'incertitude sur la mesure de la précipitation d'un pluviomètre peut varier de  $\pm 5-15\%$ . Dans le cas de la figure 4.3, l'écart est généralement sous la barre des 2%. Ceci illustre l'avantage d'utiliser les données non-biaisées du MRCC pour cette étude, mais souligne

l'importance de tenir compte de l'incertitude lorsque la méthode sera appliquée au monde réel.

Cependant, pour obtenir des résultats acceptables, il n'est pas nécessaire d'utiliser toutes les données. Par exemple, l'utilisation d'une station sur deux permet d'obtenir approximativement le même écart-type ainsi que la même moyenne que si toutes les stations étaient utilisées. Le fondement de ce mémoire repose sur cette observation. Comment déterminer la densité optimale de stations météorologiques qui permet de simuler les débits observés à l'aide de modèles hydrologiques? Tous diraient que d'utiliser 1368 stations sur 1369 donnerait des résultats à toutes fins pratiques identiques. Le même raisonnement peut être utilisé jusqu'à une certaine limite, mais où cette limite se trouve-t-elle? La méthode proposée dans ce mémoire pourrait aider à répondre à cette question, du moins permettre de démontrer qu'en contexte de modélisation hydrologique, une limite existe bel et bien quant au nombre de stations requises.

## **5.2 Méthode manuelle**

La recherche d'une densité de stations optimale basée sur une approche manuelle a été effectuée pour atteindre les mêmes objectifs que pour les schémas d'interpolation, cette fois appliqués à la modélisation hydrologique. Le comportement des modèles a donc été vérifié lorsque la densité des stations diminue. La figure 4.5 montre l'évolution du modèle lorsque le nombre de stations utilisées augmente. Les stations choisies l'ont été avec la méthode semi-aléatoire décrite à la section 3.6.

Pour le modèle HMETTS, les résultats de la figure 4.5b ainsi que ceux présentés à la figure-A III-2 de l'annexe III montrent un plateau au niveau du critère de Nash-Sutcliffe et du RMSE. Selon ces essais préliminaires, il semble qu'il est possible de réduire le nombre de stations utilisées de 39 à 10 sans conséquence marquée au niveau de la performance du modèle en calage et en validation. Le biais, également présenté à la figure-A III-2 de l'annexe III, varie

relativement beaucoup mais les valeurs sont faibles. Il ne s'agit donc pas d'un indicateur clair d'une quelconque tendance.

Pour le modèle HSAMI, le nombre de stations minimal requis pour conserver la qualité du modèle hydrologique est situé aux alentours de 5 selon la figure 4.5a. Le jeu de stations étant sélectionné au hasard, un autre jeu aurait pu donner d'autres résultats. Cependant, il est clair que d'utiliser plus de 10 stations est superflu, tant en calage qu'en validation. Les figures pour le RMSE et le biais sont présentées à la figure-A III-1 de l'annexe III. Le biais présente une particularité notable qui le différencie des autres critères pour le modèle HSAMI. En période de calage, le biais est pratiquement nul, mais en validation sa valeur augmente quelque peu (0.005 à 0.02). Ceci indique que les critères de Nash-Sutcliffe et du RMSE sont plus importants dans le calage. Les paramètres du modèle ont été optimisés pour refléter de bonnes valeurs de RMSE et de Nash tout en ayant comme « effet secondaire » de bonnes valeurs de biais. En validation, la tendance du biais ne suit pas celle des autres critères, tout comme pour le modèle HMETS.

Pour le modèle HYDROTEL, les résultats sont sensiblement les mêmes que pour HSAMI et HMETS. La figure 4.5c montre clairement que le critère de Nash-Sutcliffe est relativement constant jusqu'à 5 stations. Il y a un creux à 10 stations, mais ceci est probablement dû à la qualité des stations sélectionnées. Hydrotel étant un modèle distribué, il pourrait être plus affecté par les stations moins représentatives du reste du bassin. Le biais relatif, qui est montré à la figure-A III-3 de l'annexe III, se comporte de manière très distincte, mais il est très faible encore une fois. Le critère pondéré et le RMSE montrent également que le nombre optimal pourrait se situer à environ 5 stations. En deçà de 5 stations, la qualité diminue au point où il serait préférable d'ajouter des stations ou de sélectionner d'autres stations qui performeraient mieux.

### 5.3 Optimisation multi-objective

L'algorithme génétique multi-objectif a été utilisé comme prolongement général de la méthode manuelle décrite à la section précédente. Cet algorithme permet de trouver les meilleures combinaisons possibles sans avoir à essayer tous les arrangements possibles. La pseudo-intelligence de la méthode fait en sorte qu'il s'agit d'un outil extrêmement utile pour cette étude.

Le principal atout de l'algorithme génétique multi-objectif est la rapidité d'exécution. Pour établir un front de Pareto de manière manuelle, il faudrait, en théorie, essayer toutes les combinaisons possibles. La méthode automatique permet d'analyser les résultats et de faire des sélections intelligentes dans les stations candidates en vue de la prochaine évaluation. Ceci permet d'alléger la charge de travail de manière phénoménale. Cependant, chacune des évaluations consiste en un calage de modèle hydrologique et plus de 25 000 évaluations sont nécessaires pour que l'optimisateur retourne des résultats acceptables. Donc, malgré la rapidité de l'algorithme d'optimisation, il aura fallu près de 10 jours de calcul par exécution d'algorithme pour HSAMI. Au total, plus de 6.5 milliards évaluations de modèle (12 optimisations) auront été effectués seulement pour l'obtention des résultats présentés dans ce mémoire. Le temps de calcul d'environ 120 jours est également négligeable comparativement à si l'opération avait été faite avec HMETs (38 années de calcul) ou Hydrotel (environ 250 000 années de calcul!).

Un des désavantages de l'algorithme est qu'il ne permet pas l'utilisation de composantes aléatoires dans la fonction objectif. Ceci signifie que lors de chacun des calages de HSAMI, si un jeu de stations quelconque retourne un critère de Nash-Sutcliffe de 0.912 lors du calage, il faut absolument que le modèle retourne encore une fois 0.912 comme critère lorsque le même jeu de stations est utilisé. Il a donc fallu fixer les « seeds » (valeurs de départ pour le générateur de nombres aléatoire) initiaux pour chacun des calages de modèle. Les algorithmes d'optimisation nécessitent des valeurs aléatoires pour la sélection des paramètres initiaux lors du processus évolutif. Si le même « seed » est fourni au modèle et que toutes les

autres conditions sont identiques (même météorologie), l'algorithme retrouvera exactement les mêmes paramètres optimaux. L'aspect stochastique normalement observé en optimisation doit donc être retiré pour cette partie de l'étude.

Ceci entraîne une contrainte quant à la portée du projet, qui est un nombre insuffisant de simulations pour analyser statistiquement les résultats. Puisque les « seeds » devaient être constants, il aurait fallu refaire l'étude en entier à plusieurs reprises avec des « seeds » différents pour chaque itération.

Il est possible que différents calages retournent des résultats légèrement différents, mais la méthode proposée fonctionne avec tous les différents essais effectués. Par exemple, la figure 4.8 démontre l'efficacité de la méthode avec le modèle Hydrotel sans calage, tandis que les figures 4.9 et 4.10 mettent en évidence le potentiel de l'algorithme avec le modèle HSAMI avec recalage. Il n'y a aucune raison de croire que les modèles ne se comporteraient pas de la même manière avec des seeds différents.

#### **5.4 Robustesse de la méthode**

Étant donné la nature exploratoire de cette étude, la robustesse de la méthode n'a pas été mise à l'épreuve en l'essayant sur d'autres bassins ou sous d'autres conditions. Une seule base de données a été utilisée (MRCC15) et seul le modèle HSAMI a été utilisé pour les essais avec recalage.

Cependant, à plusieurs reprises la méthode a retourné des résultats plus qu'intéressants. Par exemple, l'utilisation des paramètres optimaux pour 39 stations en réduisant le nombre de stations utilisées (figures 4.6 à 4.8) a permis de trouver des centaines de combinaisons meilleures que la valeur de référence. Ceci est contre-intuitif, surtout pour un modèle distribué comme Hydrotel. Le fait qu'il y a moins d'informations ne permet pas seulement de conserver la performance hydrologique du modèle, mais améliore sa performance en entier. Les gains peuvent sembler limités (les valeurs absolues sont de l'ordre de 0.005 à 0.04), mais

lorsque le modèle performe déjà très bien (NS de 0.93), il est normalement difficile d'augmenter davantage sa performance.

Le même cas a été constaté avec les recalages d'HSAMI. Il aurait été sage de supposer que plus il y a de stations, mieux serait la simulation. Ceci a été contredit par les recalages d'HSAMI. Cette fois, en utilisant toute l'information disponible et en calant le modèle de surcroît, la performance du modèle hydrologique a été améliorée dans tous les cas possibles. Pour chacune des périodes sélectionnées (4 périodes de 10 ans, période de 40 ans) et avec les différents artifices de production d'intrants météorologiques (Interpolation linéaire, moyenne, Krigeage pour Hydrotel), les résultats ont démontré qu'il y a toujours moyen de réduire le nombre de stations météorologiques en ne causant aucune conséquence sur la performance de modélisation en calage. En fait, il a été montré qu'il est possible de réduire le nombre de stations et améliorer la performance du modèle. Le même exercice a été démontré avec les modèles HMETs et Hydrotel, tel qu'il a été mentionné à la section 4.4.

Il est à noter que cette méthode fonctionne pour les données virtuelles utilisées. Dans le monde réel, contrairement au monde virtuel, il y a des incertitudes liées aux données observées. Il y a notamment des erreurs de lecture des équipements, le vent cause des sous-estimations des précipitations, la neige au sol est difficile à quantifier, etc. Même les apports observés en rivière sont approximés et peuvent être jusqu'à 25% trop faibles ou trop élevés, dépendamment du débit et du type d'équipement (Baldassarre & Montanari, 2009). Toutes ces sources d'erreur rendent le processus beaucoup plus complexe dans le monde réel. Dans le monde virtuel dans lequel ce projet est imbriqué, il n'y a aucune erreur de mesure et toutes les variables météorologiques sont cohérentes. De plus, la méthode de routage assure une certaine représentativité pour que le modèle puisse simuler des débits acceptables.

Il est donc possible d'affirmer que cette méthode fonctionne sous certaines conditions dans le monde virtuel, mais qu'il faudrait tenir compte des incertitudes dans le monde réel avant d'en mettre à l'épreuve la robustesse et de pouvoir en faire une application directe.

## 5.5 Différences et similitudes entre les modèles

Les trois modèles hydrologiques se comportent de la même manière dans cette étude. Dans tous les cas, le calage avec toutes les stations disponibles donne des résultats acceptables. De plus, les modèles peuvent mieux performer avec moins de stations, tel qu'il a été montré à la figure 4.5 et à la section 4.4. Bien qu'il n'ait pas été possible de caler les modèles HMETS et Hydrotel à chaque nouveau jeu de stations, les expériences faites sur ces deux modèles ont clairement démontré que même en gardant un jeu de paramètres unique, il est possible d'améliorer leur performance en modifiant simplement les stations utilisées lors de la modélisation.

De plus, l'expérience menée sur HSAMI avec les recalages à chaque nouveau jeu de stations a permis de montrer qu'il était possible d'améliorer davantage la valeur du critère de Nash-Sutcliffe. Le même scénario a été prouvé avec Hydrotel à la figure 4.15, lorsque le meilleur jeu de deux stations avec paramètres fixes a été utilisé dans un recalage du modèle. L'amélioration du critère de Nash-Sutcliffe de 0.02 alors qu'il est déjà très acceptable démontre clairement qu'il y a place à amélioration par rapport à l'utilisation de toutes les stations. Ces résultats sont d'autant plus remarquables pour Hydrotel étant donné le principe de distribution physique des stations utilisée par le modèle.

## 5.6 Résultats globaux

Tous les essais effectués donnent des résultats très parlants. Il semblerait qu'une couverture totale du bassin peut non seulement mener à une stagnation de la performance du modèle, mais peut également réduire sa performance. En effet, les multiples fronts de Pareto avec différentes périodes et les différents modèles montrent qu'un nombre relativement restreint de stations peut permettre une modélisation plus performante qu'avec 39 stations. La figure 4.9 est la plus parlante à cet égard : il existe un grand nombre de combinaisons de moins de 10 stations qui retournent des résultats meilleurs que lorsque les 39 stations sont utilisées. Les combinaisons de 2 et 3 stations ont toutes été vérifiées et plusieurs d'entre elles sont effectivement meilleures que si la totalité des stations avait été utilisée. Le même

raisonnement peut être appliqué aux figures 4.6 à 4.8. Le modèle Hydrotel peut performer beaucoup mieux avec les deux stations formant le point optimal sur le front de Pareto. À preuve, les stations ont été trouvées en gardant le même jeu de paramètres trouvé lors du calage du modèle avec 39 stations. Un recalage sur ces deux stations a permis d'améliorer davantage la performance du modèle. Étant donné les contraintes de temps, il a été impossible de trouver les stations optimales et il est fort possible que plusieurs combinaisons de stations auraient pu permettre d'augmenter encore plus la performance du modèle en le recalant.

Les résultats montrés à la figure 4.10 démontrent clairement qu'une augmentation du nombre de stations est la seule façon de réduire les biais sur la lecture des données et ainsi réduire la variance de la performance des modèles hydrologiques. Il est aussi discernable que d'utiliser un faible nombre de stations aléatoirement choisies prédispose à obtenir de mauvaises performances en calage. Cependant, si les bons jeux de stations sont utilisés, il est possible d'obtenir de bonnes performances en calage en n'utilisant que très peu de stations. Il a également été démontré que les stations optimales varient selon la période choisie, mais qu'il y a des stations qui sont représentées dans plusieurs périodes. Celles-ci sont automatiquement pondérées plus fortement dans le calcul final sur toute la période.

Un autre point important est celui de l'échelle temporelle utilisée. Dans le cadre de cette étude, les séries utilisées étaient longues, de l'ordre de 10 années ou plus. Dans le cas d'orages estivaux, l'événement pluvieux peut durer moins de quelques heures sur le bassin versant. Il est donc évident que pour ces orages localisés, une utilisation restreinte de stations pourrait générer des biais dans les quantités d'eau disponible. Pourtant, lorsque les données sont étudiées à long terme, les apports quotidiens sont très bien simulés même avec un faible nombre de stations. Le critère de Nash-Sutcliffe utilisé ajoute plus de poids aux crues et aux périodes de fort débit, mais les étés et les périodes d'étiage sont tout de même pris en compte dans le calcul. Puisque ces critères peuvent être aussi bons avec 3 qu'avec 39 stations, il est clair que les stations choisies doivent être en mesure de prendre en considération les précipitations estivales. Cependant, cette étude ne montre pas clairement l'avantage d'utiliser

un petit nombre de stations pour des études hydrologiques de courte durée, telle que les étiages ou l'étude des événements pluvieux de courte durée et de haute intensité (orages).

L'information principale découverte dans cette étude est donc qu'il est possible d'utiliser un nombre restreint de stations, tant qu'elles sont placées intelligemment et de manière à capter le maximum d'information sur le bassin versant en vue de la modélisation hydrologique.

## **5.7 Explications possibles de la bonne performance des modèles avec une faible densité de stations**

Les résultats trouvés dans cette étude démontrent qu'il est possible d'utiliser un nombre restreint de stations et d'obtenir tout de même de bonnes performances en modélisation hydrologique distribuée et globale. Ces résultats contrastent avec les normes généralement acceptées en hydrologie et en météorologie. Cependant, il peut y avoir des raisons pour ce comportement. Cette section expose les différentes pistes possibles.

### **5.7.1 Explications générales**

Les essais effectués avec HSAMI ont clairement démontré le bienfondé de cette méthode. De plus, les modèles HMETS et Hydrotel semblent se comporter de manière semblable à HSAMI, ce qui signifie que la méthode devrait également fonctionner pour ces modèles. Si l'optimisateur multi-objectif avait pu être mis en place pour Hydrotel et HMETS, les stations trouvées auraient fort probablement été différentes de celles trouvées pour HSAMI, mais il est clair que le nombre de stations serait du même ordre de grandeur que pour ce dernier. L'approche manuelle a permis de déterminer que 10 stations sélectionnées au hasard permettaient de retrouver la même performance qu'avec 39 stations, alors que l'optimisateur multi-objectif permet de sélectionner des stations de manière optimale. Ce dernier a fait ses preuves pour tous les modèles lorsque leurs paramètres ont été fixés et que le nombre de stations a été varié. La tâche de trouver quelles stations à utiliser pour les modèles plus complexes nécessitera sans aucun doute de la programmation parallèle à haute performance.

Un des mystères restants est pourquoi un si faible nombre de stations peut être meilleur, ou du moins aussi bon, que lorsque la densité de stations est très élevée. Premièrement, il y a certainement une portion de chance dans l'équation. La moyenne de la météorologie de ces stations est donc ce qu'il y a de plus près de la perfection pour les apports observés. Cependant, les stations choisies sont inter-reliées entre-elles et avec l'écoulement de surface et souterrain produit par le MRCC. Il se pourrait également que les stations sélectionnées soient simplement une simplification spatiale de la météorologie du bassin versant. Dans le cas où plusieurs stations rapprochées sont très similaires, le fait d'utiliser celle qui représente le mieux cette portion du bassin pourrait être sélectionnée. Ceci ne réduirait pas la performance en modélisation hydrologique mais aurait comme impact de réduire considérablement le nombre de stations à utiliser en premier lieu. Ceci est facilement compréhensible pour les modèles qui moyennent les stations sur le bassin (HSAMI et HMETs), mais pour Hydrotel, la structure distribuée peut également utiliser cet artifice. Il suffit de séparer le bassin versant en sous-sections et d'utiliser la station la plus représentative de cette section, simulant ainsi des grappes d'UHRH assez similaires pour être moyennées entre-elles. Les essais montrés à la section 5.7.3 vont aussi dans le même sens.

Il est important de rappeler que les résultats obtenus sont valides dans le monde virtuel qui a été utilisé. Les patrons de température et de précipitation devraient être comparés aux observations avant de les appliquer dans le monde réel. Ceci nécessiterait une base de données complète qui n'est pas disponible. Par contre, si quelques stations sont disponibles sur un bassin versant quelconque pour une période de temps assez longue, il pourrait être possible de valider les similitudes entre les quantités de précipitations simulées par le MRCC et les observations.

### **5.7.2 Qualité du routage**

Tel que mentionné à la section 3.3.2, le schéma de routage utilisé ne devrait pas avoir d'influence notable sur les résultats. Pour tenter d'expliquer les résultats, la méthodologie a tout de même été modifiée afin d'utiliser 10 zones de routage au lieu des trois zones initiales.

Le raisonnement derrière cette idée vient du fait qu'il y a peut-être une perte importante d'information lors de la prise de la moyenne des points dans chacune des zones lors du routage. En divisant le bassin en 10 zones, chacune d'entre elles contient 4 stations au lieu de 13. Cette augmentation de résolution pourrait changer quelque peu les valeurs obtenues. Le même schéma de routage décrit à la section 3.3 a été utilisé, et les résultats obtenus pour le routage à 10 zones comparativement à 3 zones sont présentés à la figure 5.1.

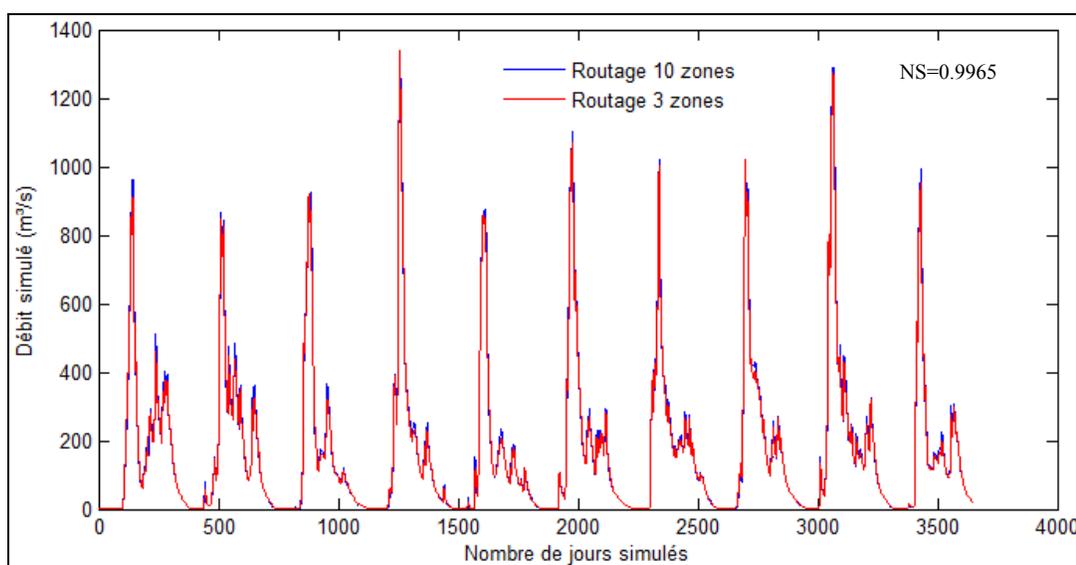


Figure 5.1 Débits reconstitués à partir des écoulements de surface et souterrains du MRCC lorsque le bassin est divisé en 3 zones et en 10 zones

L'écart entre les courbes est excessivement faible. Quelques pointes sont ratées et les courbes sont parfois décalées de très peu, mais les débits sont généralement très similaires. Le critère de Nash-Sutcliffe entre les deux courbes est de 0.9965, ce qui signifie que l'écart entre les courbes est à toutes fins pratiques inexistant. L'utilisation du débit reconstitué à partir du schéma de routage à 10 zones dans les modèles ne devrait changer en rien leur performance.

### 5.7.3 Bruit et faibles différences entre les stations

Une autre piste potentielle est que les données des stations sont très similaires entre-elles et que le moindre bruit ajouté aux données viendrait bouleverser les résultats. Un essai dans ce

sens a été effectué, en ajoutant un bruit blanc avec 2% d'écart aux données. Chacune des données pouvait donc être modifiée par un facteur variant entre +2% et -2%. Ceci ne change pratiquement rien en modélisation pour ce qui est de la température, mais l'impact sur la précipitation devrait changer la donne si cette hypothèse s'avérait vraie. Une limite de 2% a été sélectionnée pour ne pas trop changer les données tout en assurant un changement non-négligeable dans les données aux stations. Cependant, les stations météorologiques ont un biais plutôt systématique tandis que le biais appliqué dans cette partie de l'étude est aléatoire. Un bruit aléatoire de 5% ou de 10% modifierait probablement l'ordre des stations de manière aléatoire également, mais ceci n'a pas été vérifié.

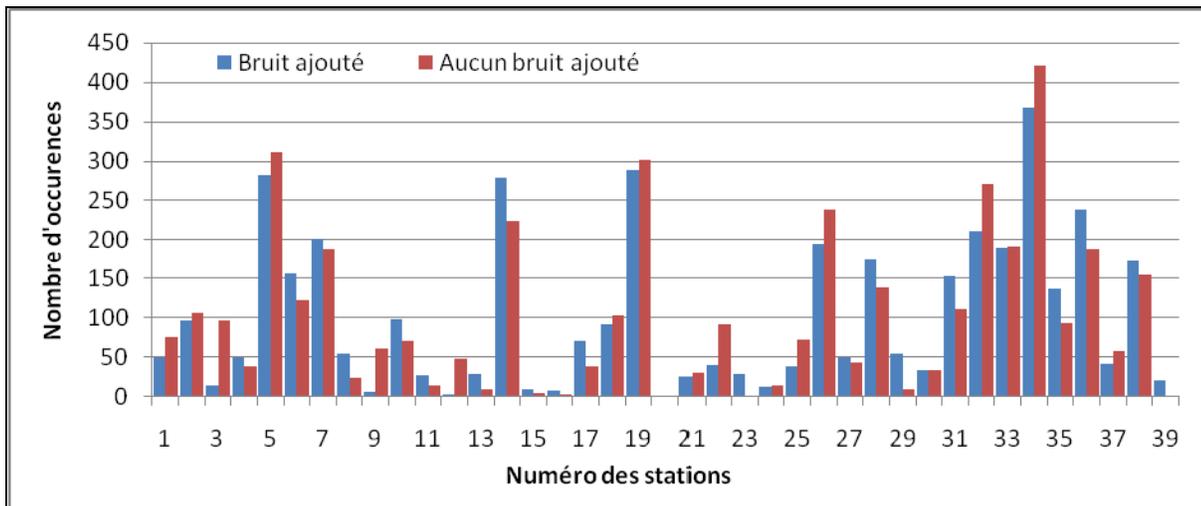


Figure 5.2 Comparaison entre les stations sélectionnées avec et sans bruit ajouté aux données

La figure 5.2 montre qu'il y a effectivement de légères modifications à l'ordre établi à la section 4.6.2, mais que les grandes lignes tiennent toujours. Les mêmes stations (5, 14, 19, 26, 32 et 34) sont les meilleures après que le bruit ait été ajouté, mis-à-part la station numéro 7 qui est pratiquement équivalente à la station 26. Évidemment, plus il y a de bruit, plus les stations sont dénaturées. Éventuellement, avec assez de bruit, les stations n'auraient plus de corrélation avec les écoulements observés. L'ajout de bruit équivalent à l'incertitude sur les mesures des stations pourrait être utilisé pour vérifier l'applicabilité de la méthode dans le monde réel.

## 5.8 Réponse à la problématique initiale

Le but du projet était de déterminer la densité optimale de stations météorologiques en contexte de modélisation hydrologique distribuée et globale. Cette étude démontre qu'il est possible d'utiliser un nombre relativement restreint de stations tout en préservant la qualité de la modélisation hydrologique de HSAMI, HMETS et Hydrotel. Les analyses ont montré qu'il serait possible de se tirer d'affaire avec deux à trois stations, mais qu'il y a un gain à faire en utilisant jusqu'à cinq stations sur le bassin de la rivière Tournustouc. Ceci correspond à une densité variant de 1 station par 1600 km<sup>2</sup> à 4000 km<sup>2</sup>. Le « WMO Guide to Hydrological Practices 2008 » considère que la densité devrait se situer entre 250 et 575 km<sup>2</sup> par station. À l'échelle du bassin versant étudié, il aurait fallu entre 14 et 33 stations sur le bassin pour être en accord avec l'OMM. En réalité, comme il a été démontré au chapitre 3, le bassin de la rivière Tournustouc est fortement sous-représenté spatialement et temporellement. Pour avoir une bonne performance en modélisation hydrologique sur le bassin, il faudrait y installer de nouvelles stations.

Cette étude montre comment cette opération pourrait être faite si le nombre de stations qu'il est possible d'installer est faible. Plutôt que d'installer les stations n'importe où, la figure 4.14 peut être utilisée pour viser des endroits à haut potentiel. Cependant, dans le monde réel, il ne serait pas possible de garantir la même efficacité que dans le monde virtuel en raison des incertitudes qui y sont présentes. Une partie des objectifs de l'étude a donc été atteint puisque la densité du réseau d'observations a été déterminée pour plusieurs circonstances, mais il faudra pousser l'étude plus loin pour généraliser les résultats en validation, aux autres bassins et, ultimement, au monde réel.

Il a également été démontré que plus le nombre de stations en place est élevé, plus la variance dans la performance du modèle est faible (voir figure 4.10). L'information à tirer de ce résultat est que dans le cas où les ressources financières ne sont pas limitées, l'installation d'un grand nombre de stations permettra d'obtenir des résultats plus près de la réalité, sans égard à l'emplacement des stations. Cependant, si tel n'est pas le cas, un emplacement

stratégique de peu de stations pourrait permettre d'obtenir des résultats semblables mais avec une variance plus grande. Une erreur d'emplacement d'une station si le réseau en comporte deux produira nécessairement plus d'erreurs que si le réseau en contient 30. Dans cette optique, pour l'instant, il est judicieux d'utiliser le plus de stations possibles dans une application réelle où il y a des erreurs de mesure et des biais.

Par contre, pour ce qui est d'une éventuelle application dans le monde réel, il est peu probable qu'une région de la taille de la moitié du bassin versant soit sans épaisseur de sol, tel qu'illustré à la figure 3.4. Il y a donc une forte probabilité que l'application des résultats de cette étude dans le monde réel ne donne pas les résultats escomptés. La méthode décrite est bel et bien valable, mais la différence entre le monde virtuel et le monde réel à l'égard du modèle de sol est trop grande pour assurer une réponse similaire. Un meilleur modèle de sol permettrait de mieux simuler le bassin versant réel dans son ensemble.

## **5.9 Vers une nouvelle approche pour la recherche en hydrologie**

La méthode proposée pourrait être utilisée partout où les données du MRCC15 (ou tout autre modèle régional à haute résolution) sont disponibles. De ce fait, des études approfondies sur les bassins versants pourraient être menées dans l'environnement virtuel. Des cartes produites à partir des ensembles Pareto-optimaux pour chacun des bassins versants pourraient être produites et utilisées pour identifier les zones à haut contenu informatif pour la modélisation hydrologique.

Si la méthode s'avère applicable dans le monde réel, il serait évidemment très pratique de pouvoir n'installer que quelques stations bien choisies et savoir que les informations recueillies sont les meilleures possibles pour un investissement donné. De plus, une amélioration des positions des stations hydrométéorologiques pourrait éventuellement permettre à des modélisateurs de travailler avec des jeux de données en sachant que ces stations représentent bel et bien la météorologie locale. La méthode utilisée dans cette étude pourra éventuellement être poussée plus loin en utilisant des MRC de meilleure résolution et en améliorant leurs bases de données de type et d'occupation du sol.

### **5.10 Estimé des retombées**

Les retombées engendrées par l'application de la méthode décrite dans ce projet de recherche sont potentiellement nombreuses. Premièrement, les avantages financiers sont imposants. Tout gestionnaire désirant réduire l'incertitude quant à la qualité des données hydrométéorologiques avec l'ajout de stations météorologiques devra éventuellement faire face au problème de l'emplacement de la station. De plus, le nombre de stations à utiliser doit être déterminé avec soin. L'utilisation d'un modèle de climat régional à haute résolution devrait permettre, à terme, de déterminer le nombre optimal de stations et leur emplacement sans avoir à dépenser des sommes considérables pour des stations temporaires ou des stations pilotes. La tendance dans l'amélioration des MRC montre qu'il sera probablement possible d'établir de tels liens dans un futur rapproché.

Le second axe de retombées est scientifique. La quantité d'informations qui pourront être mises à la disposition des chercheurs est très grande. De plus, l'utilisation d'un schéma de routage comme il a été question dans cette recherche pourra être utilisé par d'autres hydrologues pour trouver des réponses à des questions restées sans réponse dans le domaine de l'hydrologie. La réduction des biais et de l'incertitude dans les données d'entrée aux modèles améliore la fidélité et la robustesse de ces derniers. En effet, les variables d'état des modèles pourraient être comparées aux variables simulées par le MRCC, validant ou infirmant ainsi les processus internes des modèles. La présente étude a démontré la faisabilité d'utiliser des modèles hydrologiques dans le monde virtuel et a prouvé que l'efficacité du modèle est augmentée dans le monde virtuel.



## CONCLUSION

La présente étude a permis de mettre sur pied une méthode pour déterminer le nombre ainsi que l'emplacement optimal de stations météorologiques sur un bassin versant donné. En utilisant la base de données MRCC15, il a été possible de créer un environnement virtuel riche en données qui ont servi à établir un réseau d'observations météorologiques dense. Il a été démontré que la zone d'optimalité se situe entre 1 station par 1600 à 4000 km<sup>2</sup>. L'emplacement de ces stations est intrinsèquement lié à leur contribution à la performance en modélisation hydrologique.

Il a également été démontré qu'il est toujours mieux d'utiliser toute l'information disponible si on ne connaît pas la valeur de celle-ci à un endroit donné, et donc d'installer le maximum de stations lorsque l'emplacement est sélectionné de manière aléatoire. Toutefois, en fonction des contraintes financières, il vaut mieux placer les stations aux endroits où l'information qu'elles obtiendront aura plus de valeur d'un point de vue hydrologique. La méthode proposée dans cette étude pourrait servir à cet égard.

De plus, ces résultats sont valides pour une utilisation à long terme de l'hydrologie régionale. Les résultats pourraient être différents si le but final était de capter le ruissellement dû à une cellule orageuse. Cependant, sur le long terme, cette méthode semble prometteuse pour l'avenir. Au fur et à mesure que les incertitudes liées aux mesures de précipitation et de débit seront réduites, cette méthode sera de plus en plus applicable avec de bons résultats. Il faudra s'assurer que le bassin versant et les processus climatiques soient bien modélisés dans le MRCC ou le modèle régional de climat utilisé pour avoir une chance d'en faire une utilisation dans le monde réel. L'aspect virtuel de la méthode présentée dans cette étude fera en sorte qu'il sera possible de caractériser la météorologie locale du bassin sans avoir à déboursier des sommes pour l'achat, l'installation et l'entretien de dizaines de stations météorologiques temporaires.

Finalement, cette étude a permis de valider l'applicabilité des données du MRCC à 15 km de résolution en modélisation hydrologique. La science de l'hydrologie pourra définitivement évoluer pour mieux comprendre ses fondements en utilisant un monde virtuel, comme le fut le cas dans cette étude.

## ANNEXE I

### STATIONS MÉTÉOROLOGIQUES

Cette annexe regroupe les particularités des stations météorologiques présentes sur le bassin versant. Les noms et numéros des stations, leurs emplacements géographiques respectifs ainsi que les périodes pendant lesquelles les stations ont été actives sont y sont compilées.

Tableau-A I-1 Propriétés des stations météorologiques existantes sur le bassin versant de la rivière Touloustouc

<b>Numéro et nom de la station</b>	<b>Latitude (deg) / Longitude (deg) / Altitude (m)</b>	<b>Périodes couvertes</b>
7046297 – QCMR MILE 106	51.07 / -67.5 / 485.2	08/1963
7046298 – QCMR MILE 146	51.55 / -67.43 / 582.2	12/1963 à 01/1964 08/1964 à 11/1964
704A0J9 – CAOTIBI	50.72 / -67.52 / 609.6	06/1977 à 09/1977 06/1978 à 09/1978 06/1979 à 09/1979 06/1980 à 09/1980
704HE0G – TOULNUSTOUC	50.08 / -67.95 / 274.3	06/1973 à 09/1973 06/1974 à 09/1974 06/1975 à 09/1975 06/1976 à 09/1976
704J01E – CAMP M-107	51.07 / -67.48 / 487.7	06/1974 à 12/1974 04/1975



## ANNEXE II

### PARAMÈTRES DES MODÈLES HYDROLOGIQUES

Cette annexe présente les paramètres des trois modèles hydrologiques utilisés dans cette étude, soient Hydrotel, HMETs et HSAMI, ainsi que les bornes de calage.

Tableau-A II-1 Paramètres et bornes de calage pour le modèle Hydrotel

Numéro	Description des paramètres	Borne inférieure	Borne supérieure
1	Température de passage pluie-neige	-6	2
2	Taux de fonte pour conifères	0.01	12
3	Seuil de fonte pour conifères	0	5
4	Taux de fonte à découvert	0	6
5	Seuil de fonte à découvert	-5	5
6	Coefficient multiplicatif de l'ETP	0.7	1.2
7	Bilan Vertical – 3 Couches (BV3C) zone 1	0.05	1
8	Bilan Vertical – 3 Couches (BV3C) zone 2	0.05	2
9	Bilan Vertical – 3 Couches (BV3C) zone 3	0.05	5
10	Bilan Vertical – 3 Couches (BV3C) coefficient de récession	1e-8	1e-4
11	Taux de fonte pour feuillus	0	6
12	Seuil de fonte pour feuillus	0	5
13	Hydrogramme Géomorphologique (HGM)	0.0002	0.04
14	Coefficient de Manning pour la forêt ( $n_f$ )	0.1	0.8

Tableau-A II-2 Paramètres et bornes de calage pour le modèle HMETS

<b>Numéro</b>	<b>Description des paramètres</b>	<b>Borne inférieure</b>	<b>Borne supérieure</b>
1	Temps de pointe de l'hydrogramme de surface (Nombre entier)	1	4
2	Temps de base pour l'hydrogramme de surface (Nombre entier)	2	15
3	Facteur de forme de l'hydrogramme de la zone vadose	1	5
4	Facteur degré-jour minimum	0	20
5	Différence entre ddf_min et le facteur degré-jour maximum	0	20
6	Température de base pour la fonte	-1	3
7	Paramètre pour le calcul des degrés-jours.	0.01	0.2
8	Fraction minimale de rétention de l'eau dans le couvert nival	0	0.1
9	Différence entre Fcmin et la fraction de rétention de l'eau dans le couvert nival maximale	0.01	0.25
10	Paramètre pour le calcul de la capacité de rétention de la neige	0.005	0.05
11	Température de base pour le regel	-5	2
12	Facteur degré-jour du gel	0	5
13	Exposant dans l'équation du gel	0	1
14	Taux de correction de l'évapotranspiration potentielle	0	1
15	Portion de ruissellement	0	1
16	Taux de vidange vers le réservoir 1	0.00001	0.02
17	Taux de vidange vers le réservoir 2	0.00001	0.01
18	Niveau maximal du réservoir 1	0	500
19	Niveau maximal du réservoir 2	0	2000

Tableau-A II-3 Paramètres et bornes de calage pour le modèle HSAMI

Numéro	Description des paramètres	Unités	Borne inférieure	Borne supérieure
1	Efficacité ETP pour l'été	---	0.6	3
2	Efficacité ETP pour l'hiver	---	0	0.3
3	Taux de fonte le jour	cm/ $\Delta$ C/jour	0.05	0.4
4	Taux de fonte la nuit	cm/ $\Delta$ C/jour	0.05	0.5
5	Température de fonte le jour	°C	-6	7
6	Température de fonte la nuit	°C	-6	6
7	Température de référence pour calculer le flux de chaleur pluie-couvert nival	°C	-6	4
8	Effet du redoux sur l'aire enneigée	---	0.8	5
9	Effet du gel sur le ruissellement	---	0.8	15
10	Effet du sol sur le ruissellement	cm	10	45
11	Seuil minimum du sol saturé sur le ruissellement	cm	1	8
12	Sol minimum – quantité d'eau dans la zone vadose qui ne peut s'écouler par gravité	cm	0	7
13	Sol maximum – quantité maximale d'eau que peut contenir la zone vadose	cm	3	25
14	Espace disponible dans la nappe (Nappe maximum)	cm	4	30
15	Portion de ruissellement de surface	---	0.15	0.7
16	Portion de ruissellement dans le sol maximum (hydrogramme intermédiaire)	---	0.3	1
17	Taux de vidange de la zone vadose vers la nappe	jour <sup>-1</sup>	0.009	0.07
18	Taux de vidange de la nappe	jour <sup>-1</sup>	0.006	0.018
19	Taux de vidange de la réserve intermédiaire	jour <sup>-1</sup>	0.6	1.2
20	Mode de l'hydrogramme de surface	jour	0.3	5
21	Forme de l'hydrogramme de surface	---	0.4	5
22	Mode de l'hydrogramme intermédiaire	jour	1.5	13
23	Forme de l'hydrogramme intermédiaire	---	0.15	1.5



### ANNEXE III

#### MESURES D'EFFICACITÉ EN CALAGE POUR DES RÉSEAUX ALÉATOIRES DE DENSITÉ VARIABLE

Cette section expose les valeurs des différentes fonctions objectif utilisées dans la réduction manuelle et aléatoire de la densité du réseau d'observations sur le bassin versant. Les valeurs du critère pondéré, du biais relatif ainsi que du coefficient de variation de l'erreur quadratique sont présentées pour chacun des modèles utilisés.

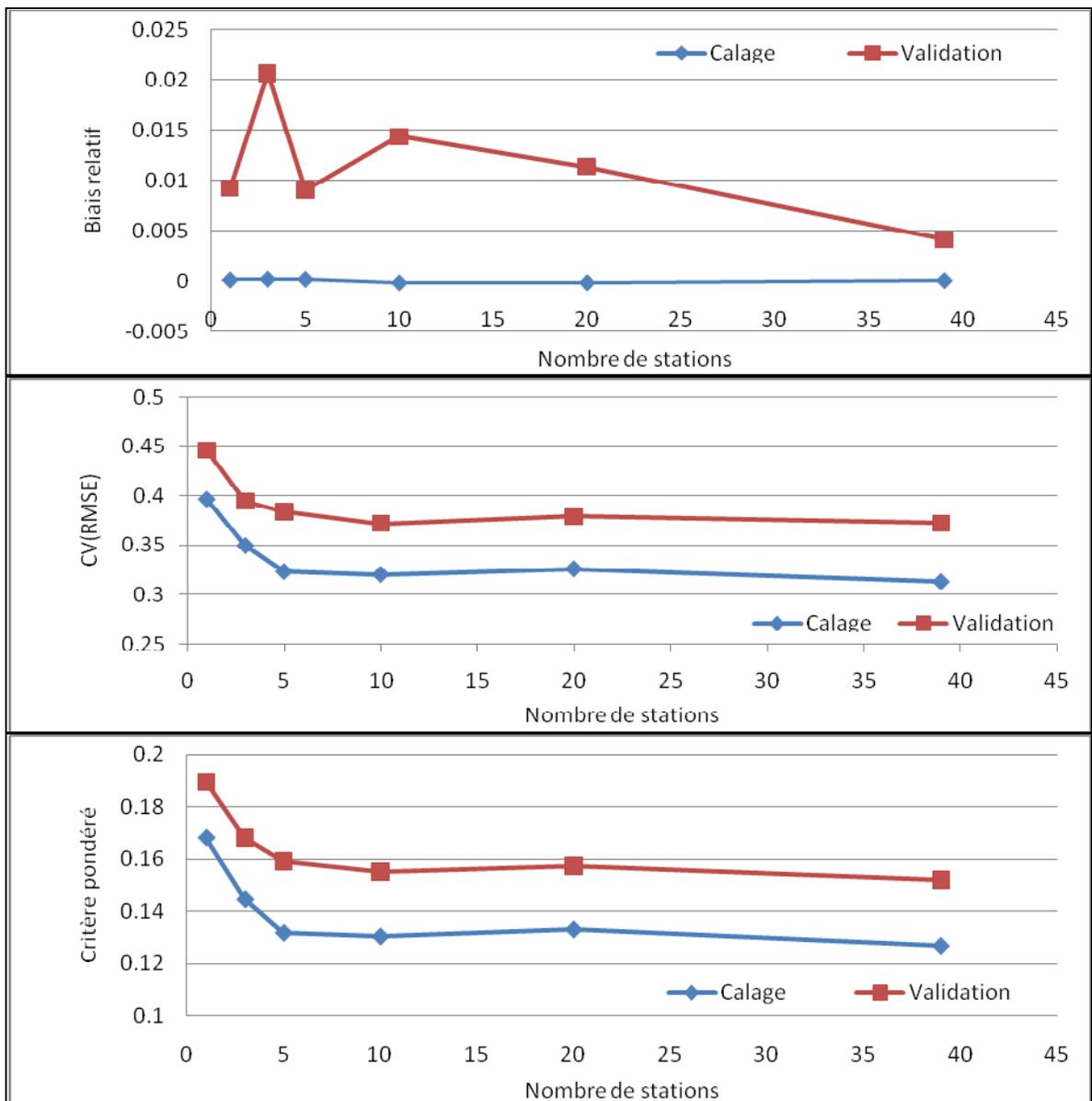


Figure-A III-1 Critères d'efficacité en calage pour le modèle HSAMI

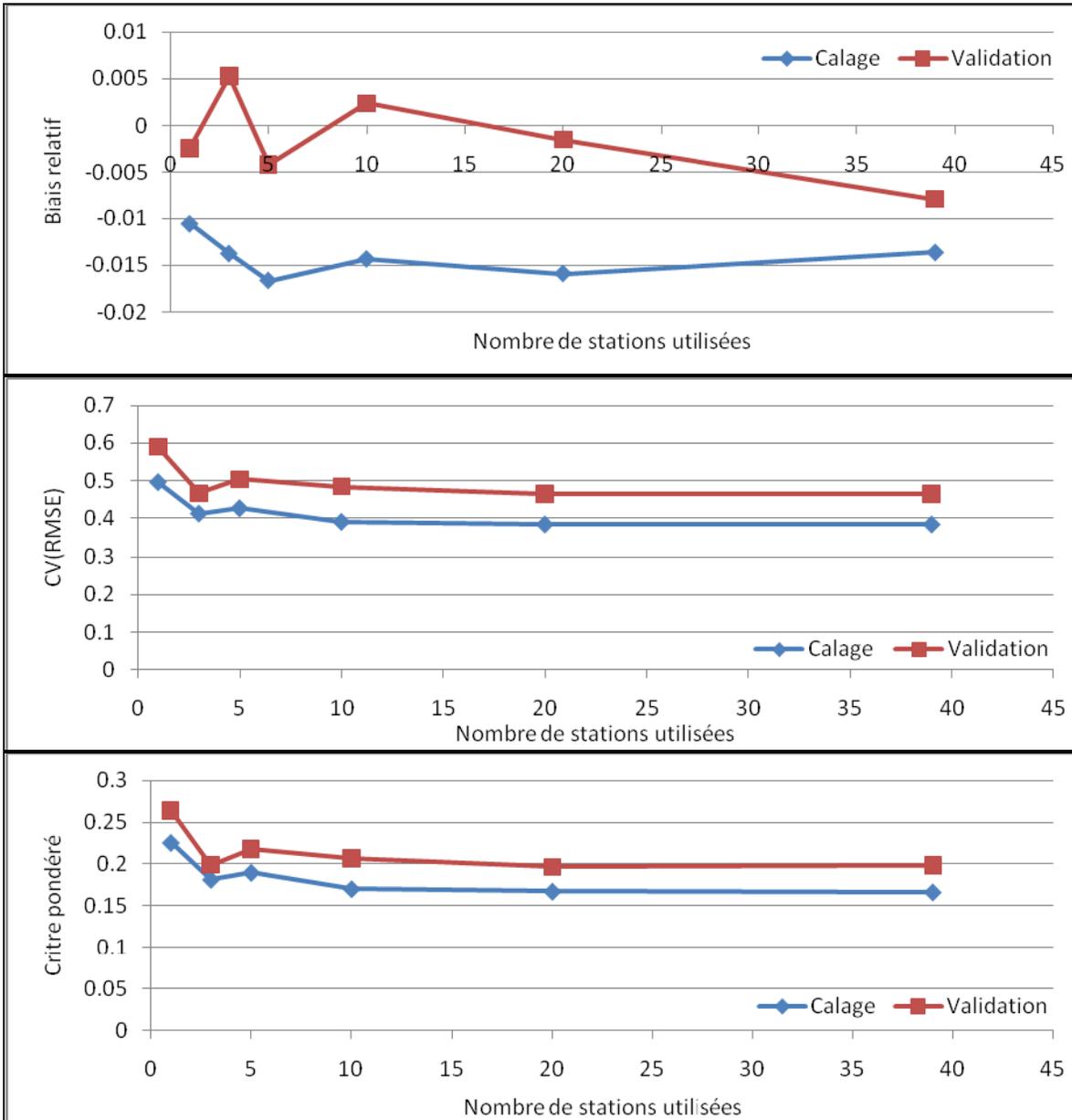


Figure-A III-2 Critères d'efficacité en calage pour le modèle HMETS

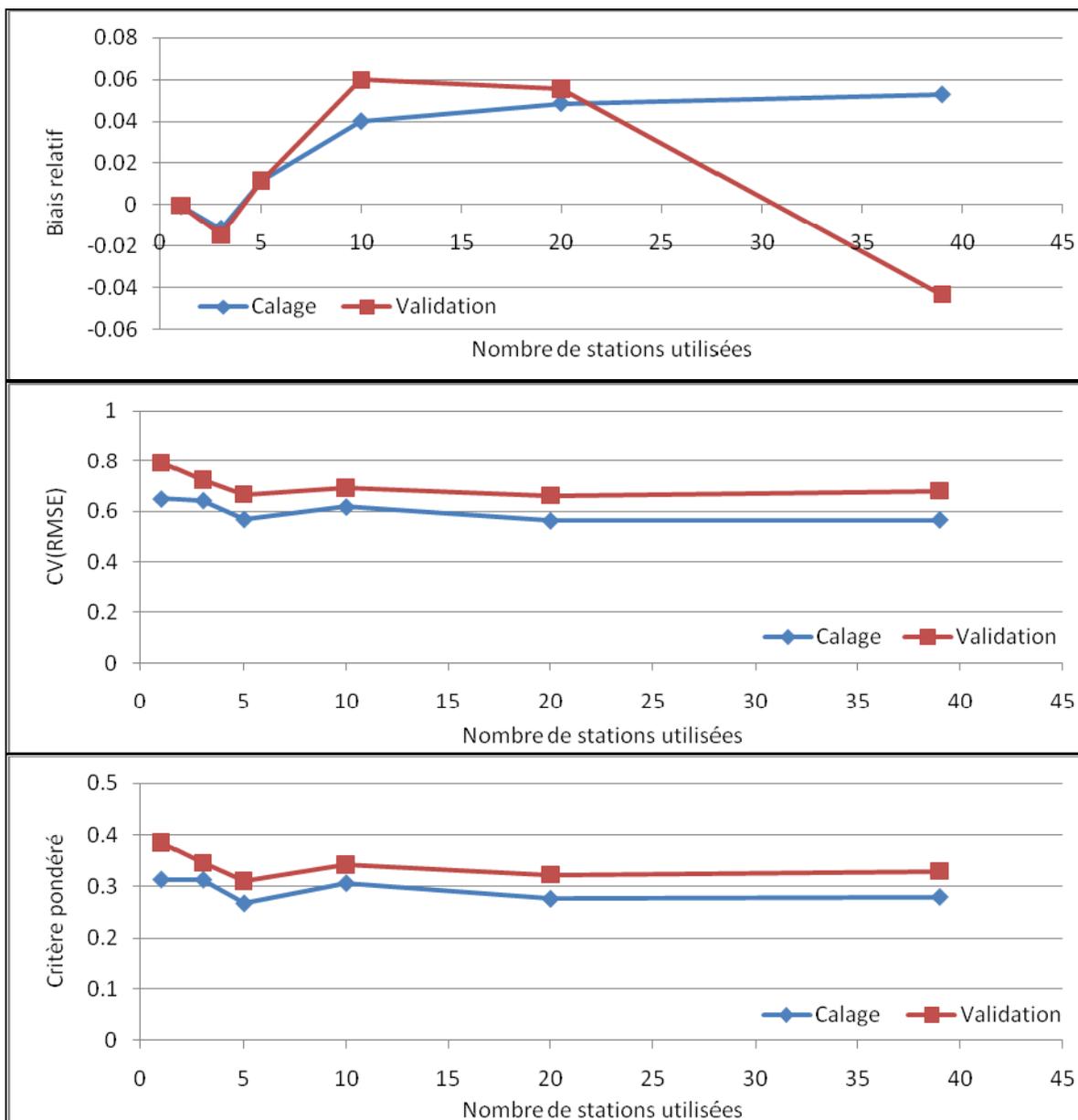


Figure-A III-3 Critères d'efficacité en calage pour le modèle Hydrotel



## LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Amani, A., et Lebel, T. (1997). Lagrangian kriging for the estimation of Sahelian rainfall at small time steps. *J. Hydrol.*, 192, p. 125–157.
- Amatya, D.M. et Trettin, C.C. (2007). Development of watershed hydrologic research at Santee Experimental Forest, coastal South Carolina. In: Furniss, M.J.; Clifton, C.F.; Ronnenberg, K.L., eds. *Advancing the fundamental sciences: Proceedings of the forest science national earth sciences conference*. PNR–GTR–689. Portland, OR: U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station: p.198–208. Vol. I.
- Arsenault, R., Poulin, A., Côté, P. et Brissette, F. (2011). A multi-comparison of global optimization algorithms in hydrological model calibration. *Soumis à J.Hydrol.* le 30 septembre 2011. 58 p.
- Bastin, G., Lorent, B., Duqué, C. et Gevers, M. (1984). Optimal Estimation of the Average Areal Rainfall and Optimal Selection of Rain Gauge Locations, *Water Resour. Res.*, 20(4), p.463–470, doi:10.1029/WR020i004p00463.
- Beven, K.J. (2006). A manifesto for the equifinality thesis. *J. Hydrol.* 320, p.18–36.
- Bisson, J. L., et Roberge, F. (1983). Prévisions des apports naturels: Expérience d'Hydro-Québec. Proc., Workshop on Flow Predictions, Institute of Electrical and Electronics Engineers (IEEE), Toronto.
- Brissette, F., Caya, D., Mailhot, A. et Roy, R. (2011). ALICE : A new Paradigm shift in hydrology research. *World Environmental & Water Resources Congress - Palm Springs, CA*, 22-26 mai.
- Burn, D.H. et Goulter, I.C. (1991). An approach to the rationalization of streamflow data collection networks. *J. Hydrol.*, 122, p.71-91.
- Burns, D. M. (1965). A summary of the Caspar Creek watershed study. Unpublished report presented to the California State Board of Forestry, Sacramento, California, 26 February 1965. 6 p.
- Cafferata, P. H. et Spittler, T. E. (1998). Logging impacts of the 1970's vs. the 1990's in the Caspar Creek watershed. In: Ziemer, Robert R., technical coordinator. *Proceedings of the conference on coastal watersheds: the Caspar Creek story*, 1998 May 6; Ukiah, CA. General Tech. Rep. PSW GTR-168. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture; p.103-115.

- Caya, D. et Laprise, R. (1999). A semi-implicit semi-Lagrangian regional climate model: The Canadian RCM. *Mon. Wea. Rev.* 127 (3), 341-362.
- Chen, J., Brissette, F. P., Poulin, A. et Leconte, R. (2011). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, *Water Resour. Res.*, 47, W12509, doi:10.1029/2011WR010602.
- Cheng, K.-S., Lin, Y.-C. et Liou, J.-J. (2008). Rain-gauge network evaluation and augmentation using geostatistics, *Hydrol. Process.*, 22, p.2554–2564, 2087.
- Creutin J.D., Delrieu G. et Lebel T. (1988). Rain measurement by rainage-radar combination: a geostatistical approach. *J. Atmos. Oceanic Technol.*, 5 (1), p.102-115.
- Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*, 1ère édition. Coll. « Wiley-Interscience series in systems and optimization ». Chichester, Angleterre : John Wiley & Sons, Ltd, 497 p.
- de Elia R, Caya D, Côté H, Frigon A, Biner S, Giguère M, Paquin D, Harvey R, Plummer D. (2008). Evaluation of uncertainties in the CRCM-simulated North American climate. *Clim Dyn* 30(2), 113–132
- Di Baldassarre, G. et Montanari, A. (2009). Uncertainty in river discharge observations: a quantitative analysis, *Hydrol. Earth Syst. Sc.*, 13, p.913–921.
- Duan, Q., Sorooshian, S. et Gupta, V. K. (1992). Effective and efficient global optimization for conceptual rainfall runoff models. *Wat. Resour. Res.* 24(7), 1163-1173.
- Epanechnikov, V.A. (1969). Non-parametric estimation of a multivariate probability density. *Theor. Probab. Appl.*, 14, p.153–158. doi:10.1137/1114019.
- Fortin, J.-P., Turcotte, R., Massicotte, S., Moussa, R., Fitzback, J. et Villeneuve, J.-P. (2001). Distributed watershed model compatible with remote sensing and GIS data. 1: Description of the model. *J. Hydraul. Eng.*, 6(2), p.91-99.
- Fortin, V. (2000). *Le modèle météo-apport HSAMI: historique, théorie et application*. Varennes: Institut de recherche d'Hydro-Québec. 68 p.
- Fortnow, L. (2009). The status of the P versus NP problem, *Communications of the ACM* 52 no. 9, p.78–86. doi:10.1145/1562164.1562186.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley.
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. The University of Michigan

- Hunter, S. (2009). WSR-88D Radar Rainfall Estimation: Capabilities, limitations and potential improvements. *NWA Digest* ;20:26-36.
- Husain, T. (1989). Hydrologic uncertainty measure and network design. *J. Am. Water Resour. As.*, 25, p.527–534. doi: 10.1111/j.1752-1688.1989.tb03088.x.
- Jedermann, R., Palafox-Albarran, J., Robla, J. I., Barreiro, P., Ruiz-Garcia, L. et Lang, W. (2011). Interpolation of spatial temperature profiles by sensor networks. *Sensors*, 2011 IEEE pp.778-781, 28-31 Oct. 2011 doi: 10.1109/ICSENS.2011.6127148
- Krajewski, W.F. et Smith, J.A. (2002). Radar hydrology: rainfall estimation, *Adv. Water Resour.*, Volume 25, Issues 8-12, August-December 2002, p.1387-1394, ISSN 0309-1708, 10.1016/S0309-1708(02)00062-3.
- Kruskal, W. H. et Wallis, W. A. (1953). Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Assn.* 47:p.583-621, 1952; and Errata. *J. Amer. Statist. Assn.* 48:p.907-911.
- Minville, M., Brissette, F. et Leconte, R. (2008). Uncertainty of the impact of climate change on the hydrology of a Nordic watershed. *J. Hydrol.*, 358(1–2), p.70–83.
- Moriassi D.N., Arnold J.G., Van Liew M.W., Bingner R.L., Harmel R.D. et Veith T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50(3), p.885-900.
- Music, B., Frigon, A., Slivitzky, M., Musy, A., Caya, D. et Roy, R. (2009). Runoff modelling within the Canadian Regional Climate Model (CRCM): analysis over the Quebec/Labrador watersheds. In : *New Approaches to Hydrological Prediction in Data Sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009)*. International Association of Hydrological Sciences (IAHS) Red Book Series Publ. 333, p.183-194.
- Music B. et Caya, D. (2007). Evaluation of the Hydrological Cycle over the Mississippi River Basin as Simulated by the Canadian Regional Climate Model (CRCM). *J. Hydrometeorology*, 8(5), p.969-988. DOI: 10.1175/JHM627.1.
- Osborn, H.B et Keppel, R V (1965). Dense Rainguage Network as a Supplement to Regional Network in Semi-arid Regions. *Symposium on the Design of Hydrological Networks, Part II Quebec*. International Association of Hydrological Science, Publication No. 68.
- Osborn, H. B., Lane, L. J. et Hundley, J. F. (1972). Optimum gaging of thunderstorm rainfall in southeastern Arizona, *Water Resour. Res.*, 8(1), p.259–265, doi:10.1029/WR008i001p00259.

- Ouarda, T. B. M. J., Rasmussen, P. F., Bobée, B. et Morin, J. (1996). Ontario hydrometric network rationalization. Statistical considerations, Res. Rep. R-470, Cent. Eau, Terre, et Environ., Inst. Natl. de la Rech. Sci., Univ. of Quebec, Quebec, Que., Canada.
- Pardo-Iguzquiza, E. (1998). Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, *J. Hydrol.*, 210, p.206 – 220, doi:10.1016/S0022-1694(98)00188-7.
- Pedersen, M.E.H. (2010). Good parameters for differential evolution. Technical Report HL1002 (Hvass Laboratories).
- Ramli, S. et Tahir, W. (2011). Radar Hydrology : New Z/R Relationships for Quantitative Precipitation Estimation in Klang River Basin, Malaysia. *International Journal of Environmental Science and Development*, Vol. 2(3).
- Schmitt, L. M. (2001). Theory of Genetic Algorithms, *Theor. Comput. Sci.*, 259 1–61.
- Schneebeili, M. et Laternser, M. (2004). A Probabilistic Model to Evaluate the Optimal Density of Stations Measuring Snowfall. *J. Appl. Meteor.*,43, p.711–719.
- Schneider-Vieira, F. (1993). Experimental Watershed project, results from 1993, year I. A report prepared for the Manitoba Model Forest by North/South Consultants Inc. 25 p.
- St-Hilaire, A., Ouarda, T. B. M. J., Lachance, M., Bobée, B., Gaudet, J. et Gignac, C. (2003). Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: a case study. *Hydrol. Process.* 17, p.3561–3580. doi: 10.1002/hyp.1350.
- Storn, R. et Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.*, 11, 341–359. doi:10.1023/A:1008202821328
- Verseghy, D. L. (1991). CLASS - A Canadian Land Surface Scheme for GCMS, I. Soil model. *Int. J. Climatol.*, 11, p.111–133.
- World Meteorological Organization. (2008). Guide to hydrological practices / World Meteorological Organization WMO, Geneva, Switzerland.