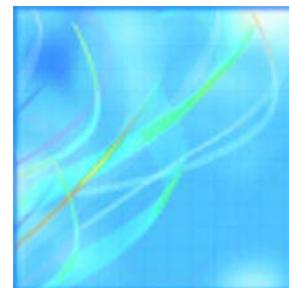


Document technique

Série de documents analytiques et techniques sur les
comptes et la statistique de l'environnement

Le modèle d'apport en eau pour le Canada exprimé en tant que moyenne de trente ans (1971 à 2000) : concepts, méthodologie et résultats initiaux



par Robby Bemrose, Laura Kemp, Mark Henry et François
Soulard

Section recherche et développement

Téléphone : (613) 951-4591



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca. Vous pouvez également communiquer avec nous par courriel à infostats@statcan.gc.ca ou par téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

Service de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Télécopieur	1-877-287-4369

Appels locaux ou internationaux :

Service de renseignements	1-613-951-8116
Télécopieur	1-613-951-0581

Programme des services de dépôt

Service de renseignements	1-800-635-7943
Télécopieur	1-800-565-7757

Comment accéder à ce produit

Le produit n° 16-001-M au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de choisir la rubrique « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de *normes de service à la clientèle* que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Offrir des services aux Canadiens ».

Le modèle d'apport en eau pour le Canada exprimé en tant que moyenne de trente ans (1971 à 2000) : concepts, méthodologie et résultats initiaux

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2009

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juin 2009

N° 16-001-M au catalogue, n° 7

ISSN 1917-9707

ISBN 978-1-100-90998-1

Périodicité : hors série

Ottawa

This publication is available in English upon request (catalogue no. 16-001-M).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Information pour l'utilisateur

Signes conventionnels

Les signes conventionnels suivants sont employés uniformément dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- p provisoire
- r révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- E à utiliser avec prudence
- F trop peu fiable pour être publié

Remerciements

Les auteurs remercient chaleureusement les nombreuses personnes qui ont contribué à notre compréhension des questions mises en cause par la production des estimations de l'apport en eau et de ce document. Ces personnes sont les suivantes : Elwood Scott, Gord Bell, Dave Kiely et Ron Woodvine, d'Agriculture et Agroalimentaire Canada; Jean-François Bibeault, David Boivin, André Bouchard, Willard Boutot, Jean-François Cantin, Malcolm Conly, Dave Harvey, Vincent Mercier, Al Pietroniro, Michel Villeneuve et Gillian Walker, d'Environnement Canada; Richard Fernandes, de Ressources Naturelles Canada; et Rowena Orok de la Division des comptes et de la statistique de l'environnement (DCSE) de Statistique Canada. Bahram Daneshfar, d'Agriculture et Agroalimentaire Canada, et Mike Sawada, de l'Université d'Ottawa, nous ont fait part de certaines connaissances spécialisées essentielles en matière de géostatistique. Claude Girard, de la Division des méthodes d'enquêtes auprès des entreprises, et Caroline Fric, anciennement de la DCSE, ont également apporté une contribution importante. De plus, nous voulons remercier la contribution de Monique Deschambault, Laurie Jong, Doug Trant et Hélène Laniel de la Section de production et diffusion de la DCSE ainsi que l'équipe de la Division de la diffusion composée de Lynne Durocher, Bradley Parsons et Philippe Thiboutot Simard.

Table des matières

Résumé	5
Le modèle d'apport en eau pour le Canada exprimé en tant que moyenne de trente ans (1971 à 2000) : concepts, méthodologie et résultats initiaux	
1 Introduction	6
2 Données et méthodes	7
3 Résultats	26
4 Conclusion et orientation future	39
Références	42

Résumé

Nous présentons ici la méthodologie élaborée par Statistique Canada pour calculer l'apport en eau annuel moyen au Canada. L'apport en eau, aux fins du présent document de recherche, se définit comme la quantité d'eau douce dérivée des mesures de débits ($\text{m}^3 \text{s}^{-1}$) non régularisés dans une zone géographique donnée pendant une période précise. Cette méthodologie a été appliquée sur la période allant de 1971 à 2000.

Cette recherche a été effectuée afin de combler des lacunes du programme de la statistique de l'eau de Statistique Canada. Ces lacunes existent du fait que les estimations du débit d'eau douce au Canada n'ont pas été calculées régulièrement et qu'elles ont été produites à l'aide de diverses méthodes qui ne génèrent pas nécessairement des résultats comparables. La méthodologie élaborée dans le cadre de cette étude produit des résultats cohérents dans le temps et dans l'espace. Ces résultats seront utilisés à l'avenir pour étudier les changements d'apports en eau de façon plus détaillée.

Pour obtenir les estimations d'apport en eau, une base de données renfermant les observations du débit naturel entre 1971 et 2000 a été constituée. Les valeurs du débit ont ensuite été converties en profondeur d'écoulement, puis interpolées à l'aide de la méthode du krigeage ordinaire afin de produire des estimations spatiales de l'écoulement. Ces estimations spatiales ont ensuite été mises à l'échelle pour créer une estimation de l'apport en eau s'échelonnant sur trente ans. Cette méthodologie et ces résultats ont ensuite été validés à l'aide de plusieurs techniques comprenant des analyses de stabilité et d'incertitude. Les résultats de cette méthodologie indiquent que l'apport en eau annuel moyen au Canada sur une période de trente ans est de $3\,435 \text{ km}^3$.

Le modèle d'apport en eau pour le Canada exprimé en tant que moyenne de trente ans (1971 à 2000) : concepts, méthodologie et résultats initiaux

par Robby Bemrose, Laura Kemp, Mark Henry et François Soulard

1 Introduction

La présente étude décrit la méthodologie mise au point pour calculer l'apport en eau au Canada. L'élaboration de cette méthodologie avait comme objectif de combler des lacunes statistiques des Comptes d'eau de Statistique Canada, notamment le compte des ressources en eau douce (Souillard, 2003). Ces lacunes découlent du fait que les estimations du débit d'eau douce au Canada ne sont pas calculées régulièrement et ont été produites à l'aide de méthodes qui ne génèrent pas des résultats comparables (Sanderson et Phillips, 1967; Laycock, 1985; Pearse et coll., 1985). Il s'agit là d'un obstacle à l'analyse, car pour effectuer l'étude adéquate des ressources en eau, particulièrement du point de vue de l'offre et de la demande, il est impératif de pouvoir réaliser une comparaison temporelle et spatiale des données. Les résultats de la méthodologie élaborée ici représentent une moyenne annuelle de l'apport en eau au Canada sur une période de trente ans (1971 à 2000) qui est cohérente dans le temps et dans l'espace, et qui nous permettra d'étudier la disponibilité et le renouvellement des ressources en eau de manière plus détaillée.

Plus précisément, les objectifs de cette étude consistaient à :

1. créer des surfaces d'écoulement annuel en utilisant une interpolation spatiale;
2. quantifier les estimations spatiales d'apport en eau annuel au Canada au moyen de surfaces d'écoulement.

Le présent document décrit en détail la méthodologie (section 2) ainsi que les résultats (section 3).

1.1 Apport en eau

Dans le présent document, l'apport en eau se définit comme la quantité d'eau douce dérivée des mesures de débits ($m^3 s^{-1}$) non régularisés dans une zone géographique donnée pendant une période précise. Le débit d'eau douce est le produit d'une combinaison du débit de base, de l'écoulement hypodermique et du ruissellement en surface provenant de l'eau souterraine, des précipitations ou de l'accumulation annuelle de neige. Le débit englobe les processus hydrologiques (par exemple, interception, infiltration et évapotranspiration), l'état du stockage d'eau (par exemple, lacs, aquifères, accumulation annuelle de neige et humidité du sol) dans un bassin hydrographique (Mosley et McKerchar, 1993) et dépend de variables climatiques (par exemple, température) et physiographique (par exemple, topographie). L'apport en eau s'appuie sur le débit non régularisé en ce sens que seules les mesures démontrant le débit naturel (c'est-à-dire, non endigué ou détourné) sont utilisées dans son calcul. Le débit est utilisé exclusivement pour estimer l'apport en eau parce qu'il peut être mesuré avec un niveau de précision connu (Dingman, 1994), comparativement à d'autres parties du cycle de l'eau, telles que les précipitations et l'évapotranspiration, qui peuvent présenter des taux d'incertitude plus élevés.

1.2 Comptes d'eau

La méthodologie d'apport en eau a été élaborée pour fournir des estimations des réserves de ressources en eau pour la composante Comptes d'eau du cadre de comptabilité de l'environnement de Statistique Canada, soit le Système des comptes de l'environnement et des ressources du Canada (SCERC).¹ Les Comptes d'eau offrent un aperçu intégré de l'eau et de son utilisation par la société. Cet aperçu intégré est fondé sur une compilation de données environnementales qui est cohérente dans l'espace et dans le temps, qui couvre l'ensemble du pays et qui est compatible avec les comptes économiques qui appuient notre capacité à comprendre et à traiter les enjeux liés à l'eau au Canada. Ce système a été créé afin de répondre au besoin d'amélioration de la surveillance des liens entre l'activité anthropique et l'environnement.

2 Données et méthodes

Nous avons exécuté quatre grandes étapes de traitement pour produire une estimation de l'apport en eau annuel moyen au Canada sur une période de trente ans. La première étape comprenait la création d'un cadre national de données pour estimer l'apport en eau mensuel. La deuxième étape consistait à calculer l'écoulement par unité de surface à l'aide du cadre de données. La troisième étape consistait à utiliser les valeurs d'écoulement pour produire une distribution spatiale de l'écoulement mensuel à l'échelle du pays annuellement, sur une période de trente ans. Pour terminer, à la quatrième étape, nous avons mis à l'échelle les surfaces d'écoulement mensuel pour produire des estimations d'apport en eau qui, pour les fins de cette étude, sont présentées en tant que moyenne annuelle. Les sections suivantes expliquent de façon détaillée chacune de ces étapes de traitement.

2.1 Création du cadre de données

Il a fallu recourir à divers ensembles de données et effectuer le traitement de ces données pour créer un cadre de données national pour le Canada sur la période allant de 1971 à 2000. Les sections suivantes décrivent ces ensembles de données, de même que les étapes qui ont été suivies pour la création du cadre. En bref, les grandes étapes de la création de ce cadre sont :

1. assemblage de stations hydrométriques de débit rapportant des débits naturels;
2. la compilation des superficies des bassins « contribuant » pour chaque station;
3. le déplacement des stations au centroïde des bassins contribuant.

2.1.1 Débit

La première étape de la création du cadre de données comprend l'assemblage des données hydrométriques. Les données de débit mensuel ($m^3 s^{-1}$) utilisées provenaient de la base de données HYDAT (HYDEX) 2005 (Environnement Canada, 2007).² Cette base de données 2005 contient des données sur le débit tirées de plus de 2 100 stations en service et de près de 4 200 sites abandonnés au Canada. Dans la base de données, chaque station est accompagnée de champs qui indiquent son identificateur unique, sa latitude et sa longitude, la superficie du bassin versant (km^2), sa période d'exploitation, l'indicateur de régularisation ainsi que des remarques faisant état de renseignements tels que la qualité des données et le régime hydraulique.

La base de données HYDAT a été filtrée afin d'inclure seulement les stations ayant un débit naturel en utilisant le champ HYDEX qui indique le régime. Après tous les filtrages, les champs des remarques de chaque station ont été évalués : si les remarques indiquaient qu'il y avait des caractéristiques de régularisation dans un bassin, la station

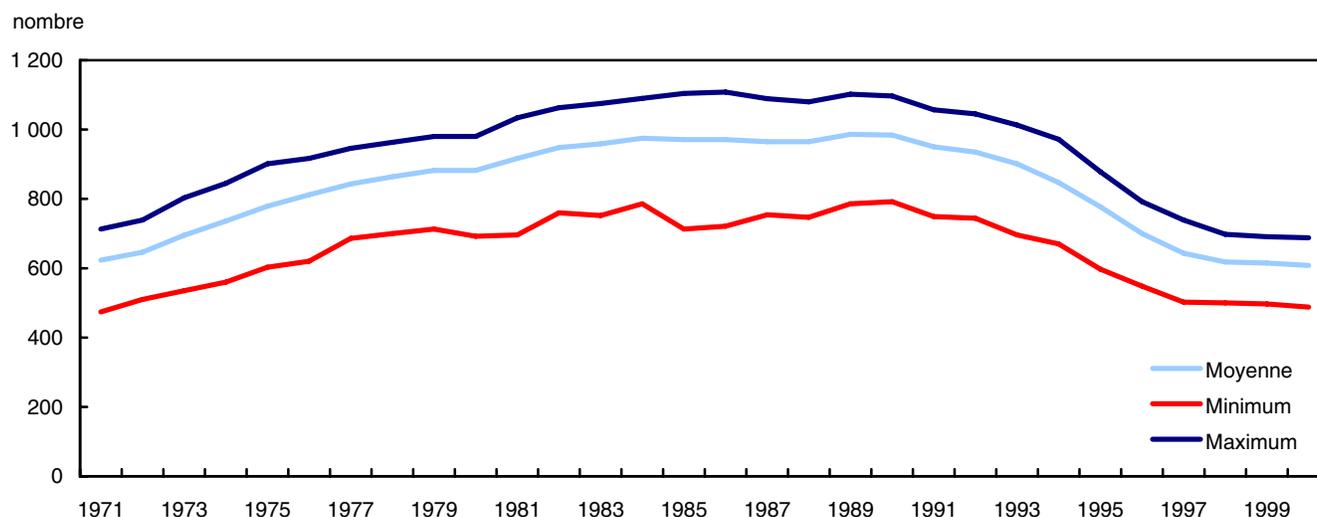
1. Une description du SCERC est fournie dans Statistique Canada (1997).

2. HYDEX est une base de données relationnelle qui contient les données recueillies dans les diverses stations (en service ou abandonné) du Canada sur le débit, le niveau d'eau, la superficie du bassin versant et la concentration de sédiments. HYDAT est une base de données relationnelle qui contient les mesures calculées à partir des données enregistrées aux stations figurant à la liste de HYDEX. On y trouve notamment les valeurs quotidiennes et mensuelles moyennes sur le débit, le niveau d'eau et la concentration de sédiments (Environnement Canada, 2007).

était supprimée de la base de données. Toutefois, les stations où le régime était régularisé ont été retenues lorsque la remarque mentionnait que la régularisation avait une incidence minimale sur le débit. De plus, on a fait des exceptions dans le cas des stations de l'Arctique lorsque les remarques contenaient un avertissement, car il n'y avait que quelques stations en service. De plus, il y avait un certain nombre de stations qui interrompaient les enregistrements pendant l'hiver. Ces stations se voyaient attribuer une valeur de zéro pour les mois de non-enregistrement si la dernière journée du mois précédant l'interruption et la première journée suivant cette interruption présentaient des valeurs de zéro. Le graphique 1 présente le nombre de stations moyen, le maximum et le minimum qui a été inclus chaque année après le filtrage de la base de données HYDAT pour la durée de l'étude.

Graphique 1

Stations ayant un débit naturel : nombre maximal, moyen et minimal de stations incluses pour chaque année après le filtrage de la base de données, de 1971 à 2000



Source(s) : Environnement Canada, 2007. Archives nationales des données hydrologiques, base de données hydrométriques (HYDAT), www.wsc.ec.gc.ca/products/hydat/main_f.cfm?cname=hydat_iso_f.cfm (site consulté le 7 janvier 2009) et Statistique Canada, Division des comptes et de la statistique de l'environnement.

2.1.2 Bassins « contribuant »

La deuxième étape de la création du cadre de données comportait la délimitation du bassin versant pour chaque station. Ces bassins versants, ou bassins « contribuant », ont été compilés à partir de trois sources principales :

1. les bassins versants de l'Administration du rétablissement agricole des Prairies (ARAP) [ARAP, 2004];
2. les Données-cadres à l'échelle nationale sur l'hydrologie – Aires de drainage, version 5.0 (NSDF) [Ressources naturelles, 2003];
3. la base de données HYDEX (Environnement Canada, 2003).

La projection de toutes les données spatiales a été effectuée à l'aide de la projection conique équivalente d'Albers.³ Les superficies des bassins contribuant de HYDEX ont été comparées à celles de l'ARAP et des NSDF; lorsque des problèmes ou des écarts majeurs ont été relevés, une inspection visuelle a été réalisée.

La base de données ARAP couvre toutes les provinces de l'Alberta, de la Saskatchewan et du Manitoba ainsi que des parties des régions voisines faisant partie de bassins hydrographiques des Prairies (ARAP, 2004). Pour les régions du pays qui font partie du territoire de l'ARAP, les données sur les bassins versants produites et mises à

3. La projection conique équivalente d'Albers convient à la projection des masses terrestres importantes ayant une orientation est-ouest parce qu'elle conserve les superficies de façon précise. Cette projection a été utilisée pour maximiser la précision des valeurs de l'apport en eau.

jour par l'ARAP ont été utilisées. Cet ensemble de données fournit les superficies totale et effective des bassins versants. La superficie effective est la partie du bassin qui pourrait contribuer au ruissellement parvenant au cours d'eau principal pendant la crue à période de récurrence de deux ans. La superficie totale correspond au plan limité par son bassin hydrographique, qui contribuerait au cours d'eau principal du bassin dans des conditions d'humidité extrême (Godwin et Martin, 1975). Les superficies effectives des bassins versants étaient considérées comme les plus appropriés pour cette étude. Un traitement important s'est avéré nécessaire pour incorporer cet ensemble de données dans le cadre, qui a consisté principalement à supprimer les stations qui ne faisaient pas partie de HYDAT, puis à supprimer ou à fusionner leurs bassins contribuant. Nous avons également trouvé de nombreux emplacements dans la base de données de l'ARAP pour certaines stations qui ont été déplacées au fil du temps. Seuls les emplacements les plus récents ont été retenus.

Les NSDF comprennent des polygones qui définissent les bassins versants pour toutes les stations HYDAT dont la superficie de drainage est supérieure à 200 km². Pour toutes les régions du Canada qui ne font pas partie du territoire de l'ARAP, nous avons utilisé les NSDF. Une première comparaison des superficies des bassins versants, telles que rapportées dans les NSDF, et des superficies publiées dans HYDEX a montré que bon nombre des bassins générés pour les NSDF n'étaient pas bien définis. Selon une évaluation visuelle des polygones des NSDF, de nombreux bassins avaient des « trous » dans leur couverture ou leur délimitation semblait plus grande qu'elle aurait dû l'être à cause de l'échelle du réseau hydrographique qui avait été utilisée pour les définir. La comparaison des superficies des bassins des NSDF à celles de la base de données HYDEX nous a permis de constater que les cas où les superficies des NSDF étaient plus petites que celles fournies dans HYDEX étaient fréquemment expliqués par des « trous » dans la couverture. Ces « trous » sont le résultat du fait que des lacs n'étaient pas reliés au réseau hydrographique et, par conséquent, n'avaient pas été pris en compte dans la technique utilisée pour délimiter la superficie des bassins. Lorsque les superficies des bassins indiquées dans les NSDF étaient plus grandes, c'était vraisemblablement parce que le réseau hydrographique généralisé utilisé pour générer les limites du bassin versant n'était pas suffisamment détaillé pour délimiter son emplacement réel. C'est pourquoi, dans la section 2.2, les superficies des bassins versants publiées dans HYDEX ont été utilisées pour calculer les valeurs de l'écoulement par unité de surface plutôt que les superficies des NSDF.

Une description détaillée de la méthodologie utilisée pour produire les bassins de l'ARAP, des NSDF et de la base de données HYDEX dépasse la portée de ce document; on peut toutefois trouver une description exhaustive de cette méthodologie dans l'ARAP (2004), Brooks et coll. (2002) et Environnement Canada (2007), respectivement.

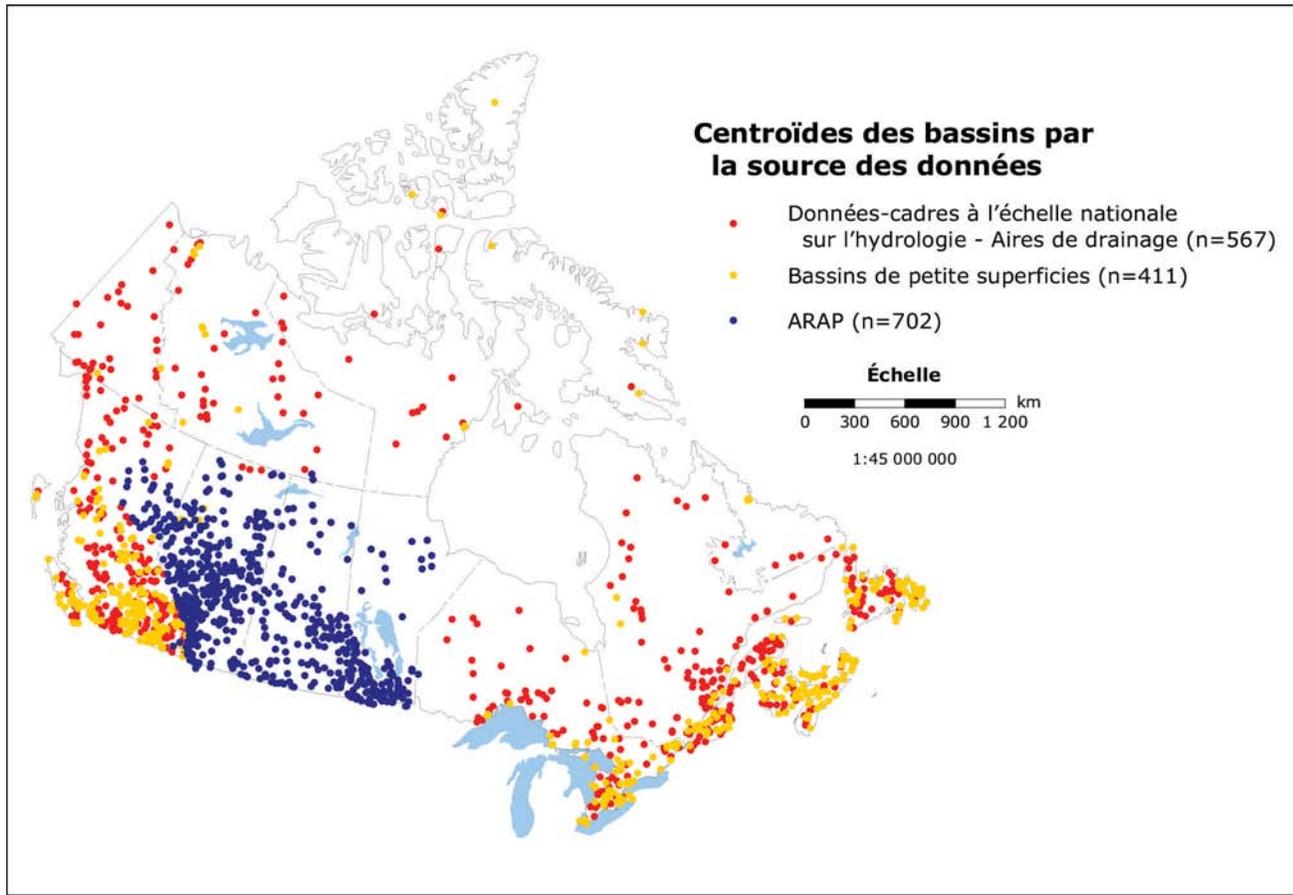
2.1.3 Positionnement au centroïde

La troisième étape de la création du cadre de données consiste à trouver l'emplacement auquel attribuer la valeur unitaire de l'écoulement. La superficie des bassins des NSDF a été utilisée pour établir le positionnement au centroïde dans les régions où la couverture de l'ARAP n'était pas offerte pour deux raisons principales : 1) les NSDF constituent la seule couverture spatiale numérique à l'échelle nationale des superficies des bassins versants de la base de données HYDAT; 2) l'alignement et la forme générale des polygones des superficies des bassins versants des NSDF étaient jugés appropriés pour cette tâche.

Toutes les stations HYDAT ont été déplacées au centroïde de chacun de leur bassin versant correspondant. Ce déplacement a été effectué puisque les valeurs de l'écoulement positionnées au centroïde des bassins versants sont plus représentatives du bassin que l'emplacement de la jauge lorsqu'on estime l'écoulement par unité de surface (Krug et coll., 1990). La superficie du bassin versant était considérée comme toute superficie en amont de la station, même si celle-ci se superposait à celle d'autres stations situées en amont. Finalement, étant donné que l'ensemble de données NSDF ne contenait pas de bassins plus petits que 200 km² mais qu'il y avait de nombreuses stations HYDAT qui correspondaient à ce critère, les stations ont été intégrées dans l'ensemble de données sans être repositionnées au centroïde de leur bassin versant. On a considéré que cela n'aurait aucune incidence sur l'estimation finale étant donné que le changement de position serait minime. Le cadre de données finales des centroïdes du débit mensuel par source de données est présenté sur la carte 1.

Carte 1

Cadre de données finales des centroïdes du débit mensuel, par source



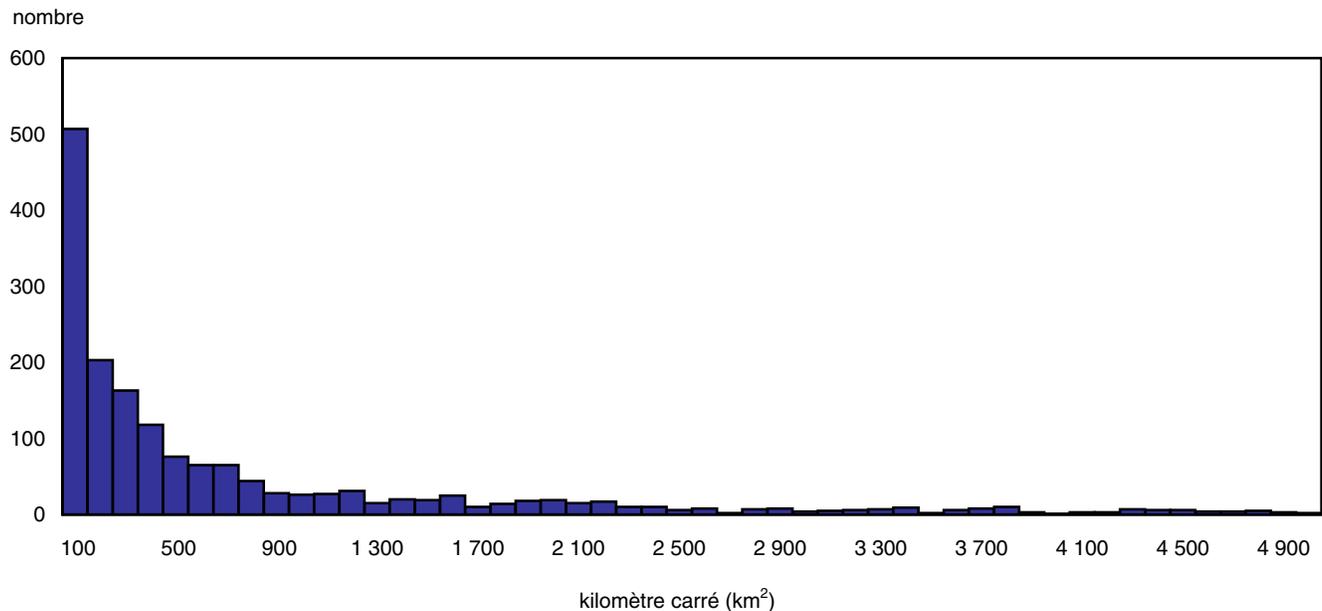
Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

2.2 Écoulement par unité de surface

Une fois le cadre de données élaboré, l'étape suivante du modèle comprenait le calcul d'un écoulement par unité de surface (mm) [c'est-à-dire, profondeur de l'écoulement]. Ce calcul a été effectué en divisant les valeurs de débit mensuel par la superficie du bassin versant correspondant, puis en attribuant les résultats aux centroïdes. Nous avons fait cette opération seulement pour les bassins ayant une superficie inférieure à 5 000 km². Cette condition suppose que la réponse hydrologique est plus susceptible d'être liée à la mesure du débit à une station pour les bassins ayant une superficie $\leq 5\ 000\ \text{km}^2$. La documentation existante fournit des preuves suffisantes qu'il s'agit d'un critère acceptable, et ce, malgré ses limites potentielles.

Le seuil de $\leq 5\ 000\ \text{km}^2$ est fondé sur l'étude Annual Unit Runoff of the Canadian Prairies menée par Bell (1994), dans laquelle on supposait que des bassins versants plus étendus pourraient contenir un mélange de conditions hydrologiques plus diversifié. Bien que cette condition ait été établie pour les bassins des Prairies, et qu'il soit fort probable que la taille des bassins n'indique pas nécessairement une plus grande complexité pour les bassins des autres régions du pays, on considère raisonnable l'hypothèse de limiter l'étendue de la superficie pour déterminer l'écoulement unitaire. Krug et coll. (1990) utilisent une condition semblable pour cartographier l'écoulement annuel moyen dans certaines régions des États-Unis. Par conséquent, dans cette étude, les superficies des bassins versants varient de 0,16 km² à 4 950 km², la taille médiane étant de 278 km² (graphique 2).

Graphique 2
Distribution des fréquences des bassins versants correspondants



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

2.3 Interpolation de l'écoulement

De nombreuses études ont été réalisées afin de comparer le rendement des méthodes d'interpolation pour la production d'estimations spatiales. Les études qui ont été examinées pour rédiger ce document (Tabios et Salas, 1985; Abteu et coll., 1993; Borga et Vizzaccaro, 1997; Syed et coll., 2003) mettent l'accent sur l'interpolation des données de précipitation obtenues à l'aide de pluviomètres, mais on considère que cette méthodologie est semblable à celle qui serait employée pour l'interpolation des données d'écoulement. Ces études sont unanimes sur le fait que le krigeage fournit les prévisions de surface optimales. Afin de s'assurer que le krigeage était approprié pour les données utilisées dans la présente étude, une analyse comparative du krigeage et des interpolateurs de la fonction de base radiale a été réalisée (par exemple, multiquadratique, plans splinés et courbe splinée avec tension). La méthode de krigeage a produit les résultats les plus concluants dans les régions où les valeurs sur de courtes distances étaient très variables ou là où la densité des points diminuait ou était faible ou nulle près des limites de la région à l'étude. Cette méthode a également permis d'obtenir une surface plus lisse qui reproduisait la variation locale sans effet « cible »⁴, qui se produisait lorsqu'on utilisait les interpolateurs de la fonction de base radiale. De plus, il était avantageux d'utiliser le krigeage plutôt que les autres interpolateurs, car on pouvait produire une erreur-type de la prévision, qui définissait le degré d'incertitude lié à la surface interpolée. Eaton et coll. (2002) ainsi que Krakauer et Fung (2008) sont des exemples d'études dans lesquelles le krigeage a été utilisé pour l'interpolation de l'écoulement à l'échelle régionale et nationale.

Comme dans le cas d'autres méthodes d'interpolation, le krigeage fait appel à des valeurs d'observation proches des endroits n'ayant pas fait l'objet de mesures afin d'en estimer leurs valeurs. L'estimateur du krigeage est une moyenne pondérée des échantillons de données du voisinage. Il existe plusieurs variantes du krigeage qui sont utilisées pour le lissage (par exemple, krigeage simple, indicateur krigeage). Les méthodes de krigeage simple, ordinaire et universel (krigeage avec modèle de tendance) sont les trois types de krigeage qui auraient pu être utilisés dans la présente étude. Le krigeage ordinaire est la méthode la plus fréquemment employée, et on l'a

4. L'effet cible se produit lorsque certaines observations influent excessivement sur l'interpolation, entraînant des points localisés d'une valeur extrême sur la surface concentrée sur l'observation, qui diminue de façon radiale en fonction de la distance, formant ainsi un semblant de « cible ».

préférée dans le cadre de cette étude en raison des résultats d'une comparaison des trois méthodes : il a été déterminé que le krigeage ordinaire était plus efficace que les krigeages simple et universel dans les régions pour lesquelles on ne disposait que peu de données.

Avant de pouvoir appliquer le krigeage ordinaire, il a fallu calculer la semi-variance de l'ensemble de données afin de déterminer les poids utilisés dans l'interpolation. La semi-variance sert à mesurer la dépendance spatiale entre les observations sur des distances précises (décalages). La semi-variance est calculée selon la formule suivante :

$$(1) \quad \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z(x_i) - z(x_i + h)\}^2$$

où $\hat{\gamma}(h)$ est la semi-variance estimée à un décalage (h), $z(x_i)$ et $z(x_i + h)$ correspondent aux valeurs observées aux points x_i et $x_i + h$ séparés par une distance h , pour lesquels il y a un nombre de paires de données $N(h)$ (Borga et Vizzaccaro, 1997). Les points $\hat{\gamma}(1), \hat{\gamma}(2), \hat{\gamma}(3), \dots$, pour une distance h donnée sont graphiqués afin de produire ce qu'on appelle un semi-variogramme empirique où une fonction continue est ajustée aux valeurs de façon à ce que l'on puisse effectuer des prévisions. La fonction la plus couramment utilisée est un modèle sphérique dont la formule est la suivante :

$$(2) \quad \gamma(h) = \begin{cases} c \left\{ \frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\} & \text{pour } h \leq a \\ c & \text{pour } h > a \end{cases}$$

où c et a représentent le seuil et la fourchette, respectivement (Webster et Oliver, 2000). La fourchette caractérise la limite de la dépendance spatiale (autocorrélation spatiale) entre les valeurs de l'écoulement et se définit par le seuil, qui indique le degré d'hétérogénéité maximal. La fourchette calculée à l'aide du modèle sphérique détermine également le rayon du cercle de recherche des points avoisinants utilisé pour réaliser le krigeage. Le krigeage ordinaire calcule la valeur de Z à l'emplacement x_0 par $\hat{Z}(x_0)$ en utilisant la moyenne pondérée des valeurs avoisinantes de la variable Z à l'emplacement x_i comme suit :

$$(3) \quad \hat{Z}(x_0) = \sum_{i=1}^{N(h)} \lambda_i z(x_i)$$

où $z(x_i)$ correspond aux valeurs mesurées aux i_e emplacements, x_0 est l'emplacement prédit, λ_i sont les poids de l'estimateur qui minimisent la variance (c'est-à-dire, $E\{\hat{Z}(x_0) - Z(x_0)\} = 0$, où E est le ruissellement attendu) et N est le nombre de valeurs mesurées (Borga et Vizzaccaro, 1997; Webster et Oliver, 2000). La structure spatiale du semi-variogramme est utilisée pour produire un système d'équations linéaires de krigeage combinant les données sur les points du voisinage comme suit :

$$(4) \quad \sum_{i=1}^{N(h)} \lambda_i \gamma(x_i - x_j) + \psi = \gamma(x_i - x_0)$$

avec les contraintes suivantes :

$$(5) \quad \sum_{i=1}^N \lambda_i = 1$$

où ψ est le multiplicateur de Lagrange pour la contrainte sur les poids des conditions non-biaisées (St-Hilaire et coll., 2003), $\gamma(x_i, x_j)$ est le semi-variogramme de Z entre les données x_i et x_j , et $\gamma(x_i, x_0)$ est le semi-variogramme entre l'observation i_e et l'emplacement estimé x_0 (Webster et Oliver, 2000). Les sommation des poids utilisés dans

le krigeage ordinaire doit être égale à 1 afin que le prédicteur, equation (3) soit sans biais pour $z(x_0)$ (Isaaks et Srivastava, 1989).⁵

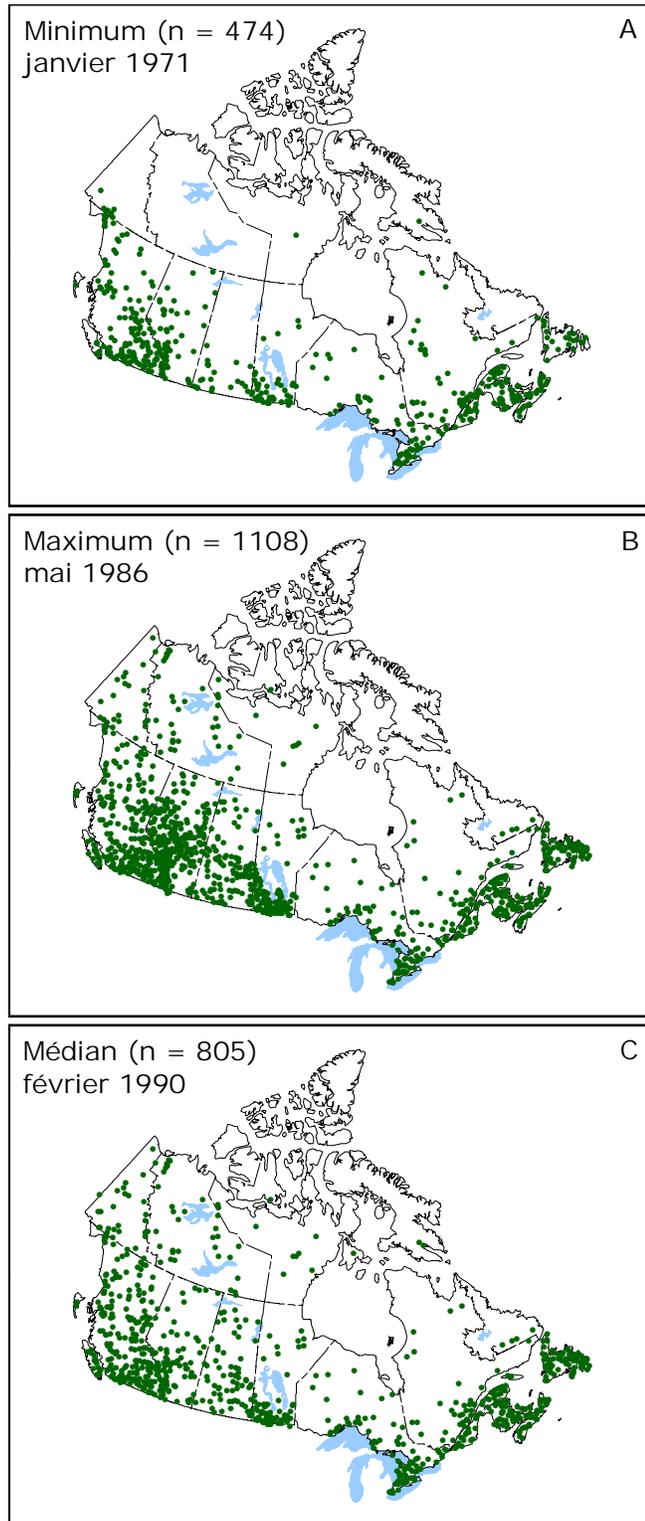
Avant la modélisation des semi-variogrammes, les histogrammes des données d'écoulement mensuel ont été évalués aux fins de normalisation. Le krigeage ordinaire est reconnu comme la méthode optimale lorsque les données suivent une distribution normale (Armstrong et Boufassa, 1987) étant donné que, selon l'hypothèse sous-jacente de l'analyse des semi-variogrammes, la distribution des valeurs se rapproche d'une distribution normale (Houlding, 2000). Les données de l'écoulement présentaient une forte asymétrie positive, de même que plusieurs valeurs de zéro pour les observations de l'écoulement dans les Prairies et l'Arctique. Pour transformer les données de façon à ce qu'elles suivent une distribution normale, l'équation $Y = \text{Log}_{10}(\text{runoff} + 1)$ a été appliquée avec une constante de 1 qui a été ajoutée afin d'éliminer les valeurs de zéro. Les constantes de 1 ou moins sont recommandées, car des valeurs élevées peuvent éliminer l'avantage de la transformation logarithmique (Yamamoto, 2007). Selon des études menées par Wu et coll. (2006) ainsi que par Kravchenko et Bullock (1999), la transformation de données fortement asymétriques permet d'améliorer les résultats obtenus à partir du krigeage ordinaire.

Puisque la surface interpolée varie en fonction de l'agencement spatial des données d'entrée (carte 2), celui-ci a été évalué afin de déterminer comment les observations étaient distribuées dans l'étendue visée par l'étude.

5. Pour obtenir une description plus détaillée du système de krigeage avec sa notation matricielle, consulter Borga et Vizzaccaro (1997) ou Webster et Oliver (2000).

Carte 2

Les mois ayant le nombre d'enregistrements minimal (A), maximal (B) et médian (C) des mois se trouvant dans les 360 mois d'observations entre 1971 et 2000

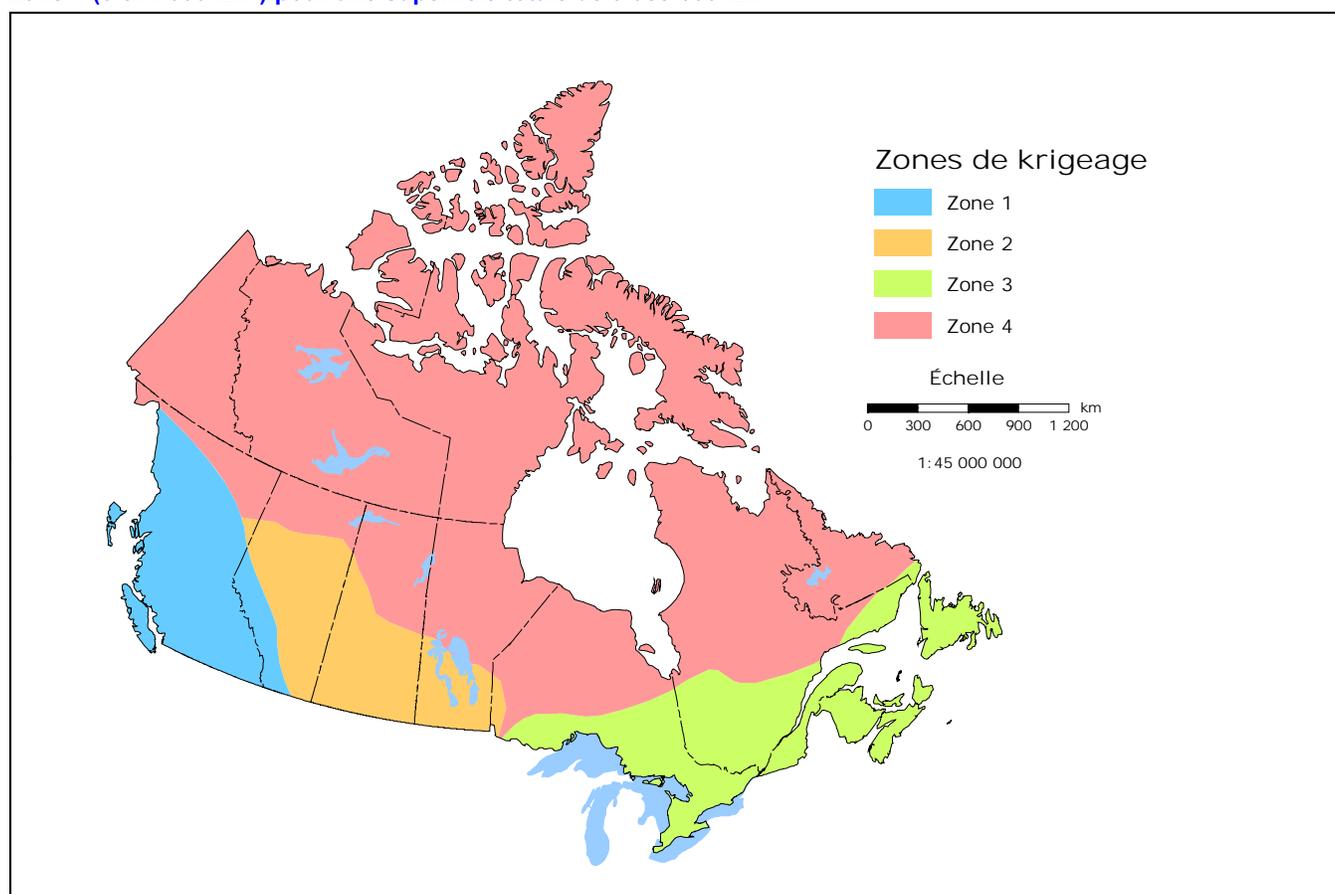


Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

La structure spatiale des données présentait un réseau assez dense dans le Sud avec des zones définies de regroupement, alors que le Nord présentait un réseau peu dense. Comme les modèles de semi-variogramme se caractérisent par la distance et le degré de variabilité entre les observations, des modèles élaborés pour s'ajuster à tout l'ensemble de données n'étaient pas appropriés. On a également constaté que les observations relatives à l'écoulement présentaient une certaine forme de directionnalité, qui pouvait être relevée où la variabilité entre les observations relatives à l'écoulement était plus marquée dans une certaine direction au sein de regroupements précis. Ces résultats indiquent que les modèles de semi-variogramme pour les régions individuelles seraient plus appropriés. Quatre zones ont été définies (carte 3) à l'aide d'une évaluation des caractéristiques physiographiques où les points étaient regroupés et d'une évaluation globale de l'agencement spatial.

Carte 3

Zones définies pour le krigeage : zone 1 (894 500 km²), zone 2 (1 003 900 km²), zone 3 (1 276 200 km²) et zone 4 (6 814 000 km²) pour une superficie totale de 9 988 600 km²



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

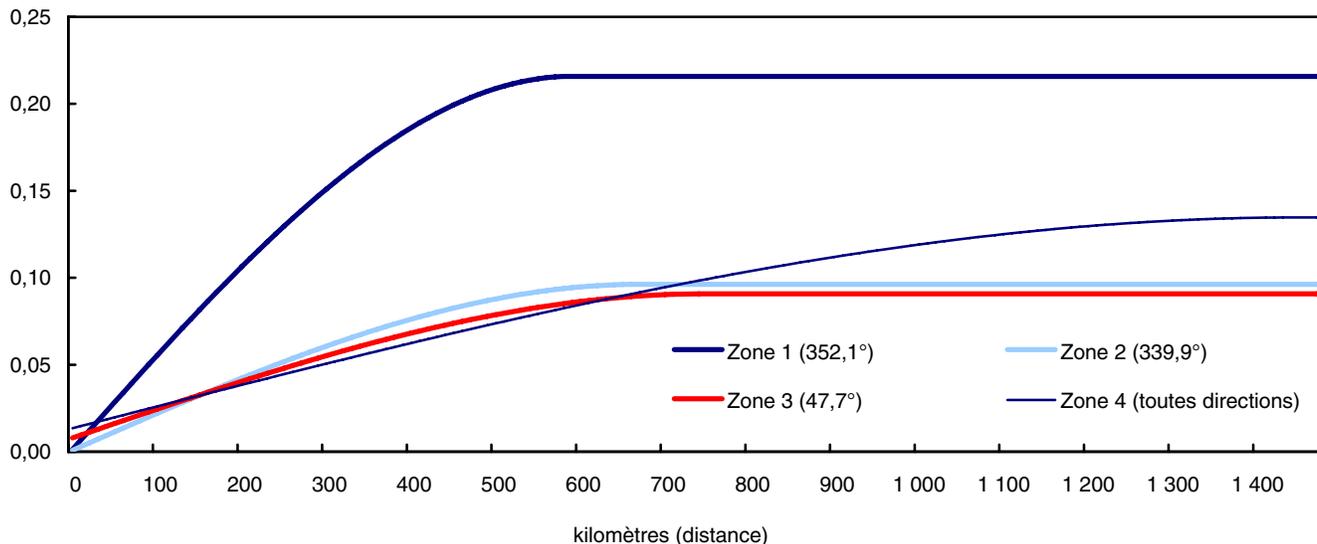
Comme l'ensemble de données était divisé en différentes zones, il était important de veiller à ce que l'interpolation ne s'effondre pas à la division des zones, puisque le krigeage offre un rendement optimal lorsque la zone faisant l'objet des prévisions se situe dans une enveloppe convexe définie par les observations (Olea, 1999). Afin d'éviter cet effondrement, les points qui se trouvaient dans une bande de 300 km autour de chaque zone étaient inclus pour produire une transition en douceur entre les zones et maintenir les tendances globales. Cette distance de 300 km

était la plus appropriée, car il y avait un nombre suffisant de données comprises dans cette bande pour permettre l'interpolation adéquate à la division des zones.⁶

Après l'attribution des données à chaque zone, la distribution de chacun des ensembles de données a été vérifiée aux fins de normalisation, et ce, pour s'assurer que la division de l'ensemble de données ne créait pas d'ensembles de données asymétriques. Des semi-variogrammes ont été créés pour chaque zone et chacun des mois des trente années à l'étude (1 440 modèles) étant donné que les observations de l'écoulement peuvent varier d'une période à l'autre. L'étendue du décalage de tous les semi-variogrammes a été établie à la distance euclidienne moyenne minimale entre tous les points se trouvant dans la zone.⁷ Le nombre approximatif de décalages a été déterminé en calculant combien de décalages multipliés par l'étendue du décalage égalerait approximativement la moitié de la distance euclidienne maximale entre les observations, étant donné que les distances (décalages) plus grandes que l'étendue n'auraient pas de points correspondants (paires). Afin de déterminer la portée de l'influence des observations, les modèles sphériques ont ensuite été ajustés aux nuages de semi-variance en indiquant la position à laquelle les échantillons devenaient indépendants les uns des autres (graphique 3).

Graphique 3
Exemple de portée maximale et de directionnalité des semi-variogrammes sphériques correspondant à chaque zone de Log₁₀ (ruissellement + 1), janvier 1980

$Y(h) \text{ Log}_{10} (\text{ruissellement} + 1)$



Note(s) : Les semi-variogrammes de chaque zone sont : Zone 1 : $y(h)=0,214 [1,5(h/588\ 929)-(h/588\ 929)^3]$, Zone 2 : $y(h)=0,095 [1,5(h/668\ 698)-(h/668\ 698)^3]$, Zone 3 : $y(h)=0,082 [1,5(h/745\ 609)-(h/745\ 609)^3]$, Zone 4 : $y(h)=0,121 [1,5(h/1\ 447\ 630)-(h/1\ 447\ 630)^3]$.

Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Tel qu'il a été mentionné précédemment, une directionnalité a été observée lorsque l'ensemble de données a été divisé en différentes zones. C'est pourquoi des semi-variogrammes anisotropes ont été calculés pour les zones 1, 2 et 3. L'anisotropie indique que les caractéristiques des directions spatiales ne sont pas toutes les mêmes et que la dépendance entre $z(x_i)$ et $z(x_i + h)$ est une fonction à la fois de la grandeur et de la direction de h (Cressie, 1993) [c'est-à-dire, qu'il est possible d'ajuster différents semi-variogrammes pour différentes directions].

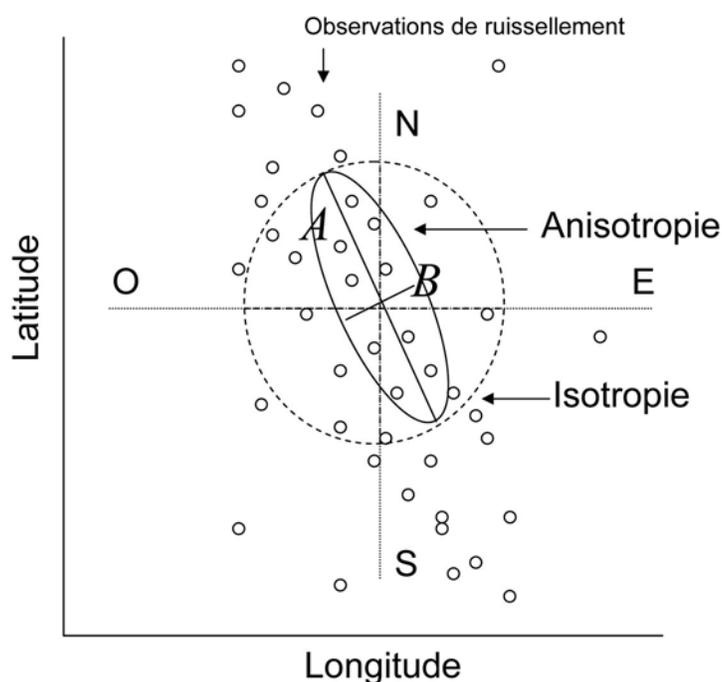
6. La distance de 300 km a été établie au moyen d'une évaluation de plusieurs distances différentes (plus grandes et plus petites) et d'une évaluation visuelle des surfaces prédites le long des divisions des zones. Les distances inférieures à 300 km ne produisaient pas une transition en douceur entre les zones, alors que les distances supérieures à 300 km ne correspondaient pas entièrement au modèle semi-variogramme élaboré à l'aide des observations dans la zone (par exemple, le modèle semi-variogramme généré pour la zone 1 ne convient pas pour les observations réalisées dans la zone 4).

7. Nous avons utilisé cette règle, car l'augmentation du décalage peut avoir un effet notable sur le semivariogramme; une augmentation moins grande que la distance minimale entre les observations de données peut entraîner un nombre insuffisant de paires d'observations, créant ainsi un semi-variogramme asymétrique (Webster et Oliver, 2001).

Par exemple, dans la zone 1, l'écoulement varie de façon semblable sur de longues distances en direction sud-est-nord-ouest, ce qui correspond à l'écoulement que l'on trouve dans les Rocheuses. Nous disposons ainsi d'une fourchette plus importante pour l'ajustement du semi-variogramme dans cette direction (figure 1). L'anisotropie dégagée par la configuration de la continuité spatiale a également défini le voisinage de recherche qui limite la superficie à rechercher pour les points. Habituellement, le voisinage de recherche est un cercle (isotrope) centré dans l'emplacement faisant l'objet de l'estimation, mais lorsque la variation est anisotrope, il est préférable d'utiliser une ellipse dont le grand axe est orienté dans la direction de la continuité maximale (Goovaerts, 1997). La zone 4 a été traitée différemment en raison du réseau épars de données. La direction faisant état de la plus grande fourchette dans le semi-variogramme a été utilisée pour le grand axe et le petit axe, formant une ellipse de recherche isotrope.

Figure 1

Exemple des ellipses de recherche définies pour les semi-variogrammes anisotropes et isotropes, A indique la portée maximale et B la portée minimale



Source(s): Statistique Canada, Division des comptes et de la statistique de l'environnement.

Le krigeage ordinaire de chaque zone a ensuite été réalisé à l'aide des paramètres établis à partir des modèles de semi-variogramme. Pour toutes les prévisions de krigeage, l'erreur-type correspondante a été calculée. L'erreur-type fournit une mesure de l'incertitude qui est fonction de l'agencement spatial des points d'observation et du modèle de semi-variogramme (Houlding, 2000). Les résultats étaient des prévisions et des erreurs-types des surfaces à une résolution de 10 km. Puisque la méthode de krigeage a été appliquée dans un espace logarithmique, les prévisions ont été rétrotransformées à l'échelle des données initiales. Cette transformation n'est pas un simple antilogarithme ($10^{Y(x_i)}$) de la prévision du krigeage, car cette méthode tend à gonfler toute erreur associée à l'interpolation de la surface (Simard et coll., 1992). La rétrotransformation de l'estimation du krigeage du $\log_{10} Z(x)$ s'effectue selon la formule suivante :

$$(6) \quad \hat{Z}(x_i) = \left(10^{\left[\frac{y(x_i) + \sigma_y^2(x_i)}{2} \right]} \right) - C,$$

où $y(x_i)$ représente la prévision du krigeage logarithmico-normale, $\sigma_y^2(x_i)$ est la variance logarithmico-normale correspondante, C est la valeur de la constante ajoutée aux données initiales avant la transformation et $\hat{Z}(x_i)$ sont les données rétrotransformées à l'échelle originale de l'écoulement (mm) (Gilbert, 1987; Deutsch et Journel, 1998). La variance $\sigma_y^2(x_i)$ est utilisée car la rétrotransformation (non linéaire) d'une estimation sans biais n'est pas sans biais en soi et, dans ce cas, il est reconnu que l'introduction de cette variance $\sigma_y^2(x_i)$ permet de réduire le biais. Même si cette méthode de rétrotransformation peut produire un léger biais, elle était considérée comme pratique dans le cadre de cette étude comme il y avait 1 440 prévisions de surface à rétrotransformer. Il existe des méthodes plus complexes qui avancent ne produire aucun biais. Consulter Yamamoto (2007) pour obtenir un exemple de ces méthodes.

Il n'y avait aucune méthode simple de transformation de l'erreur-type à l'échelle initiale des données d'écoulement. Par conséquent, l'erreur-type est présentée dans un espace logarithmique, fournissant ainsi une comparaison relative (non absolue) avec les unités d'écoulement initiales.

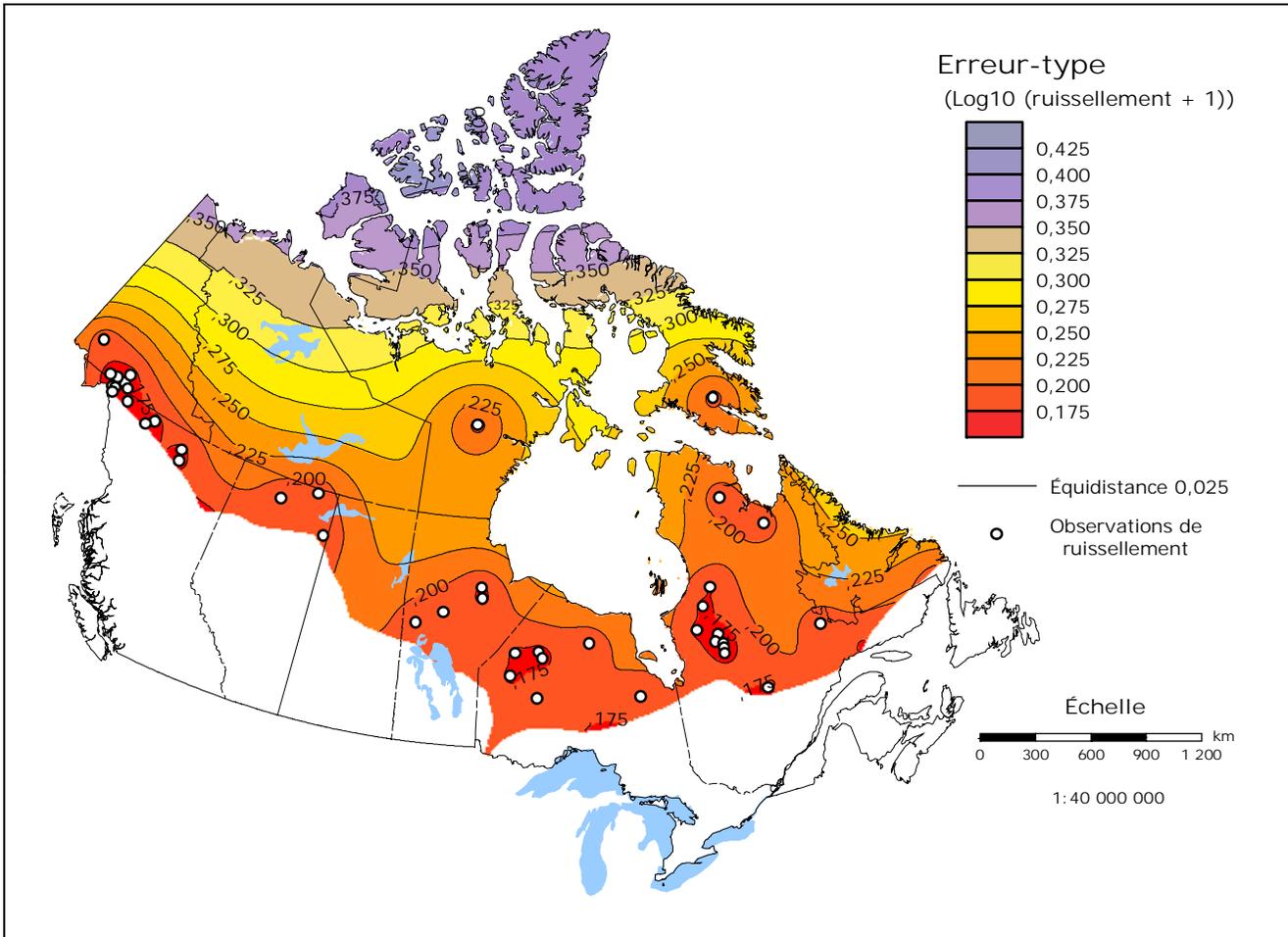
2.4 Estimation de l'apport en eau

Les zones mensuelles d'écoulement interpolé ont d'abord été fusionnées et ensuite additionnées pour produire des surfaces annuelles. Ces surfaces d'écoulement annuelles ont ensuite fait l'objet d'une moyenne pour produire la surface sur trente ans et elles ont ensuite été ramenées à un volume en fonction de la résolution (100 km²) des surfaces. Le résultat de ces calculs a donné l'estimation de l'apport en eau.

En sachant que la densité des observations utilisées pour produire l'estimation du krigeage est extrêmement variable et éparse dans certaines régions, une évaluation a été réalisée pour déterminer si certaines périodes ne contenaient pas suffisamment de renseignements pour être comprises dans l'estimation s'échelonnant sur trente ans. Les évaluations des surfaces d'erreur-type ont indiqué des régions dans lesquelles les prévisions du krigeage sont très incertaines. Par exemple, la carte des erreurs-types pour janvier 1971 pour les zones 1 et 4 peut être observée aux cartes 4 et 5, respectivement.

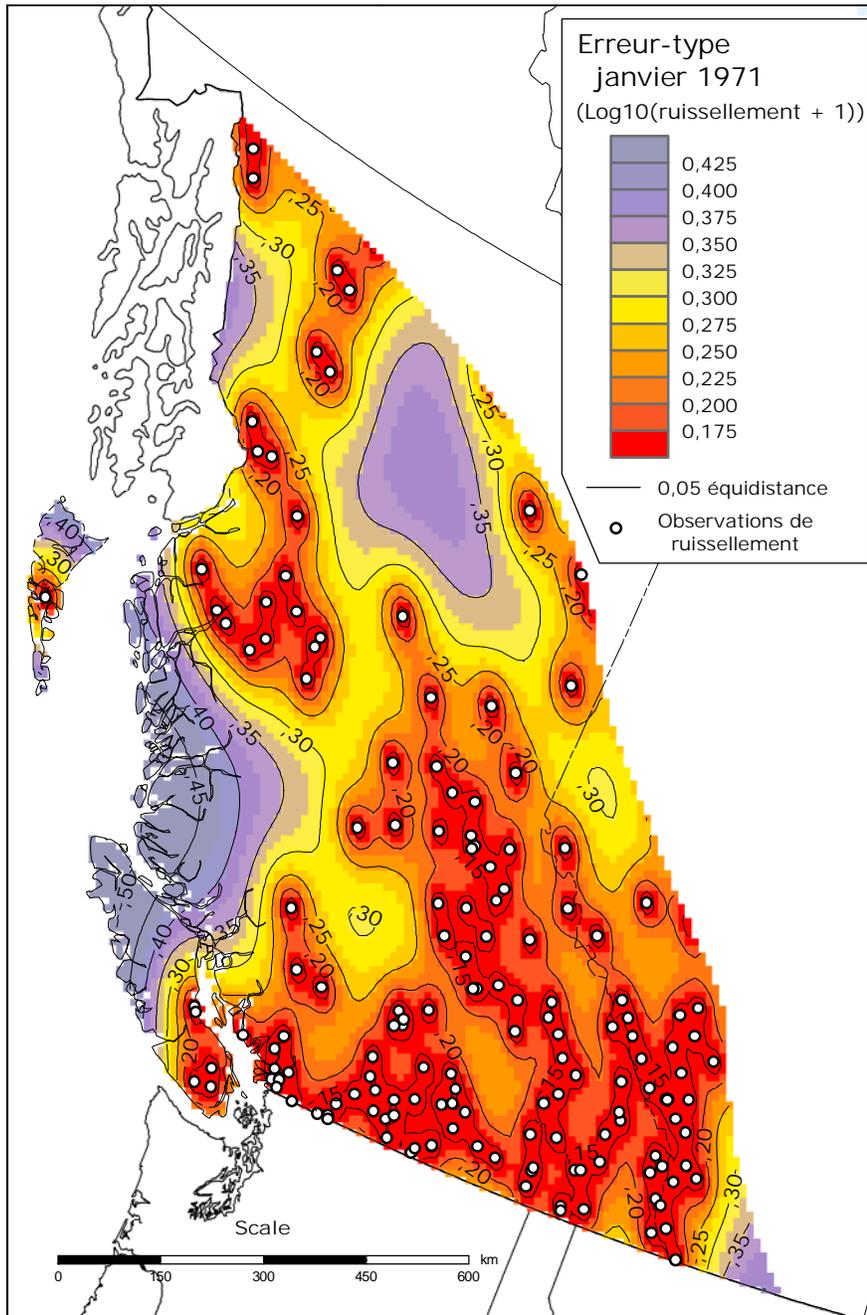
Carte 4

Carte des erreurs-types de $\text{Log}_{10}(\text{ruissellement} + 1)$ pour la zone 4, janvier 1971



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Carte 5
 Carte des erreurs-types de $\text{Log}_{10}(\text{ruissellement} + 1)$ pour la zone 1, janvier 1971



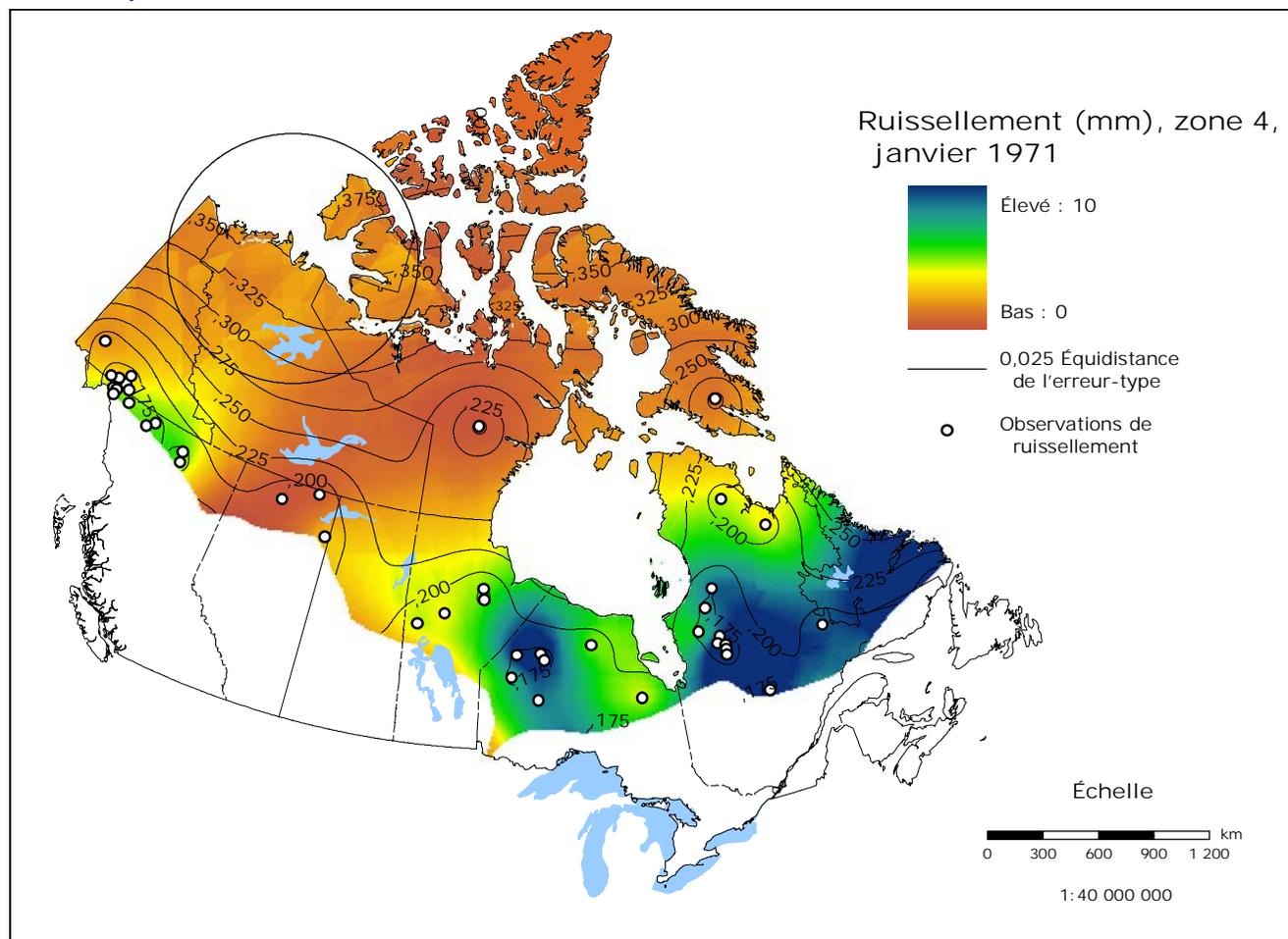
Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

D'après les cartes 4 et 5, il est clair que l'erreur-type augmente progressivement en s'éloignant des points d'observation. Toutefois, la relation entre l'erreur-type et la distance constatée dans la zone 4 (carte 4) est très différente de celle constatée dans la zone 1 (carte 5). Il est évident que le semi-variogramme correspondant influence fortement le degré d'erreur-type. Le modèle de semi-variogramme présenté dans le graphique 3 indique que les observations réalisées dans la zone 1 sont très variables sur de courtes distances par rapport à la zone 4, dans laquelle les observations sont semblables sur de grandes distances. Les prévisions du krigeage

correspondant aux zones 1 et 4 ont été évaluées et il a été déterminé que l'erreur-type élevée dans la zone 4 était liée au manque de données et était une bonne indication que des artéfacts⁸ étaient introduits dans la surface prédite (carte 6), contrairement à la zone 1, où l'interpolation ne se décompose pas et où l'erreur-type indique les régions d'incertitude dans la surface prédite (carte 7). Les surfaces d'erreur-type et prédites pour les zones 2 et 3 ont également été évaluées, mais il n'y avait ni artéfact dans les surfaces prédites semblables à celles découvertes dans la zone 4, ni point localisé d'incertitude élevée dans la surface d'erreur-type comparable à la zone 1.

Carte 6

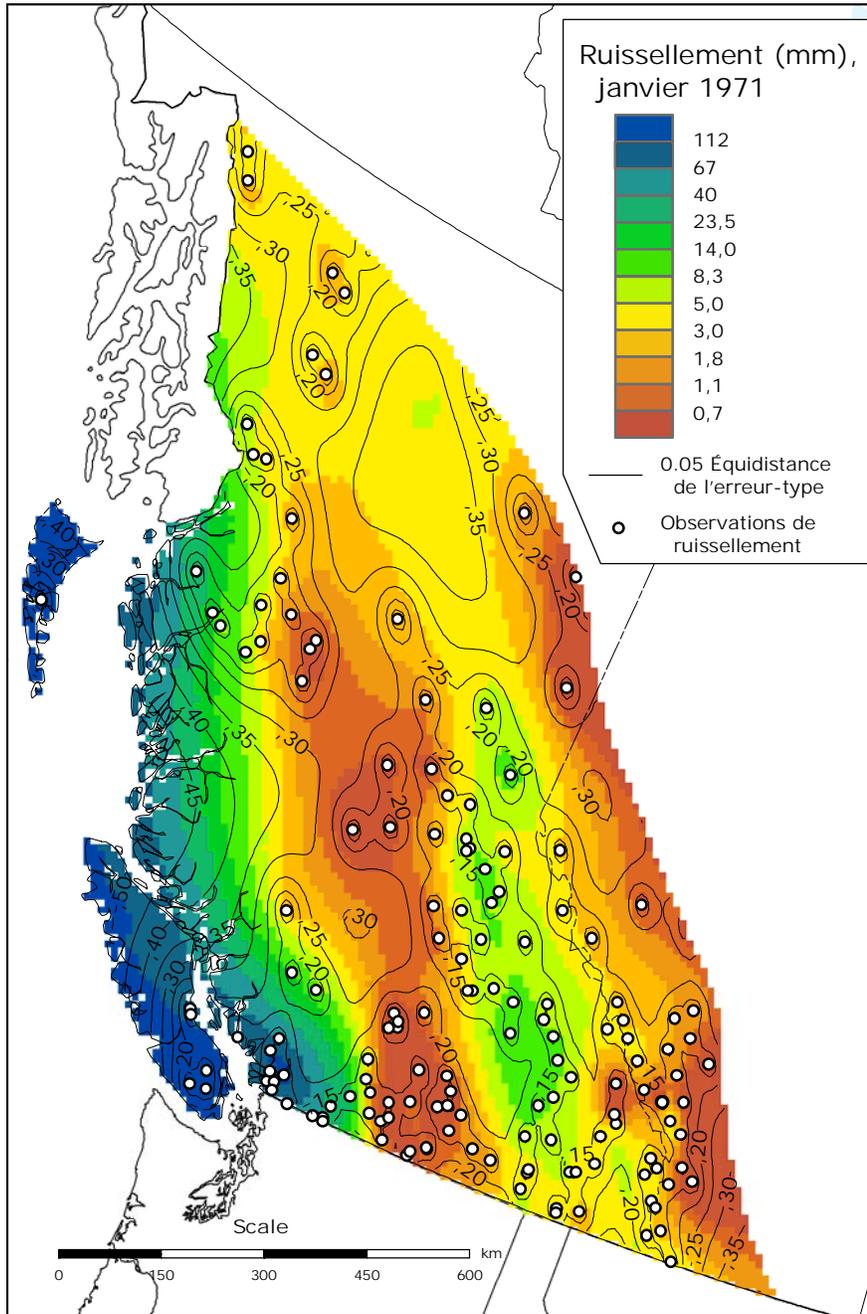
Écoulement krigé (en mm) pour la zone 4, janvier 1971, avec un calque de courbes de niveau des erreurs-types dans les unités Log_{10} (ruissellement +1). Le cercle noir délimite une région dans laquelle on peut trouver des artéfacts dans la surface prédite



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

8. Les artéfacts sont définis ici comme des ruptures de la surface continue causées par des changements abrupts dans les valeurs pour lesquelles il devient clair, visuellement parlant, que l'interpolation ne fonctionne pas bien. Ce phénomène se produit habituellement lorsque l'interpolation s'étend à des zones qui se trouvent en dehors de l'enveloppe convexe formée par les observations.

Carte 7
Écoulement krigé (en mm) pour la zone 1, janvier 1971, avec un calque de courbes de niveau des erreurs-types dans les unités Log_{10} (ruissellement +1)



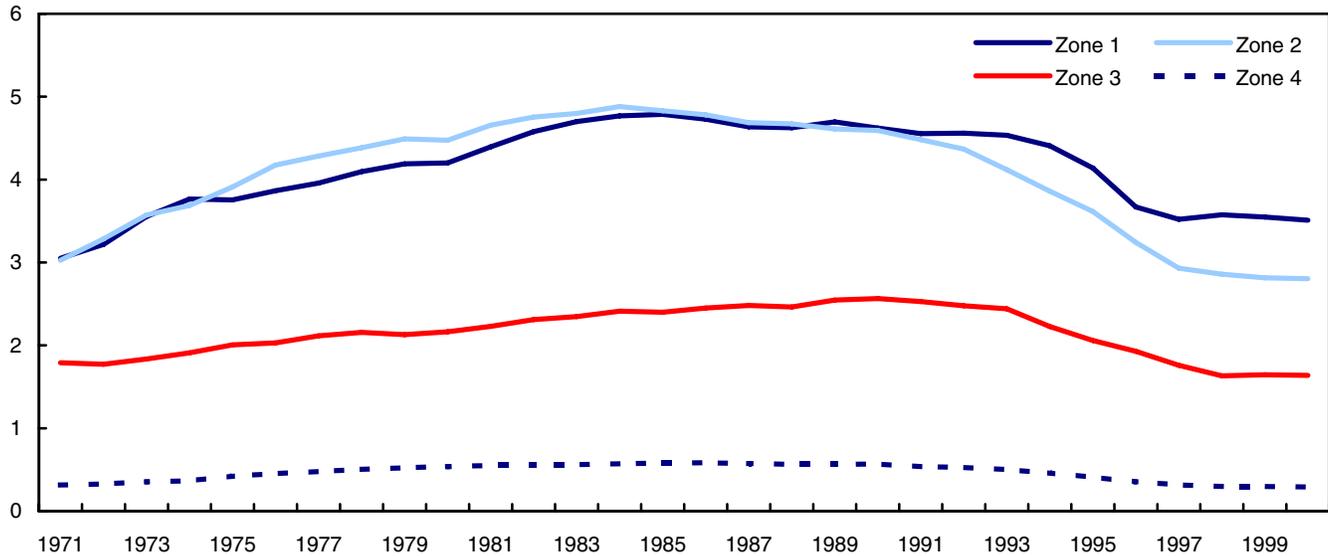
Source(s) Statistique Canada, Division des comptes et de la statistique de l'environnement.

Afin de mieux comprendre la manière dont la densité des points change au fil des années, la densité d'observation moyenne par 100 km² pour chaque année a été calculée pour chaque zone en divisant le nombre d'observations par l'aire de chaque zone (graphique 4). Celle-ci donne une estimation approximative de la densité, mais elle ne tient pas compte de l'agencement spatial (la concentration) des observations dans chaque zone. Le graphique 4

montre la très faible densité de la zone 4 par rapport aux autres zones. Il indique aussi que la densité varie au fil du temps.

Graphique 4
Densité d'observation par 10 000 km² pour chaque zone, 1971 à 2000

nombre par 10 000 kilomètres carrés (km²)

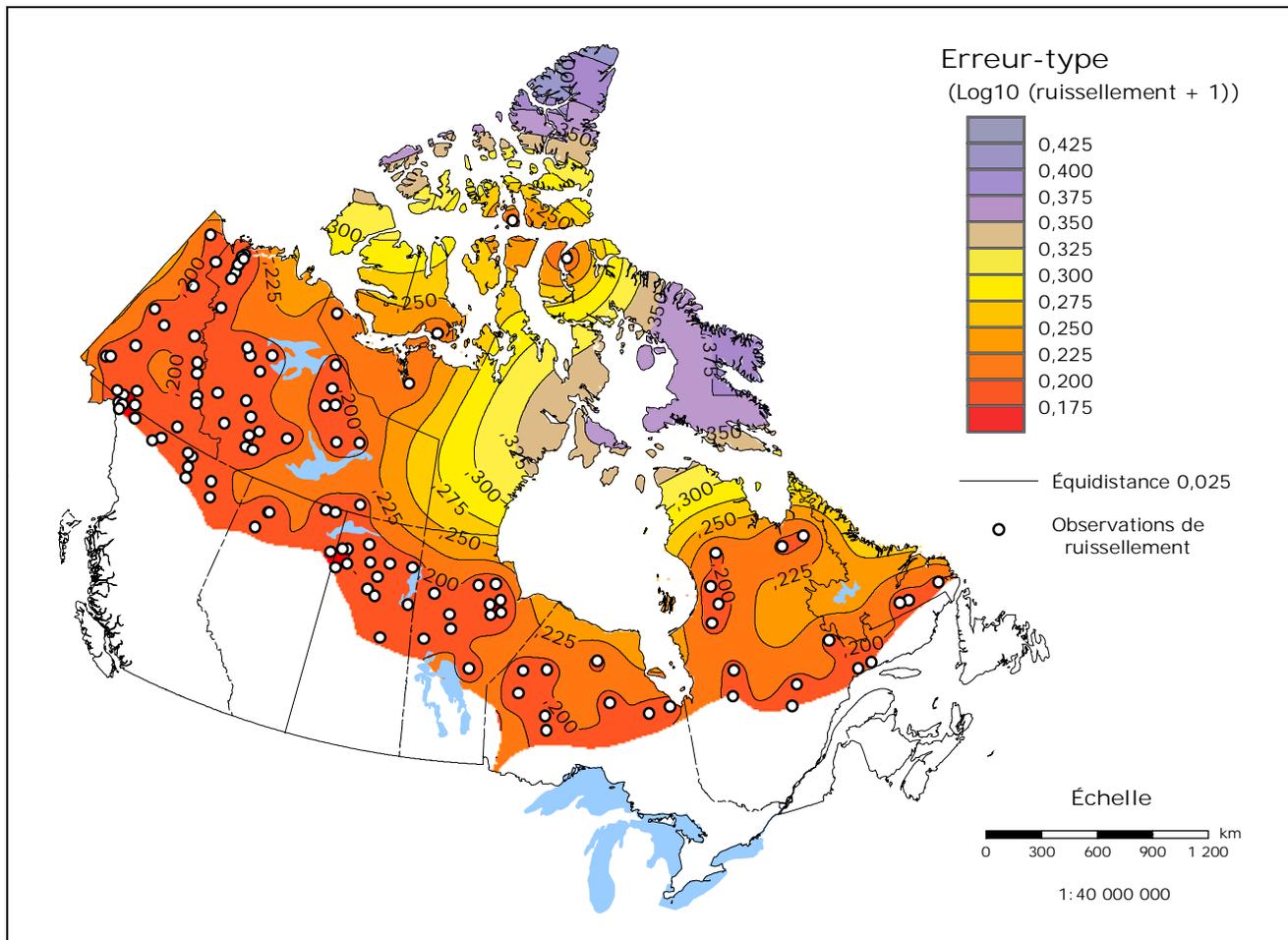


Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Les découvertes liées à l'erreur-type et à la densité des données ont indiqué qu'il serait mieux d'exclure certains endroits dans lesquels la densité des données était faible ou dans lesquels l'erreur-type était élevée. Toutefois, il n'était pas possible de définir une limite d'incertitude visant l'erreur-type pour chaque surface prédite, car l'erreur-type du krigeage dépend du semi-variogramme. Dès lors, les surfaces d'erreur-type ne sont pas directement comparables. Bien que les degrés d'incertitude soient connus dans chaque surface prédite, la préoccupation majeure était de supprimer les artéfacts de l'interpolation qui avaient été découverts dans la zone 4. Les artéfacts diminuent entre 1971 et 1976, lorsque la densité des données augmente (graphique 4). Les artéfacts deviennent pratiquement inexistantes à partir de 1977 pour la région se trouvant sous l'archipel Arctique (carte 8).

Carte 8

Exemple de carte des erreurs-types de $\text{Log}_{10}(\text{ruissellement} + 1)$ pour la zone 4, janvier 1981, lorsque la densité des données est élevée



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

De 1993 à 1996, la densité des données est de nouveau réduite, mais l'agencement spatial est bien réparti sur la zone et il n'y a pas d'artéfacts flagrants. De 1997 à 2001, la densité d'observation diminue encore et la répartition favorable est perdue. Pour cette raison, la moyenne annuelle sur trente ans pour la zone 4 ne comprend que les données comprises entre 1974 et 1997. Afin d'améliorer davantage la zone 4, et étant donné que l'archipel Arctique n'est jamais exempt d'artéfacts, les recherches menées par Spence et Burke (2008) ont été introduites comme une valeur fixe au cours des séries chronologiques. Cette valeur (211 mm ou 189 km³) est une moyenne annuelle pour une période de vingt-trois années allant de 1972 à 1994 et elle comprend environ 70 pour cent de l'archipel Arctique (895 000 km²). Cette estimation s'appuie sur les observations enregistrées entre mai et octobre, car les données étaient trop éparpillées durant les mois d'hiver.

2.5 Validation des estimations de l'apport en eau

La méthodologie et les estimations finales relatives à l'apport en eau ont été validées à l'aide de deux techniques. La première était une évaluation de la stabilité des estimations de l'apport en eau, qui a été effectuée en supprimant les données d'entrée de façon répétitive et en observant la manière dont l'interpolation fonctionnait. La seconde technique a eu recours à des mesures des différences pour quantifier le degré de désaccord entre les valeurs observées (valeurs de la base de données HYDAT divisées par les bassins contribuant de HYDEX ou ARAP) et

les valeurs prédites des surfaces d'écoulement à l'aide d'un ensemble de données retenues des données d'entrée. L'archipel Arctique a été exclu du processus de validation, comme nous avons utilisé les résultats de Spence et Burke (2008) pour cette région.

2.5.1 Analyse de la stabilité

La stabilité des estimations de l'apport en eau a été évaluée à l'aide d'une technique « bootstrap » au cours de laquelle l'interpolation a été effectuée 100 fois avec une sélection aléatoire de 10 pour cent des observations retirées de chaque zone. Cette technique visait à comprendre la manière dont les estimations de l'apport en eau calculées à partir des surfaces interpolées changent avec un ensemble de données d'entrée réduit. Cette analyse a été réalisée pour chaque zone chaque mois de l'année 1990, étant donné que cette année-là, le mois de février contenait le nombre moyen d'observations dans la base de données, ce qui devait donner une bonne indication du comportement du modèle au cours des autres années.

2.5.2 Mesures des différences

Les surfaces interpolées (profondeurs d'écoulement) ont été évaluées de manière plus approfondie en calculant l'erreur moyenne (ME) et la racine de la moyenne des erreurs quadratiques (RMSE) de la manière suivante :

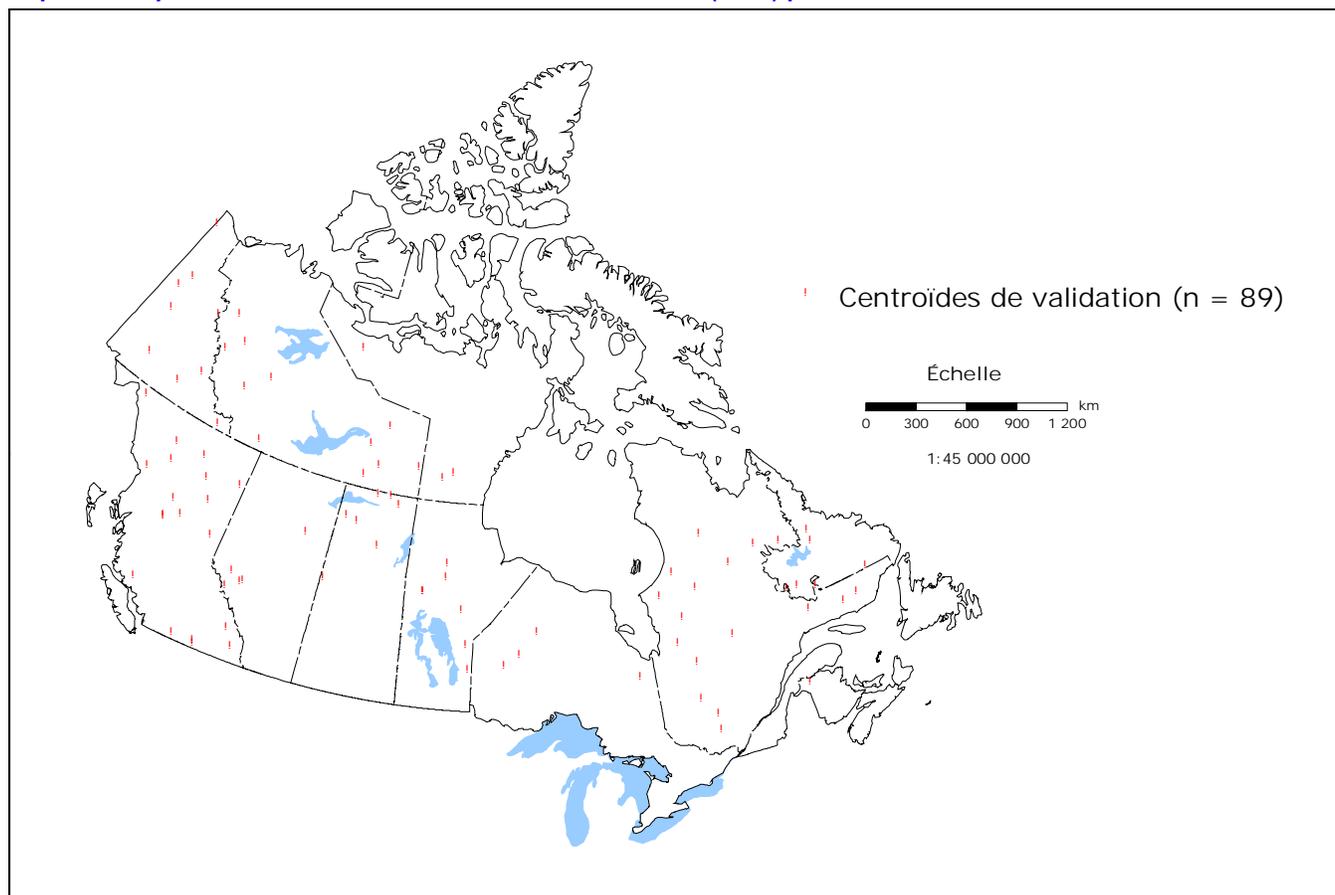
$$(7) \quad ME = \frac{1}{n} \sum_{i=1}^n [z(x_i) - \hat{z}(x_i)],$$

$$(8) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [z(x_i) - \hat{z}(x_i)]^2},$$

où $z(x_i)$ est la valeur observée (connue) et $\hat{z}(x_i)$ est la valeur prédite. La valeur ME doit être égale à zéro si la méthode d'interpolation est non biaisée. La valeur RMSE fournit une mesure de la précision où une fois encore, la valeur idéale serait égale à zéro. Ces mesures des différences ont été calculées au moyen des données qui répondaient à l'ensemble des critères dans la base de données (tels qu'ils sont présentés à la section 2.1.1), excepté le fait que la zone contribuant correspondante à ces observations dépasse 5 000 km². L'ensemble des données de validation contenait 89 stations hydrométriques avec des bassins d'une taille allant de 5 020 km² à 9 890 km². La carte 9 présente l'emplacement de ces observations de validation. Un diagramme de dispersion a également été produit de $z(x_i)$ par rapport à $\hat{z}(x_i)$ pour chaque point de données utilisé dans la validation lors de chaque période (n=19583). Les mesures des différences présument que les valeurs observées ne sont pas erronées.

Carte 9

Répartition spatiale de l'ensemble des données de validation (n=89) pour les mesures de différences



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

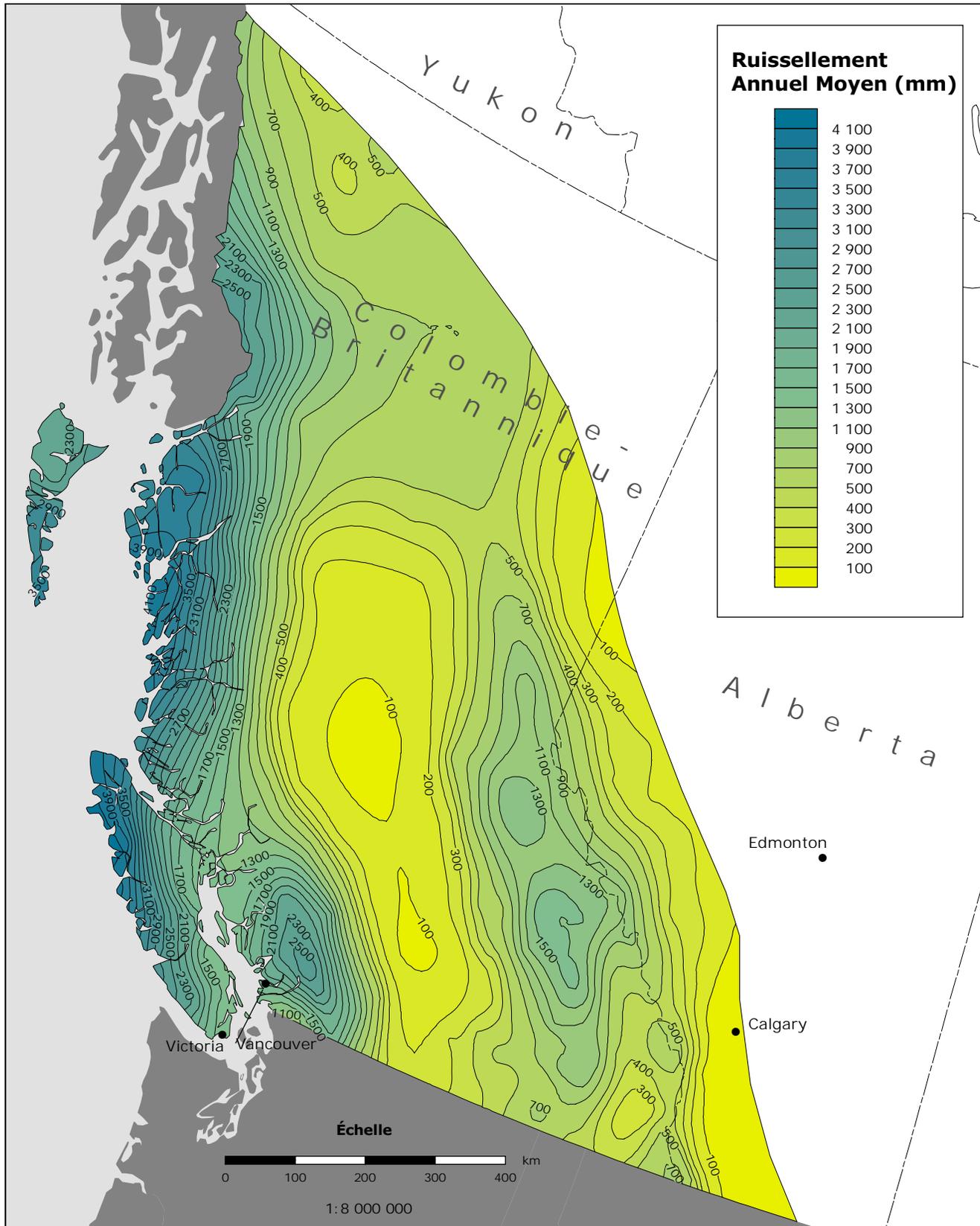
3 Résultats

3.1 Estimations de l'apport en eau

Le calcul de l'estimation annuelle de l'apport en eau sur trente ans au Canada a donné 3 435 km³/an. Les surfaces des profondeurs d'écoulement moyennes sur trente ans qui ont été utilisées pour produire l'estimation de l'apport en eau peuvent être observées sur les cartes 10, 11, 12 et 13.

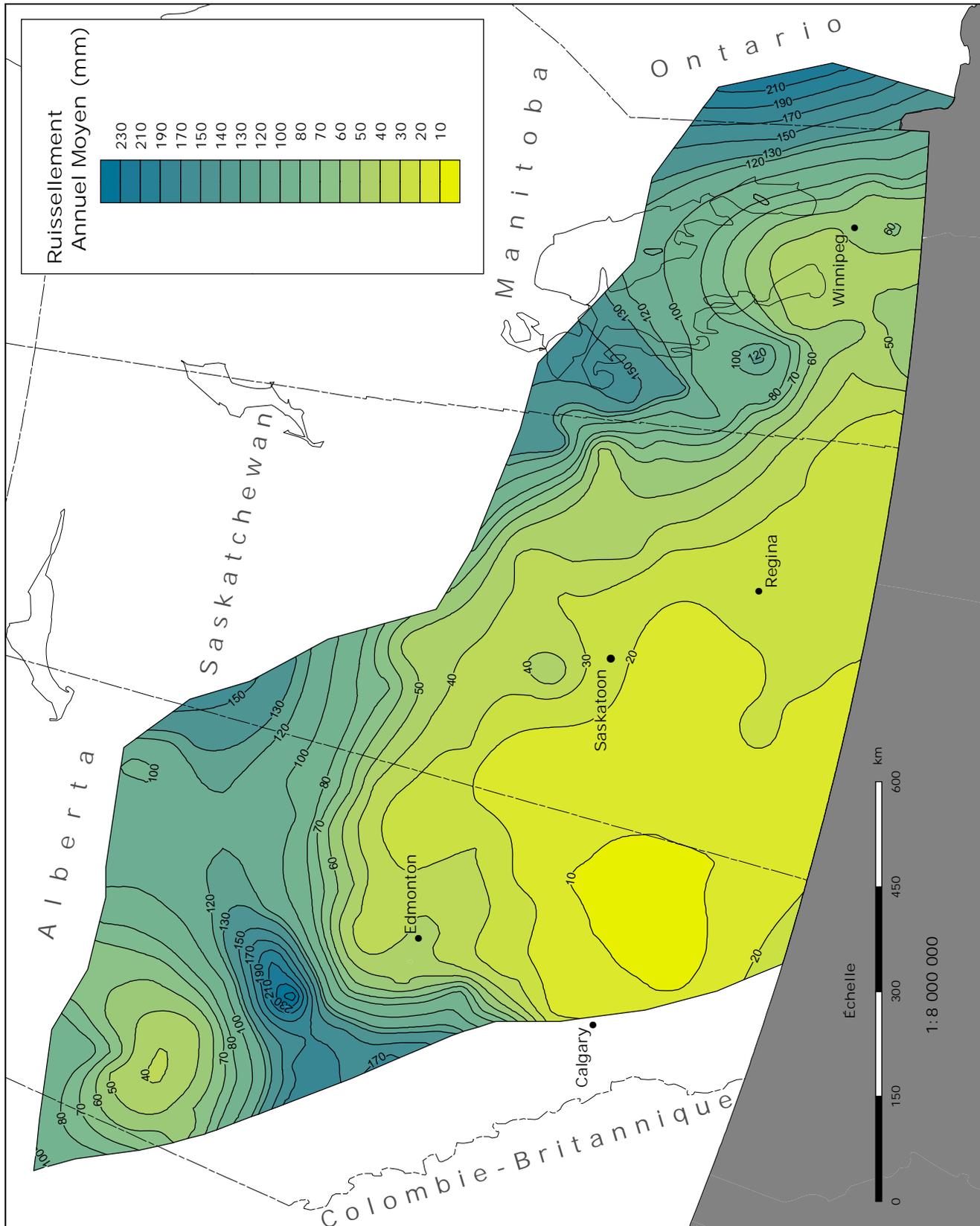
Carte 10

Profondeur moyenne de ruissellement estimée sur trente ans (1971 à 2000) pour la zone 1



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

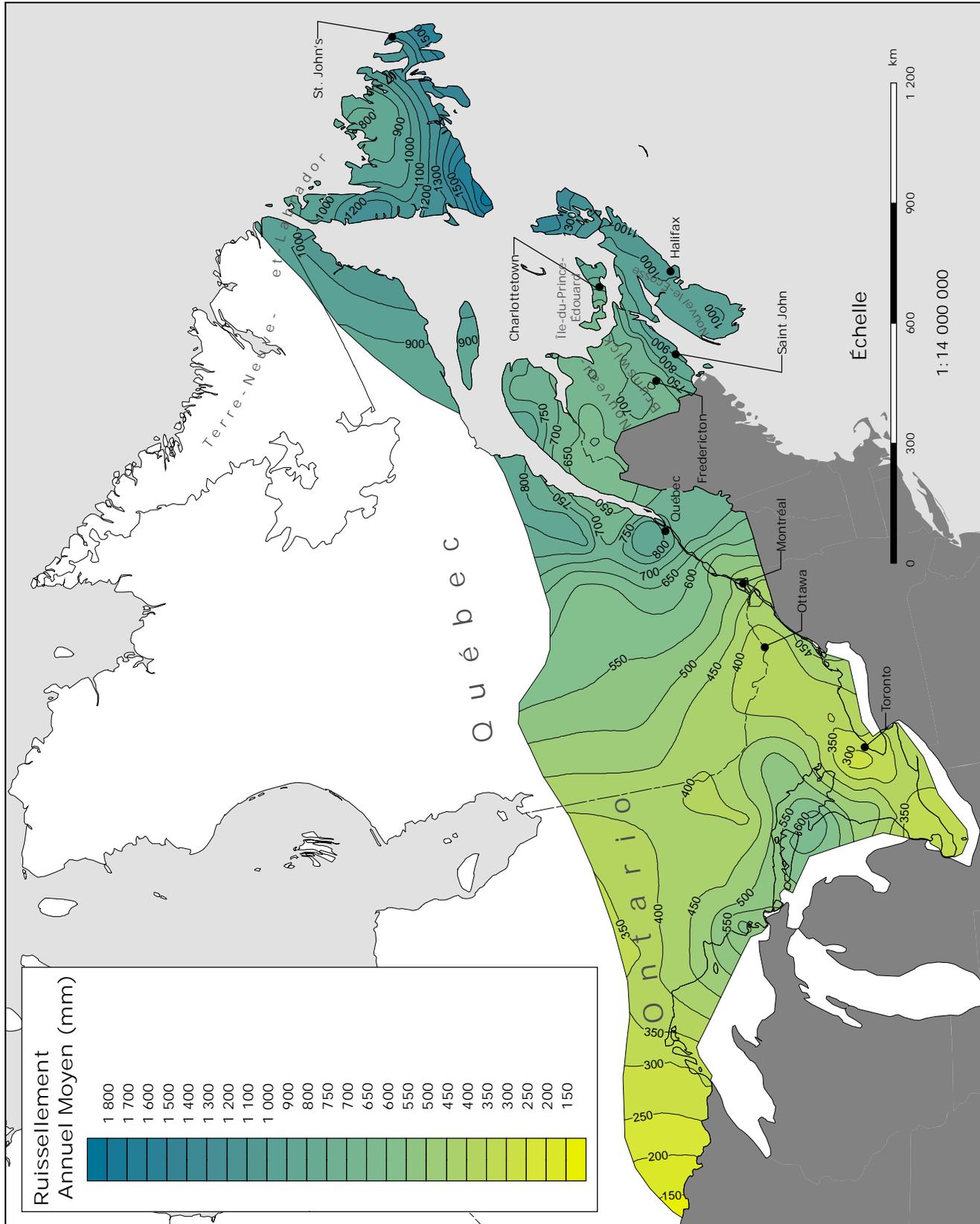
Carte 11
Profondeur moyenne de ruissellement estimée sur trente ans (1971 à 2000) pour la zone 2



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

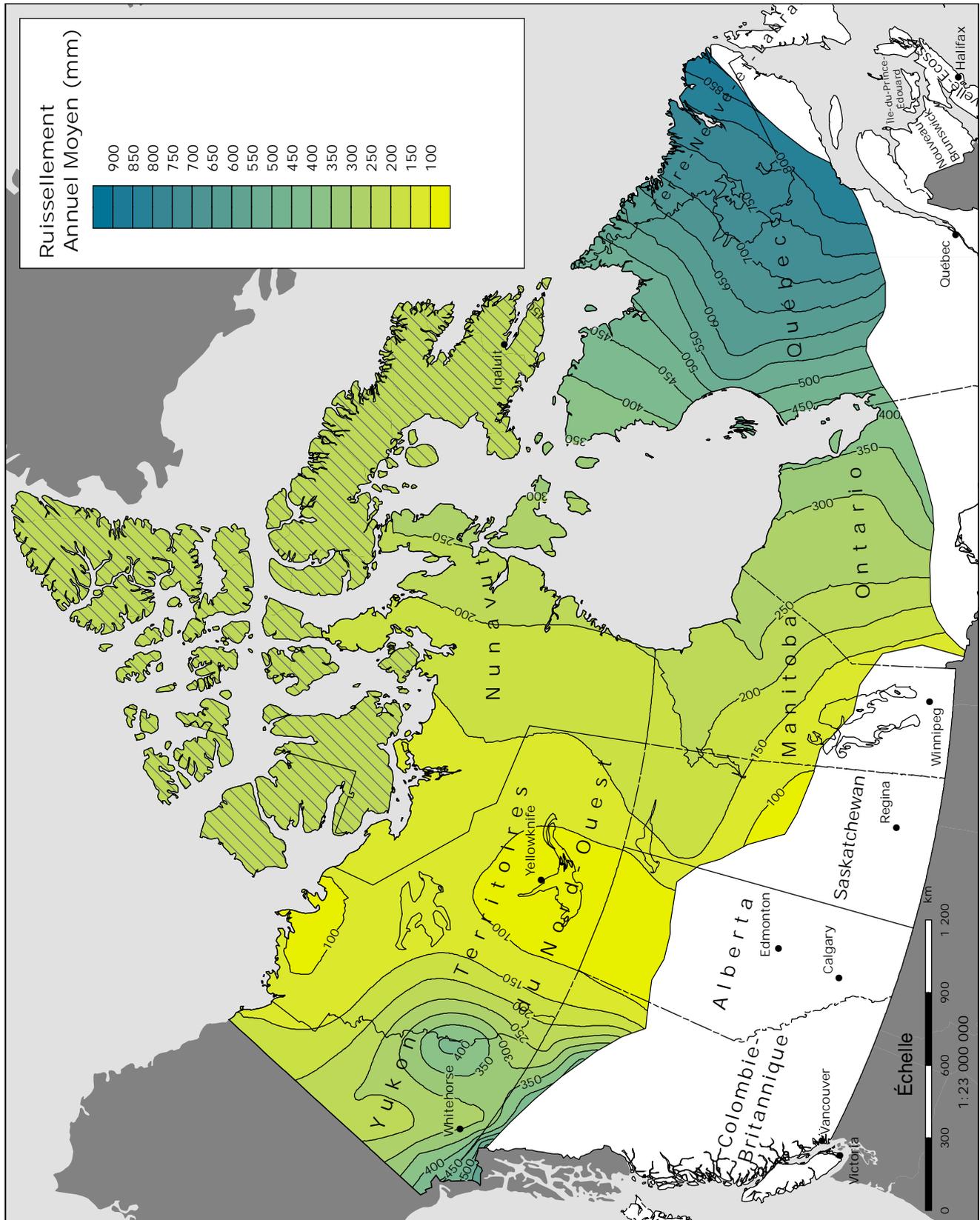
Carte 12

Profondeur moyenne de ruissellement estimée sur trente ans (1971 à 2000) pour la zone 3



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Carte 13
 Profondeur moyenne de ruissellement estimée sur trente ans (1971 à 2000) pour la zone 4



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

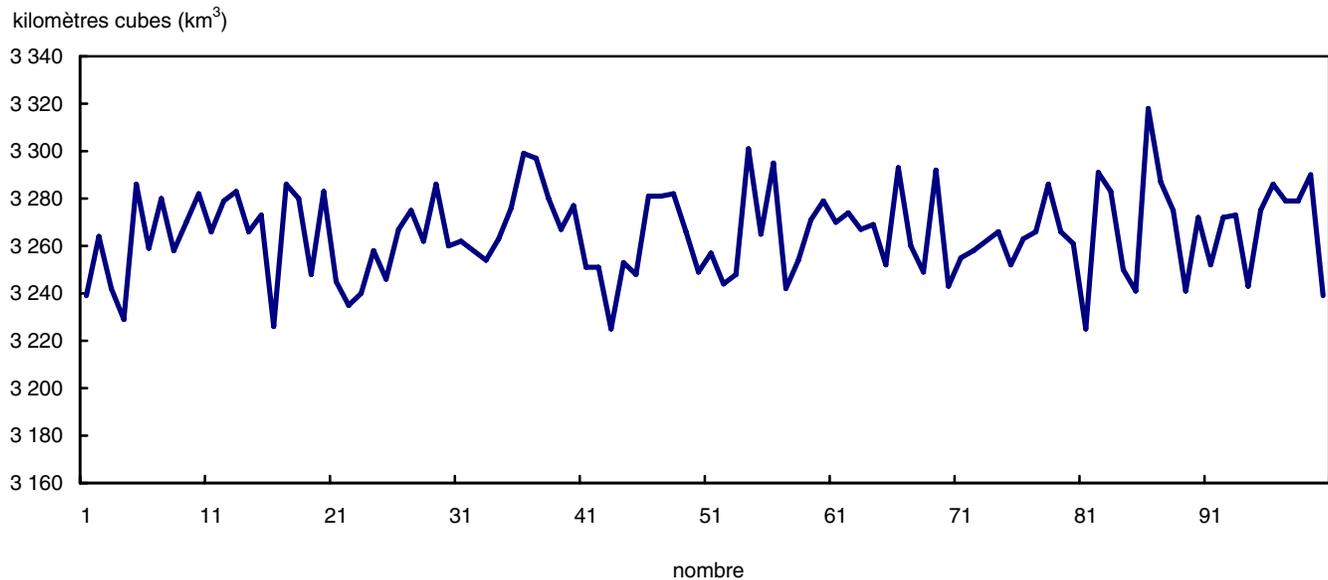
3.2 Validation

3.2.1 Analyse de la stabilité

L'analyse de stabilité a montré que l'estimation de l'apport en eau avait une portée maximale de 96 km³ pour l'apport moyen annuel en eau pour 1990 (archipel Arctique exclue), avec le retrait aléatoire de 10 pour cent des observations de chaque zone, pour tous les mois (graphique 5). L'apport en eau moyen prédit lors de cette analyse était de 3 265 km³ (graphique 6), ce qui est seulement 8km³ de moins que l'estimation actuelle de 3 273 km³. La portée maximale de 96 km³ correspond à 2,9 % de l'estimation de 3 273 km³.

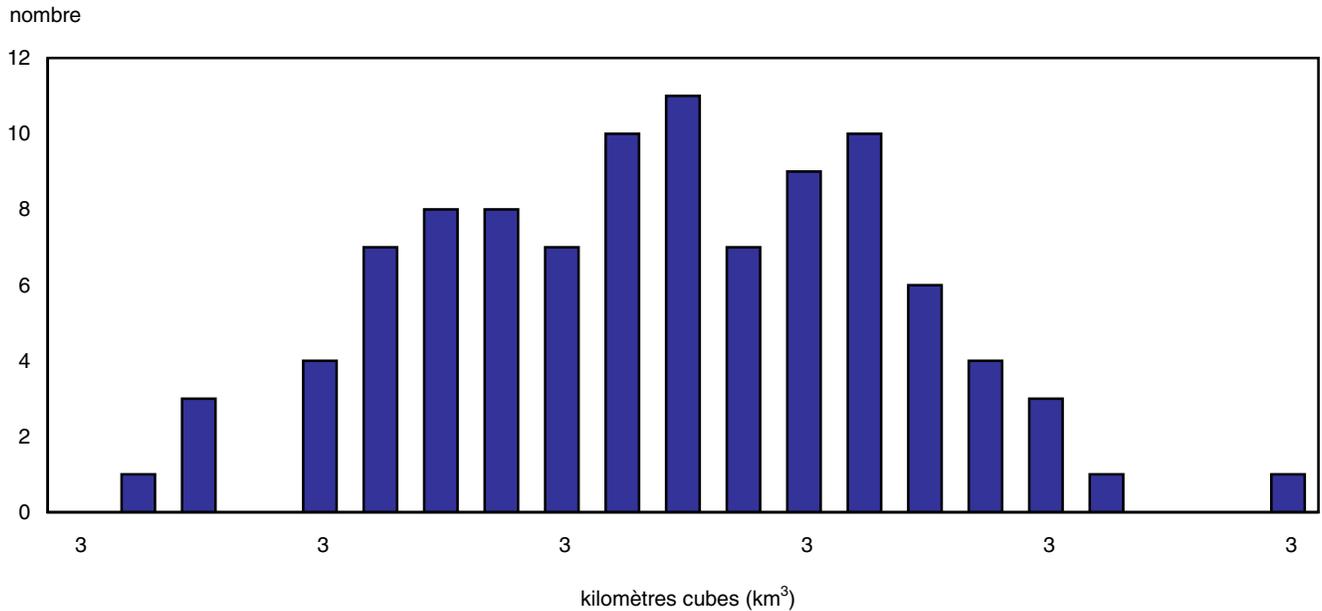
Graphique 5

Résultats de l'analyse de stabilité des estimations de l'apport en eau, 1990



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Graphique 6
Distribution des fréquences des estimations de l'apport en eau par l'entremise de 100 itérations avec un retrait aléatoire de 10 p. cent des données

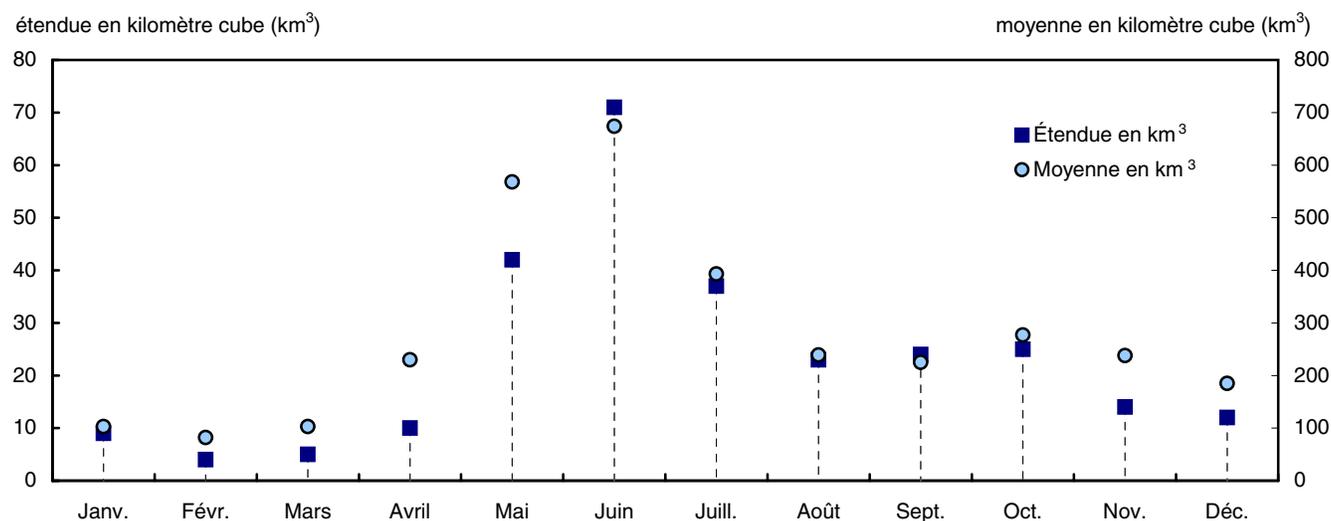


Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Le graphique 7 montre que les mois de mai, juin et juillet contribuent le plus largement à l'instabilité des estimations de l'apport annuel en 1990. Ce phénomène s'explique par le fait qu'à l'échelle nationale, cette période représente les apports les plus élevés et donc la probabilité de variation la plus élevée.

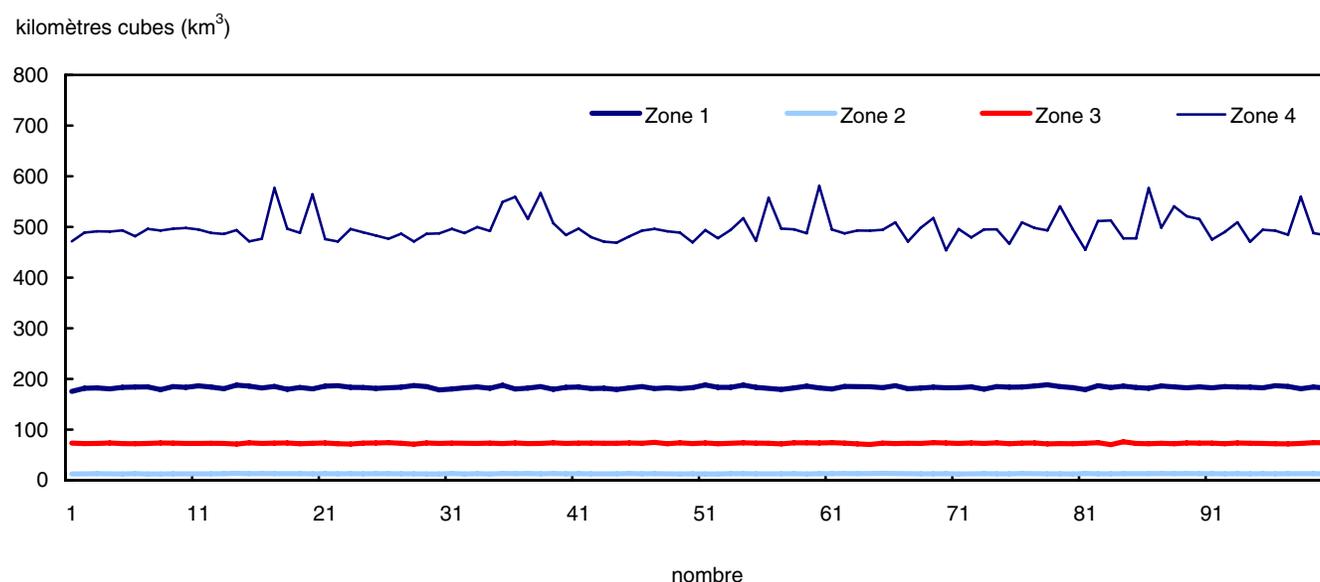
Le graphique 8 montre que la variabilité la plus élevée dans l'estimation de l'apport pour le mois de juin se trouvait dans la zone 4 (portée = 126,5 km³), suivie de la zone 1 (portée = 10 km³). La zone 4 est très variable, car les données sont très éparpillées avec la majorité des observations regroupées le long de la limite méridionale. La portée trouvée dans la zone 1 est liée à l'extrême variabilité entre les observations relatives à l'écoulement. Certaines observations sont très élevées ou faibles par rapport aux observations voisines. L'estimation variera selon que ces valeurs extrêmes seront ajoutées ou exclues. Les estimations pour les zones 2 et 3 sont restées relativement constantes avec des portées respectives de 1,2 km³ et 5,7 km³.

Graphique 7
Portée des estimations de l'apport en eau et apport en eau moyen estimé sur 100 itérations



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Graphique 8
Résultats de l'apport en eau de l'analyse de stabilité, par zone, juin 1990



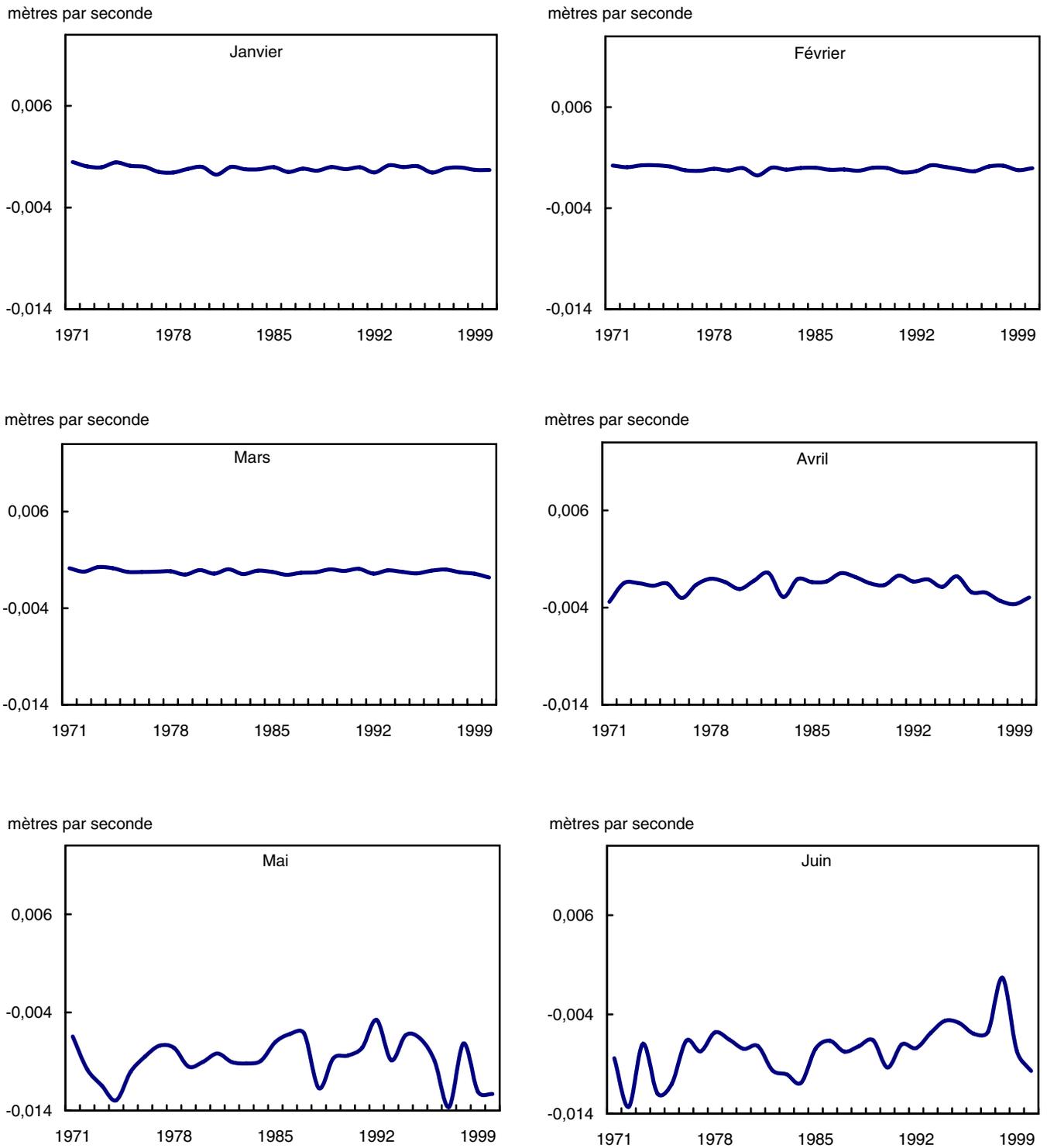
Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

3.2.2 Mesures des différences

Les résultats pour la ME et la RMSE peuvent être observés aux graphiques 9, 10 et 11, 12, respectivement. Les graphiques révèlent que les valeurs d'écoulement prédites sont en moyenne plus élevées que les valeurs observées (ME négative), en particulier pour mai, juin et juillet, tout au long des trente années. On pense que les différences sont causées par l'augmentation de la variabilité de l'écoulement au cours de ces mois, comme il y a probablement des changements localisés qui ne sont pas inclus dans les observations d'écoulement environnantes qui déterminent la valeur prédite. Globalement, les mesures des différences indiquaient que les surfaces prédites se trouvent dans

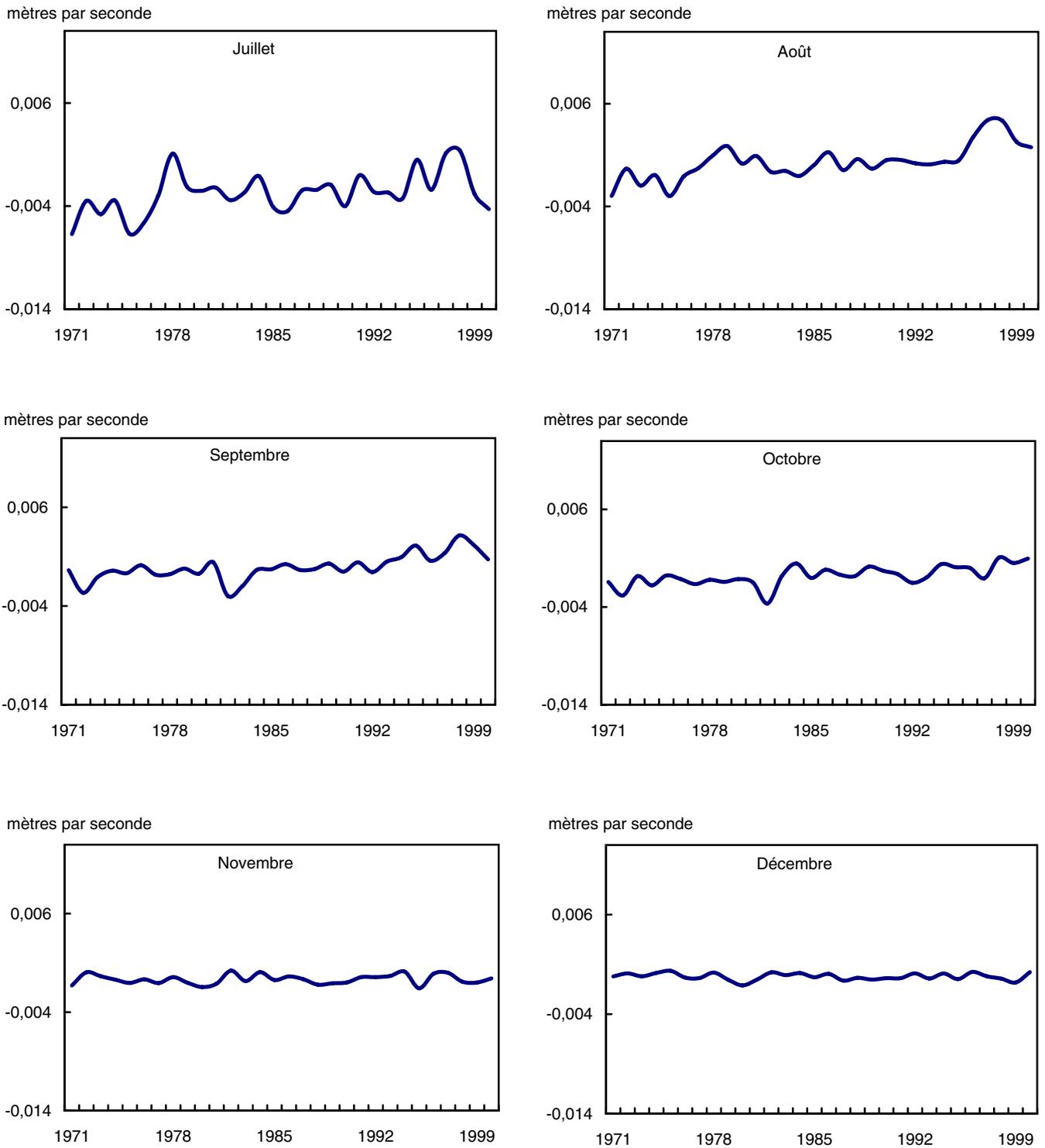
une portée raisonnable des valeurs de validation et que les différences principales sont présentes pendant les mois au cours desquels l'écoulement est reconnu comme hautement variable. Le diagramme de dispersion a produit une forte corrélation ($r = 0,78$, $p \leq 0,001$) entre toutes les valeurs observées et prédites ($n=19583$) utilisées dans l'ensemble des données de validation au cours des trente ans de données (figure 2).

Graphique 9
Erreur moyenne (ME) des profondeurs d'écoulement interpolées, de janvier à juin, 1971 à 2000



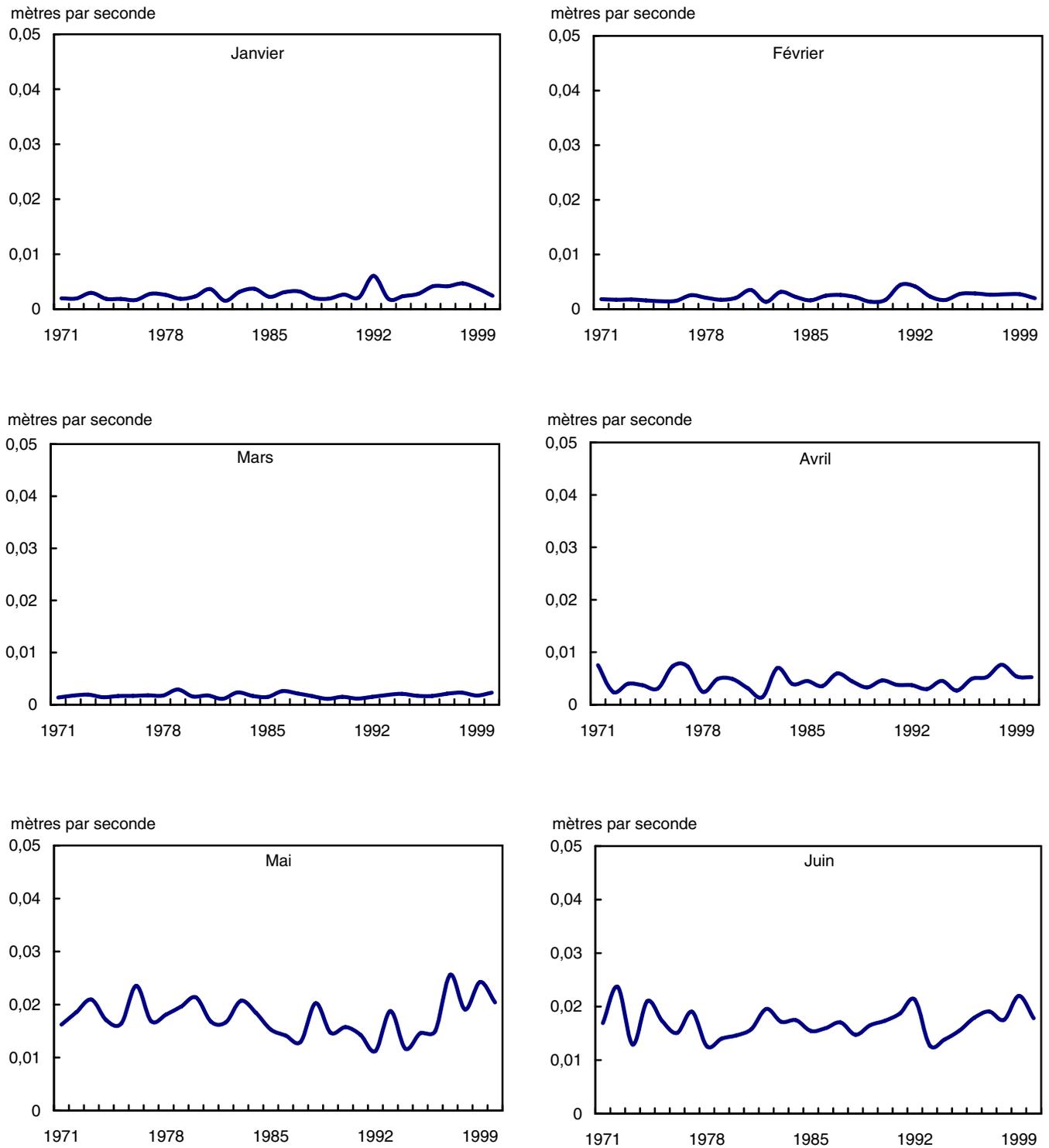
Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Graphique 10
Erreur moyenne (ME) des profondeurs d'écoulement interpolées, de juillet à décembre, 1971 à 2000



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Graphique 11
Erreur quadratique moyenne (RMSE) des profondeurs d'écoulement interpolées, de janvier à juin, 1971 à 2000

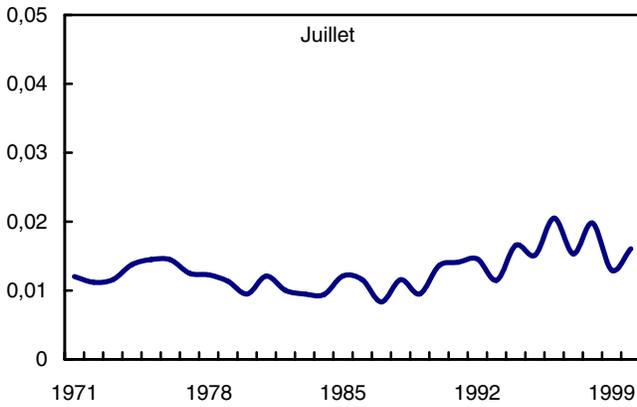


Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

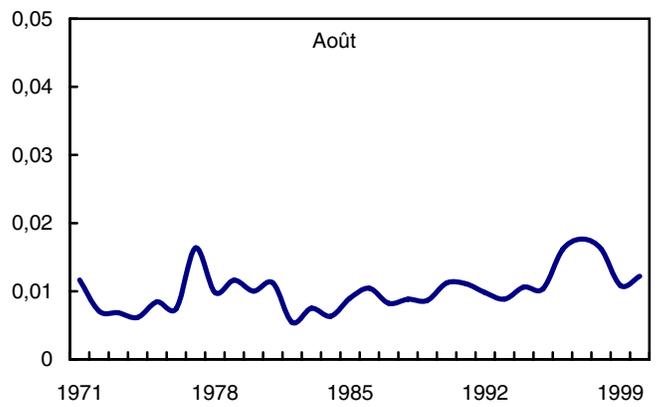
Graphique 12

Erreur quadratique moyenne (RMSE) des profondeurs d'écoulement interpolées, de juillet à décembre, 1971 à 2000

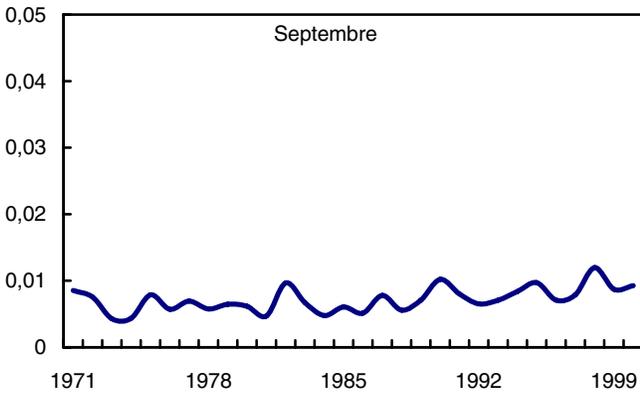
mètres par seconde



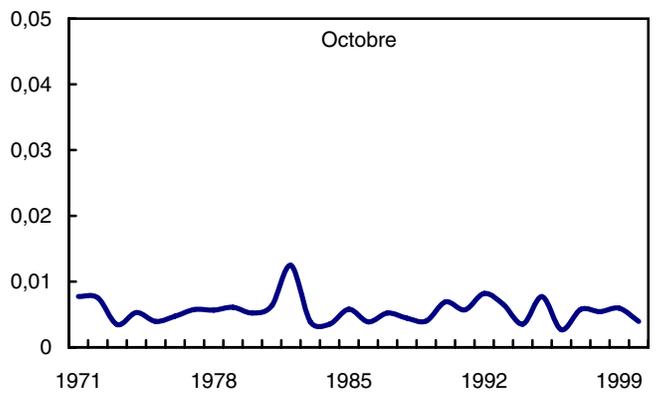
mètres par seconde



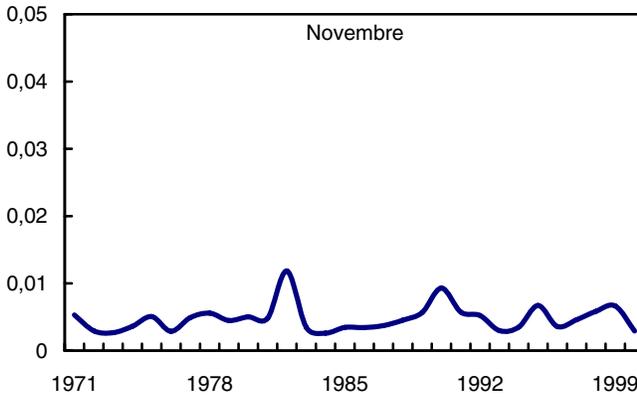
mètres par seconde



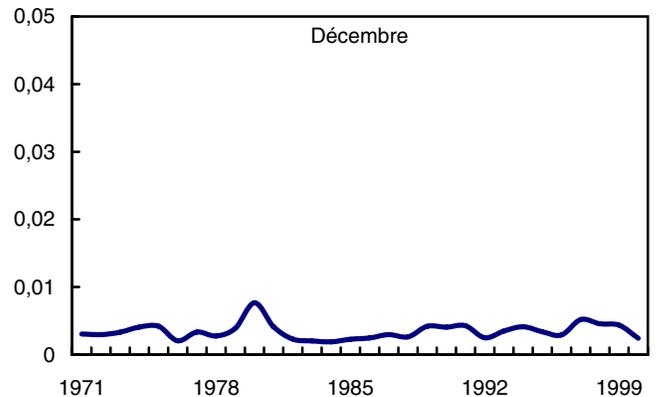
mètres par seconde



mètres par seconde



mètres par seconde

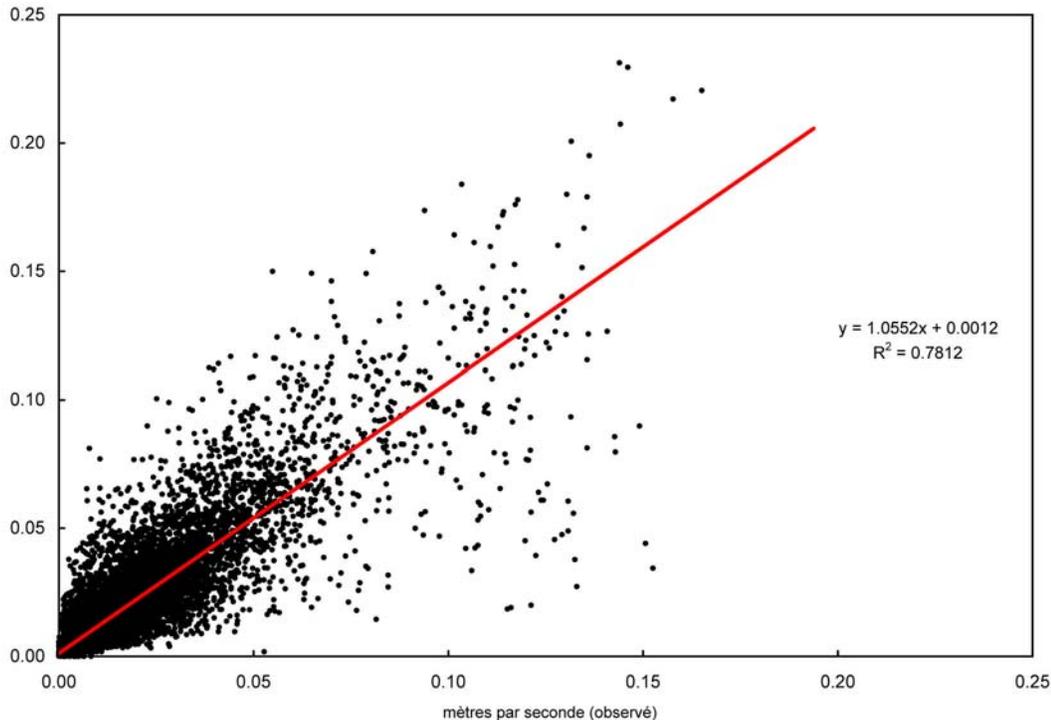


Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Figure 2

Profondeur d'écoulement prédite par opposition à profondeur d'écoulement observée aux stations de validation, 1971 à 2000

mètres par seconde (prévu)



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

4 Conclusion et orientation future

Ce papier présentait une description détaillée de la méthodologie de Statistique Canada visant à produire une estimation annuelle sur trente ans de l'apport en eau au Canada. Les résultats de cette étude révèlent que l'apport en eau annuel moyen au Canada sur une période de trente ans allant de 1971 à 2000 est de 3 435 km³. Il s'agit de la première estimation exhaustive du débit d'eau douce à l'échelle nationale qui utilise une méthodologie systématique semi-automatisée appliquée de manière uniforme dans tout le pays.

L'estimation finale a été validée au moyen de deux approches. La première a testé la stabilité du modèle d'apport en eau lorsque 10 pour cent des données ont été retirées de manière répétitive du modèle. L'analyse de stabilité a montré que l'interpolation des surfaces était généralement stable et que la plus grande instabilité se produisait en mai, juin et juillet. Le manque d'observations dans la zone 4 était la raison principale pour la variabilité des estimations de l'apport, plus particulièrement durant les mois enregistrant un débit élevé.

La seconde approche a montré la manière dont les valeurs d'écoulement observées et prédites diffèrent sur l'étendue de la région d'étude, en utilisant un ensemble de sites de validation et en donnant une approximation de l'erreur dans l'interpolation. Elle a également indiqué que les principales différences entre l'écoulement observé et prédit se sont manifestées durant les mois de mai, juin et juillet. Lors de cette validation, on a découvert que les prévisions dans certains bassins étaient supérieures à l'écoulement observé, plus particulièrement dans la zone 1, où le débit

peut être « torrentiel ». Dans ces quelques cas, les valeurs d'écoulement utilisées dans l'interpolation des bassins avoisinants n'avaient pas toujours de régimes d'écoulement comparables.

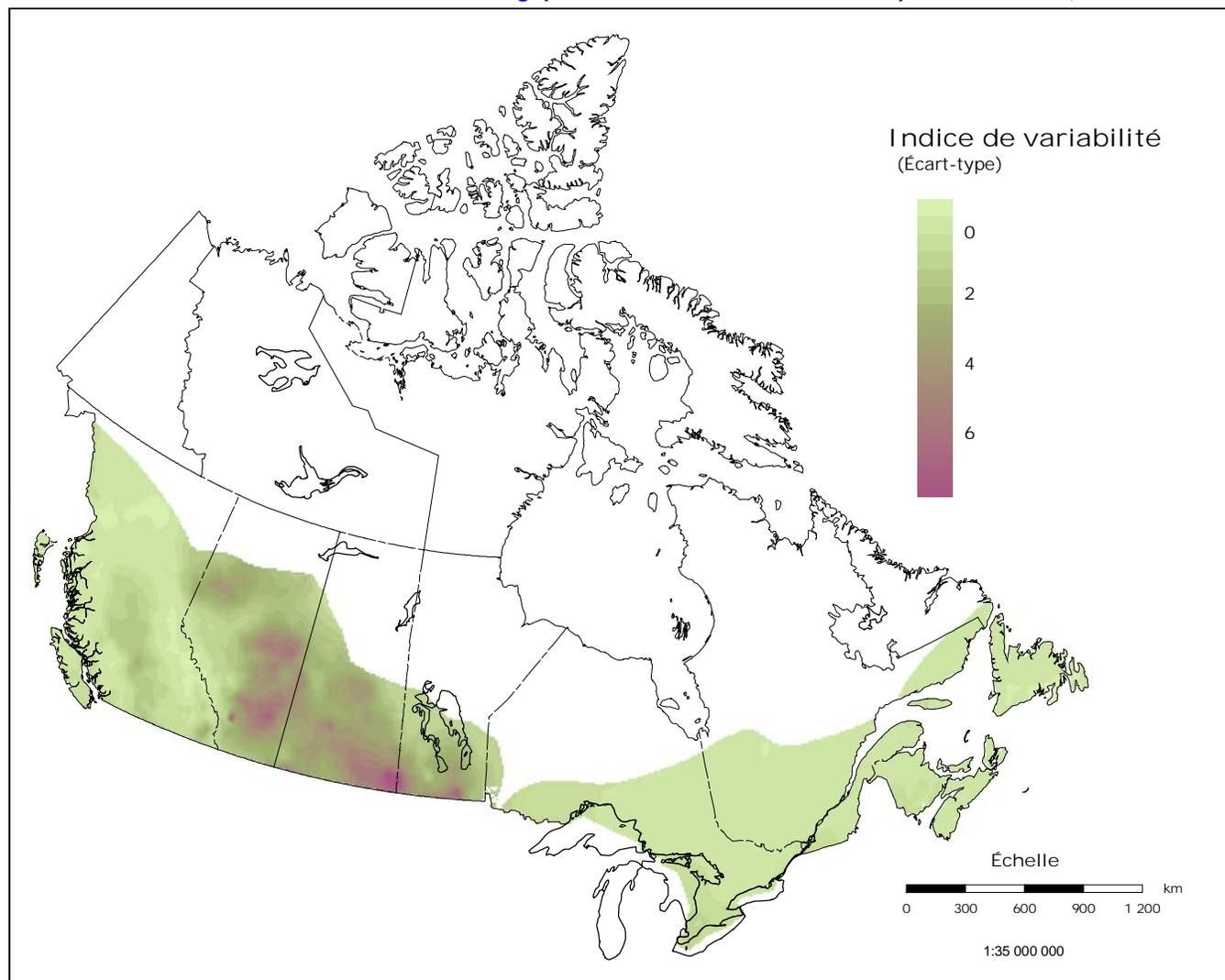
L'estimation de l'apport en eau produite par cette méthodologie est considéré comme une estimation raisonnable à l'échelle nationale et convient pour combler une lacune statistique importante dans les comptes des ressources en eau douce. Toutefois, en raison des défis liés à des estimations basés sur un réseau d'observations hydrométriques éparses (en particulier dans le Nord), plusieurs régions feront l'objet de recherches ultérieures. Par exemple, des recherches supplémentaires en matière de définition d'un réseau national de stations de jaugeage *représentatif sur le plan statistique* seront nécessaires afin d'améliorer davantage les estimations actuelles. Ceci aidera également à minimiser le biais potentiel qui est associé à l'utilisation d'une partie des bassins jaugés, comme les programmes de jaugeage ont des objectifs particuliers qui pourraient ne pas saisir complètement le régime d'écoulement principal pour une région donnée.

Des recherches plus approfondies seront également requises afin de déterminer la manière dont les prévisions relatives à l'apport en eau peuvent être utilisées pour des évaluations régionales ou des analyses des tendances, étant donné que plusieurs endroits contiennent des degrés d'incertitude élevés, en particulier dans la zone 4 et dans certaines régions de la zone 1. D'autres recherches sont nécessaires pour déterminer si l'erreur-type (variance) de la prévision peut être rétrotransformée à l'échelle initiale des données, ce qui permettrait une comparaison directe de l'erreur avec les prévisions de krigeage. L'erreur-type pourrait ensuite être utilisée pour établir des intervalles de confiance de la prévision qui pourraient alors être utilisés pour déterminer des différences significatives entre les périodes dans une analyse des séries chronologiques. Des recherches doivent également être effectuées pour déterminer si l'erreur-type pourrait établir une limite à laquelle la prévision est trop incertaine pour donner une estimation de l'apport en eau pour les études régionales.

Nonobstant ces préoccupations, un indice de variabilité a été calculé pour les zones 1, 2 et 3 afin de présenter un aperçu de l'endroit où se produisait la variabilité relative de l'écoulement la plus élevée au cours de la période, par rapport à la moyenne sur trente ans (carte 14). L'indice pour chaque cellule de la grille a été calculé pour les trente années comprises dans la période d'étude, en divisant l'écart-type par la moyenne. L'indice indique les régions dans lesquelles règne une grande variabilité, ce qui nous guidera dans nos recherches futures visant à déterminer si la variabilité est le résultat de fluctuations de l'écoulement d'une année à l'autre ou une tendance à augmenter ou à diminuer au fil du temps.

Carte 14

Indice de variabilité à travers les séries chronologiques s'échelonnant sur trente ans pour les zones 1, 2 et 3



Source(s) : Statistique Canada, Division des comptes et de la statistique de l'environnement.

Finalement, des mesures préliminaires ont été prises pour évaluer les volumes prédits et observés à l'échelle d'un bassin au moyen de comparaisons d'hydrogrammes. Les résultats tirés de cette analyse seront publiés lorsque l'évaluation des apports à une échelle régionale sera complétée. L'objectif central de ce document visait à présenter la méthodologie utilisée pour produire l'estimation de l'apport en eau annuel sur trente ans. Des travaux supplémentaires sont nécessaires pour expliquer la manière dont les résultats de l'apport en eau sont liés à l'hydrologie d'une région.

Références

- ABTEW, W., J. OBEYSEKERA et G. SHISH. « Spatial Analysis for Monthly Rainfall in South Florida », *Water Resources Bulletin, American Water Resources Association*, vol. 29, n° 2, 1993, p. 179-188.
- ARAP. Projet « Bassins-versants » de l'Administration du rétablissement agricole des prairies (ARAP), version 8, Agriculture et Agroalimentaire Canada, Regina (Saskatchewan), Canada, 2004. Consulté le 3 mai 2004. <<http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1226330737632&lang=fra>>
- ARMSTRONG, M. et A. BOUFASSA. « Comparing the Robustness of Ordinary Kriging and Lognormal Kriging : Outliers Resistance », *Mathematical Geology*, vol. 20, n° 4, 1987, p. 447-457.
- BELL, B. J. *Annual Unit Runoff on the Canadian Prairies*, Administration du rétablissement agricole des Prairies, Division de l'hydrologie, Hydrology Report n° 135, Regina (Saskatchewan), 1994.
- BORGA, M. et A. VIZZACCARO. « On the Interpolation of Hydrologic Variables : Formal Equivalence of Multiquadric Surface Fitting and Kriging », *Journal of Hydrology*, vol. 195, 1997, p. 160-171.
- BROOKS, R., K. D. HARVEY, D. W. KIRK, F. SOULARD, A. MURRAY et P. PAUL. *Building a Canadian Digital Drainage Area Framework*, 55^e conférence annuelle de l'Association canadienne des ressources en eau, Winnipeg (Manitoba), Canada, 14 juin 2002. Consulté le 3 septembre 2008. <http://www.cim.mcgill.ca/%7Eerbrook/downloads/Drainage_Area_Framework.pdf>
- CRESSIE, N. A. C. *Statistics for Spatial Data*, John Wiley and Sons Inc., New York (New York), États-Unis, 1993.
- DEUTSCH, C. V. et A. G. JOURNEL. *GSLIB : Geostatistical Software Library and User's Guide*. Oxford University Press, Oxford, Royaume-Uni, 1998.
- DINGMAN, S. L. *Physical Hydrology*. Prentice Hall, New Jersey, États-Unis, 1994.
- EATON, B., M. CHURCH et D. HAM. « Scaling and Regionalization of Flood Flows in British Columbia, Canada », *Hydrological Processes*, vol. 16, p. 3245-3263.
- ENVIRONNEMENT CANADA. Archives nationales des données hydrologiques, base de données hydrométriques (HYDAT), 2007. Consulté le 7 janvier 2009. <http://www.wsc.ec.gc.ca/products/hydat/main_f.cfm?cname=hydat_iso_f.cfm>
- GILBERT, R. « Statistical Methods for Environmental Pollution Monitoring », Van Nostrand Reinhold, New York, États-Unis, 1987.
- GODWIN, R. B. et F. R. J. MARTIN. « Calculation of Gross and Effective Drainage Areas for the Prairie Provinces », *Canadian Hydrology Symposium*, Winnipeg (Manitoba), Canada, 1975, p. 5.
- GOOVERTS, P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, États-Unis, 1997.
- HOULDING, S. W. *Practical Geostatistics : Modelling and Spatial Analysis*. Springer-Verlag, Berlin, Allemagne, 2000.
- ISAAKS, E. H. et R. M. SRIVASTAVA. *Applied Geostatistics*. Oxford University Press, New York, 1989.
- KRAKAUER, N. Y. et I. FUNG. « Mapping and Attribution of Change in Streamflow in the Coterminous United States », *Hydrology and Earth System Science*, vol. 12, 2008, p. 1111-1120.

KRAVCHENKO, A. K. et D. G. BULLOCK. « A Comparison Study of Interpolation Methods for Mapping Soil Properties », *Agronomy Journal*, vol. 91, 1999, p. 393-400.

KRUG, W. R., W. A. GEBERT, D. J. GRACZYK, D. L. STEVENS, B. P. ROCHELLE et M. R. CHURCH. « Map of Mean Annual Runoff for the North-Eastern, South-Eastern, and Mid-Atlantic United States 1951-80 », *U.S. Geological Survey, Water-Resources Investigations Report*, n° 88-4094, 1990.

LAYCOCK, A. H. « The Amount of Canadian Water and Its Distribution », *Canadian Aquatic Resources*, édité by M. C. Healey et R. R. Wallace, Pêches et Océans Canada, 1987.

OLEA, R. A. *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers, Boston, États-Unis, 1999.

RESSOURCES NATURELLES CANADA. *National Scale Frameworks Hydrology – Hydrometric Gauging Stations and Station Basins*, Canada, 2003. Consulté le 1^{er} septembre 2006. <<http://www.geogratis.ca/geogratis/fr/collection/detail.do?id=27748>>

SANDERSON, M. E. et D. W. PHILLIPS. « Average Annual Water Surplus in Canada », *Climatological Studies* n° 9, Meteorological Branch, ministère des Transports, Toronto, Canada, 1967.

SIMARD, Y., P. LEGENDRE, G. LAVOI et D. MARCOTTE. « Mapping, Estimating Biomass and Optimizing Sampling Programs for Spatially Autocorrelated Data : Case Study of the Northern Shrimp *Pandalus borealis* », *Canadian Journal of fisheries and Aquatic Sciences*, vol. 49, n° 1, 1992, p. 32-45.

SOULARD, F. « Water Accounting at Statistics Canada : The Inland Fresh Water Assets Account », *London Group of Environmental Accounting*, Rome, novembre 2003. Consulté le 27 février 2009. <http://unstats.un.org/unsd/envaccounting/londongroup/meeting8/water_accounting_canada.pdf>

SPENCE, C. et A. BURKE. « Estimates of Canadian Arctic Archipelago Runoff from Observed Hydrometric Data », *Journal of Hydrology*, vol. 362, 2008, p. 247-259.

STATISTIQUE CANADA. *Concepts, sources et méthodes du Système des comptes de l'environnement et des ressources du Canada*, catalogue n° 16-505–GPF, Ottawa (Ontario), Canada, 1997.

ST-HILAIRE, A., T. B. M. J. OUARDA, M. LACHANCE, B. BOBÉE, J. GAUDET et C. GIGNAC. « Assessment of the Impact of Meteorological Network Density on the Estimation of Basin Precipitation and Runoff : a Case Study », *Hydrological Processes*, vol. 17, 2003, p. 3561-3580.

SYED, K. H., D. C. GOODRICH, D. E. MYERS et S. SOROOSHIAN. « Spatial Characteristics of Thunderstorm Rainfall Fields and Their Relation to Runoff », *Journal of Hydrology*, vol. 271, 2003, p. 1-21.

TABIOS, G. Q. et J. D. SALAS. « A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation », *Water Resources Bulletin, American Water Resources Association*, vol. 21, n° 3, 1985, p. 365-380.

WEBSTER, R. et M. A. OLIVER. *Geostatistics for Environmental Scientists*, John Wiley and Sons, Ltd., London, 2000.

WU, J., W. A. NORWELL et R. M. WELCH. « Kriging on Highly Skewed Data for DTPA-Extractable Soil Zn with Auxiliary Information for PH and Organic Carbon », *Geoderma*, vol. 134, 2006, p. 187-199.

YAMAMOTO, J. K. « On Unbiased Backtransform of Lognormal Kriging Estimates », *Computers in Geoscience*, vol. 11, 2007, p. 219-234.

YASREBI, J., M. SAFFARI, H. FATHI, N. KARIMIAN, M. MOAZALLAHI et R. GAZNI. « Evaluation and Comparison of Ordinary Kriging and Inverse Distance Weighting Methods for Prediction of Spatial Variability of Soil Chemical Parameters », *Research Journal of Biological Sciences*, vol. 4, n° 1, 2009, p. 93-102.