Louis Carrier

carl2621

[louis.carrier@usherbrooke.ca](mailto:louis.carrier@usherbrooke.ca)

Vincent Le Falher

levf2603

[vincent.lefalher@usherbrooke.ca](mailto:vincent.lefalher@usherbrooke.ca)

Cours GMQ706 Travaux pratiques d’analyse spatiale

Automne 2016

Travail pratique # 1 Analyse et interpolation spatiale de points remis à Mickaël Germain

Contenu

[1. Introduction 4](#_Toc469574935)

[2. Questions 4](#_Toc469574936)

[2.1. Question 1 : Échantillonnage de données ponctuelles 4](#_Toc469574937)

[2.2. Question 2 : Interpolations de type déterministe 4](#_Toc469574938)

[2.3. Question 3 : Analyses géostatistiques avec le krigeage 5](#_Toc469574939)

[3. Conclusion 8](#_Toc469574940)

[4. Annexes 9](#_Toc469574941)

[4.1. Préparation des données 9](#_Toc469574942)

[4.2. Système de coordonnées 9](#_Toc469574943)

[4.3. Symbologie de l’élévation des stations météorologiques du québec 9](#_Toc469574944)

[4.4. Symbologie des températures moyennes des stations météorologique du québec 11](#_Toc469574945)

[4.5. Symbologie des précipitation totales des stations météorologiques du Québec 12](#_Toc469574946)

[4.6. Bassins hydrographiques du Québec 13](#_Toc469574947)

[4.7. Zones Géologiques de l’amérique du nord 13](#_Toc469574948)

[4.8. Comparaison des méthodes déterministes pour l’interpolation de l’élévation 14](#_Toc469574949)

[4.9. Question 3 – Objectif 1 : krigeage Tm et Co-krigeage Elevation 16](#_Toc469574950)

[4.10. Question 3 – Objectif 2 – Pré-Analyse 20](#_Toc469574951)

[4.10.1. Températures Moyennes 20](#_Toc469574952)

[4.10.2. Précipitations Totales 23](#_Toc469574953)

[4.10.3. Synthèse de la pré-analyse 24](#_Toc469574954)

[4.11. Question 3 – Objectif 2-3 : Analyse krigeage Températures Moyennes 25](#_Toc469574955)

[4.12. Question 3 – Objectif 2-3 : Analyse krigeage Précipitations 29](#_Toc469574956)

[5. Références bibliographiques 33](#_Toc469574957)

**Figures**

[Figure 1 : Système de coordonnées 9](#_Toc469574958)

[Figure 2 : Symbologie des élévations 10](#_Toc469574959)

[Figure 3 : Symbologie des températures moyennes 11](#_Toc469574960)

[Figure 4 : Symbologie des précipitations totales 12](#_Toc469574961)

[Figure 5 : Vue générale des bassins hydrographiques (UQAM, 1999) 13](#_Toc469574962)

[Figure 6 : Vue détaillée bassins hydrographiques (MDDELCC, 2003) 13](#_Toc469574963)

[Figure 7 : Zones Géologiques vue en plan (Pierre-André Bourque, 2010) 13](#_Toc469574964)

[Figure 8 : Zones Géologiques vue en profil (Pierre-André Bourque, 2010) 13](#_Toc469574965)

[Figure 9 : Question 3 – Objectif 1 – Comparaison des étapes 1-2 – Températures moyennes 17](#_Toc469574966)

[Figure 10 : Question 3 – Objectif 1 – Comparaison des étapes 3-5 – Températures moyennes 18](#_Toc469574967)

[Figure 11 : Question 3 – Objectif 1 – Comparaison des indicateurs d'erreurs – Températures moyennes 19](#_Toc469574968)

[Figure 12 : Question 3 – Objectif 1 – Comparaison des surfaces des modèles d’interpolation – Températures moyennes 19](#_Toc469574969)

[Figure 13: Question 3 – Objectif 2 – Pré-analyse de la distribution pour les températures moyennes 21](#_Toc469574970)

[Figure 14 : Question 3 – Objectif 2 – Pré-analyse de la stationnarité pour les températures moyennes 22](#_Toc469574971)

[Figure 15 : Question 3 – Objectif 2 – Pré-analyse des tendances pour les températures moyennes 23](#_Toc469574972)

[Figure 16 : Question 3 – Objectif 2-3 – Étapes similaires du krigeage – Températures moyennes 26](#_Toc469574973)

[Figure 17 : Question 3 – Objectif 2-3 – semivariogramme et validation croisée – Températures moyennes 27](#_Toc469574974)

[Figure 18 : Question 3 – Objectif 2-3 – Indicateurs d'erreurs – Températures moyennes 28](#_Toc469574975)

[Figure 19 : Question 3 – Objectif 2-3 – Surfaces des modèles d’interpolation – Températures moyennes 28](#_Toc469574976)

[Figure 20 : Question 3 – Objectif 2-3 – Étapes 2-3 – Précipitations totales 30](#_Toc469574977)

[Figure 21 : Question 3 – Objectif 2-3 – Semivariogramme et validation croisée – Précipitations totales 31](#_Toc469574978)

[Figure 22 : Question 3 – Objectif 2-3 – Indicateurs d'erreurs – Précipitation totales 32](#_Toc469574979)

[Figure 23 : Question 3 – Objectif 2-3 – Surfaces des modèles d’interpolation – Précipitations totales 32](#_Toc469574980)

**Tableaux**

[Tableau 1 : Question 2 – Comparaison des modèles déterministes 15](#_Toc469572508)

[Tableau 2 : Question 2 – Comparaison des surfaces des modèles déterministes 17](#_Toc469572509)

[Tableau 3 : Question 3 – Objectif 2 – Pré-analyse pour les précipitations totales 26](#_Toc469572510)

[Tableau 4 : Question 3 – Objectif 2 – Synthèse de la pré-analyse – Températures moyennes et précipitations totales 27](#_Toc469572511)

# Introduction

Le présent TP consiste à répondre à trois questions qui nous permettrons de couvrir les objectifs souhaités, c'est-à-dire 1) apprendre à construire un semis de points en vue de réaliser une analyse spatiale, 2) connaitre l’intérêt spécifique des principales méthodes d’interpolation, et 3) savoir utiliser le krigeage et les fonctions d’estimation géostatistique des erreurs d’interpolation pour optimiser l’interpolation.

# Questions

# Question 1 : Échantillonnage de données ponctuelles

Il est important de noter que toutes les stations à notre disposition ont été volontairement conservées, même si la création d’un semis de points avec un minimum de 40 stations était le premier objectif de la question 1. En effet, nous avons considéré répondre le mieux possible à l’une des hypothèses principales de l’interpolation qui stipule d’utiliser un échantillon de données représentant la variabilité spatiale la plus variée afin de capter les patrons spatiaux (R.Fournier, 2005). De plus, nos disponibilités restreintes et la planification de nos efforts afin de répondre de manière optimale au TP, vient renforcer cette décision.

L’étape de préparations des données est documentée en annexe.

Afin de répondre au second objectif de la question 1, les températures moyennes et la précipitation totale pour le mois de Décembre 2013 pour la province de Québec ont été regroupés avec l’élévation de chaque station en utilisant l’outil « Joins » d’ArcMap. L’identifiant de la station, de type ‘text’, a été utilisé comme élément de jointure entre les deux jeux de données. Le résultat de la jointure donne une sélection de 203 stations. Les polygones de la province de Québec ont été extraits des polygones du Canada. Les semis de points pour chacune des stations ont ensuite été représentés sur les polygones de la province de Québec grâce à l’option Display XY Data de ArcMap.

Nous pouvons déjà indiquer que notre échantillon ne permet pas une bonne représentation spatiale de la province de Québec. Il est clair qu’il n’y a que la partie sud de la province, autour du fleuve St-Laurent, qui est bien représentée. La qualité des résultats d’interpolation en seront impacté en conséquence.

Les polygones de la province du Québec, le semis de points des stations, et les tableaux des stations avec l’élévation, la température moyenne et les précipitations totales pour Décembre 2013 pour la province du Québec ont été regroupés dans une geodatabase dont le système de coordonnées est le GCS North American 1983, hérité des polygones des provinces du Canada.

# Question 2 : Interpolations de type déterministe

L’objectif de la question 2 est de construire un modèle numérique d’altitude (MNA) à partir des données d’élévation de notre jeu de données.

Avant de commencer, clarifions les notations utilisées par le modèle d’interpolation déterministe.

La variable régionalisée *z* correspond à l’élévation pour cette question. Le domaine *D* d’étude est la province du Québec. Les sites *s* sont représentés par les stations météorologiques de la province du Québec, qui sont au nombre de 203 dans notre jeu de données. L’élévation à une station météorologique correspond à la variable régionalisée *z* à un site *s* et noté *z(s)*. La valeur régionalisée est l’élévation en mètre de la station au site .

Nous savons que le Québec est composé de plusieurs bassins hydrographiques. On peut penser que la délimitation de ces bassins passe par des brisures de pentes lesquelles sont marquées par une variation des élévations. À notre avis, ces marques forment des zones locales. Des images des principaux bassins hydrographiques du Québec sont fournies en annexe. Elles nous servirons à valider l’interpolation. Grace à une étude des zones géologiques, disponible également en annexe, on s’aperçoit que le semis de points est malheureusement assez dispersé pour la partie des Appalaches, où la fluctuation des élévations est plus importante. La vallée du St-Laurent, si on exclue les montérégiennes, offre peu de fluctuation et constitue une plaine uniforme. C’est là que le semis de points est le plus dense. On peut s’attendre à une meilleure interpolation pour cette région. Pour ce qui est du bouclier canadien, le semis de points est très faible pour l’ensemble de cette grande région. Heureusement, les élévations varient plus lentement pour la partie nord-ouest de la région.

Les élévations pour les stations de la province du Québec varient entre le niveau de la mer, le long du fleuve, et 791 mètres à la station de « L’étape », dans la région de la Capitale-Nationale.

Malgré la mauvaise représentation spatiale de notre échantillon, nous avons déterminé au départ que la méthode d’interpolation « *Radial Basic Function* » (RBF) était selon nous la méthode fournissant l’interpolation la plus proche de nos attentes. Cette méthode est celle qui est recommandée par la littérature (ESRI, 2014) pour produire des surfaces qui varient légèrement, ce qui est le cas pour l’élévation au Québec. Cette méthode est déterministe car elle ne modélise pas la variable régionalisée par une fonction aléatoire (S.Baillargeon, 2015), telle que désiré par la question. Étant une méthode exacte, cela nous certifie que les surfaces d’interpolation passent par les stations météorologiques et produisent une surface la plus proche possible de la réalité. Elle est locale, c'est-à-dire que les propriétés de la fonction RBF lui font exécuter sa fonction en tenant compte du point le plus proche de la variable régionalisée à prédire, dans plusieurs directions (R.Fournier, 2015).

Par rapport à l’autre méthode déterministe locale et exacte qu’est le IDW, elle présente les avantages de pouvoir donner des valeurs régionalisées pouvant dépasser l’élévation d’une station, d’avoir plus d’options et donc de flexibilité et de contrôle sur la fonction, de donner des points interpolés moins erronés, et des indicateurs d’erreurs pour évaluer les prédictions.

La méthode polynomiale globale donne le pire des résultats. Les échantillons sont tellement dispersés, d’une densité très variable et inégale, l’aplanissement ou le lissage déforme complètement le modèle interpolé.

Même si la valeur de l’erreur *«* *Root Mean Square* » (RMS) de l’interpolation par la méthode polynomiale locale semble moins fiable que le RBF avec la fonction « *Spline Tension* », la comparaison qualitative offre de meilleur résultat visuel. En effet, lorsque nous comparons le modèle interpolé avec le modèle 250k (RNCAN, 2016), on reconnait la configuration des principales variations des élévations.

La comparaison et la validation croisée des méthodes déterministes sont disponibles en annexe.

# Question 3 : Analyses géostatistiques avec le krigeage

Le premier objectif de la question 3 est de faire une analyse de type krigeage d’une de nos variables climatiques avec les paramètres par défaut et de vérifier par cokrigeage la contribution de l’élévation aux variations.

La variable régionalisée que nous avons choisie est la température moyenne. L’exécution du krigeage et du cokrigeage s’effectue avec les paramètres par défaut, sans pré-analyse ni correction de tendances potentielles, et sans optimisation logicielle. Nous nommerons le modèle du krigeage pour les températures moyennes le modèle « krigeageTmDéfautNonOptimisé », et celui du cokrigeage qui tient compte de l’altitude en plus de la température moyenne « cokrigeageTmElevDéfautNonOptimisé ».

Plusieurs indices de la validation croisée nous intéressent. Le premier est l’erreur « *Root Mean Square* » (RMS), qui vaut 2.565. Une valeur la plus basse possible est préférable (ESRI, 2016b) et nous la comparerons avec l’erreur RMS du modèle « cokrigeageTmElevDéfautNonOptimisé ». Cet indice doit aussi être le plus proche possible de l’indice « *Average Standard Error* » (ASE) car c’est uniquement l’erreur standard qui permet d’estimer l’incertitude des valeurs prédites (ESRI, 2016b). L’autre indice d’erreur qui doit être comparé est le « *Root Mean Square Standardized* » (RMSS). Celui-ci doit être le plus proche possible de la valeur un (1) (ESRI, 2016b). Dans ce cas ci il vaut 0.853 ce qui signifie que la variabilité est surestimée (ESRI, 2016b). L’outil donne aussi des graphiques permettant de visualiser le résultat de l’interpolation prédite, les erreurs, et la normalité de la distribution prédite (graphique *QQPlot*). Ils pourront être comparés avec ceux du modèle « cokrigeageTmElevDéfautNonOptimisé ».

Lorsque l’interpolation de type krigeage implique l’élévation, les indices d’erreur ont évolués. Le RMS, qui vaut 2.875, et la différence entre le RMS et le ASE, sont plus élevés que le modèle « krigeageTmDéfautNonOptimisé ». Le RMSS dans ce cas ci vaut 1.093, ce qui signifie que la variabilité est sous-estimée (ESRI, 2016b). Afin de déterminer le modèle d’interpolation le plus adéquat, la littérature (ESRI, 2016c) indique que le meilleur modèle est celui qui a l’erreur « Mean Standardized » proche de zéro (0), le RMS le plus bas, le ASE le plus proche du RMS, et le RMSS le plus proche de un (1). Si nous suivons cette recommandation aveuglément, le modèle « krigeageTmDéfautNonOptimisé » est le plus optimal à représenter les températures moyennes sur l’ensemble du territoire du Québec. Cela se confirme en visualisant les surfaces d’interpolation qui sont générées, où l’on peut observer que la partie basse du Québec, où les échantillons sont le plus nombreux, semble bien respecter les différences de températures, alors que la partie nord semble moins certaine, moins uniforme. Cette observation se vérifie avec la surface d’interpolation représentant les erreurs standards, où la partie nord possède le plus de probabilité d’erreur, et le sud semble être beaucoup plus « chanceux » en matière de prédiction. De plus le modèle « cokrigeageTmElevDéfautNonOptimisé » représente bien les échelons de températures selon l’élévation, mais c’est aussi à ce niveau que le risque d’erreur est le plus élevé. La partie basse du Québec est aussi moins uniformément représenté. Ce modèle est donc le moins recommandé.

Les détails de l’analyse, les paramètres de l’assistant, les surfaces des modèles et la comparaison des indices d’erreurs sont disponibles en annexe.

Le second objectif de la question 3 est d’effectuer un krigeage pour toutes les variables climatiques à partir d’une exploration détaillée des données et créer les meilleures prédictions possibles en effectuant des validations croisées « Cross Validation ».

La première variable régionalisée que nous avons choisi est la température moyenne. Avant de produire une interpolation par krigeage, nous validons les trois prérequis : la distribution doit être 1) normale; 2) stationnaire; 3) sans tendance systémique globale. Les détails de la pré-analyse se retrouvent en annexe.

A la suite de la pré-analyse, nous pouvons conclure que la méthode du krigeage peut être exécutée bien que la stationnarité n’est pas parfaite.

La validation croisée permet de valider la justesse de la fonction d’interpolation du semivariogramme de ce modèle. L’erreur « *Root Mean Square* » (RMS) est faible et vaut 1.567. Cet indice est très proche de l’indice « *Average Standard Error* » (ASE), qui vaut 1.519, la différence étant de 0.048, et ce qui est bon. Le « *Root Mean Square Standardized* » (RMSS) vaut 1.02 ce qui signifie que l’on peut considérer que la variabilité est très bien estimée (ESRI, 2016b). La ligne du graphique de l’interpolation est proche de la médiane, et celui du QQPlot confirme la normalité de la distribution prédite.

Nous pouvons affirmer que ce modèle est très encourageant car il répond aux critères de sélections d’un bon modèle (ESRI, 2016b). Il est bien meilleur que celui réalisé avec les paramètres par défaut du logiciel. En analysant un peu plus les possibles améliorations, nous pensons incorporer dans le modèle la prise en compte de l’anisotropie dans la fonction du semivariogramme. Le RMSS devient précisément égale à un (1) jusqu'à la 3ème décimale après la virgule, et la différence entre le RMS et le ASE s’est amoindrie pour n’avoir qu’une différence de 0.0098. L’interpolation est devenue encore meilleure.

Le dernier objectif de cette question était de comparer le meilleur modèle avec un modèle généré avec la méthode du krigeage empirique bayésienne « *Empirical Bayesian Kriging* » (EBK). Plusieurs modèles EBK ont été exécutés avec différents paramètres, et avec surprise aucun n’a pu battre le modèle tenant compte de l’anisotropie. Nous pensions que la non-stationnarité observée dans notre pré-analyse, surtout pour le nord du Québec, influencerait négativement la méthode du krigeage et qu’une méthode EBK serait donc plus appropriée. Il n’en ait rien.

Nous pouvons considérer avec confiance que ce modèle est le meilleur et le plus adéquat pour représenter les températures du mois de Décembre 2013 dans l’ensemble de la province du Québec, sans trop d’écarts de prédiction dans les zones les moins bien échantillonnées. Les valeurs des indicateurs d’erreurs tels que la littérature le recommande (ESRI, 2016b) nous démontrent que la méthode utilisée donne une interpolation juste; les paramètres personnalisés suivent une logique d’analyse et ont fait la différence, telle la suppression d’une tendance de type polynomiale de 1er ordre et la prise en compte de la direction (anisotropie).

La seconde variable régionalisée pour répondre au second objectif de la question 3 est la précipitation totale. De même que pour les températures moyennes, avant de produire une interpolation par krigeage, nous validons les trois prérequis. Les détails sont également fournis en annexe.

La validation croisée permet de valider la justesse de la fonction d’interpolation du semivariogramme de ce modèle. L’erreur « *Root Mean Square* » (RMS) est assez élevé et vaut 30.843. Cet indice est assez proche de l’indice « *Average Standard Error* » (ASE), qui vaut 27.164, la différence étant de 3.679. Le « *Root Mean Square Standardized* » (RMSS) vaut 1.12 ce qui signifie que l’on peut considérer que la variabilité est sous estimée (ESRI, 2016b). La ligne de prédiction du graphique est loin d’être alignée avec celle de la médiane, celle du QQPlot confirme la normalité de la distribution prédite.

Nous pouvons affirmer que ce modèle est encourageant mais pas parfait. Il est néanmoins meilleur que le modèle généré par défaut sans correction de la tendance mais avec une optimisation pas défaut. En analysant un peu plus les possibles améliorations, nous pensons incorporer dans le modèle la prise en compte de l’anisotropie dans la fonction du semivariogramme. Pour un angle de 75.4 degrés et une portée minime (*minor range*) de 0.97, déterminé manuellement avec la carte du semivariogramme, le RMSS devient meilleur, à 1.110, et la différence entre le RMS et le ASE s’est amoindrie pour n’avoir qu’une différence de 3.165. L’interpolation est devenue meilleure, il y a moins de probabilité d’erreur pour les valeurs prédites.

Le dernier objectif de cette question est de comparer le meilleur modèle avec un modèle généré avec la méthode du krigeage empirique bayésienne « *Empirical Bayesian Kriging* » (EBK). Plusieurs modèles EBK ont été explorés, avec différents paramètres, et le modèle EBK avec les paramètres par défaut donne le meilleur résultat et est meilleur que le modèle de kriegage. La non-stationnarité observée dans notre pré-analyse influence donc négativement la méthode du krigeage et la méthode EBK est donc plus appropriée pour cette variable régionalisée dans notre contexte.

Nous pouvons donc considérer avec confiance que le modèle EBK est le meilleur et le plus adéquat pour représenter les précipitations totales du mois de Décembre 2013 dans l’ensemble de la province du Québec, sans qu’il n’y ait trop d’écarts de prédiction dans les zones les moins bien échantillonnées. Les valeurs des indicateurs d’erreurs tels que la littérature le recommande (ESRI, 2016b) nous démontrent que la méthode utilisée donne une interpolation juste en tenant compte de la non-stationnarité.

# Conclusion

La flexibilité de l’analyse spatiale devient rapidement chronophage. Elle met l’analyste moins expérimenté face à des possibilités et des prises de décision diverses et souvent subjectives. Il ne doit pas hésiter à faire valider ses méthodes et ses résultats auprès de ses pairs. Car il ne faut pas oublier de mentionner que les choix devront pouvoir être défendus devant une audience de décisionnaires. Il est recommandé de réaliser une bonne pré-analyse, de connaitre le jeu de données empiriques à disposition, son histoire, la méthode d’échantillonnage utilisée, et d’avoir une bonne connaissance des différentes méthodes d’interpolation, de leurs paramètres d’entrés, de leurs influences, et des résultats attendus et souhaités selon les objectifs et les requis du projet. La méthode de krigeage est très intéressante et puissante, et les méthodes de cokrigeage et EBK viennent renforcer l’analyse spatiale.

L’assistance du logiciel est non négligeable mais les résultats obtenus avec les paramètres par défaut doivent être remis en question et comparé avec d’autres méthodes. La flexibilité du logiciel est une force mais aussi une faiblesse : les erreurs de manipulation et d’interprétation sont très faciles.

Il existe encore des possibilités d’optimisation que nous avons rencontrée lors de notre analyse. Il existe deux stations très à l’écart de la normalité. Une étude plus approfondie pourrait aider à expliquer ce phénomène et appliquer une stratégie d’intégration de ces stations dans les méthodes d’interpolation. L’étude de l’anisotropie pourrait aussi être approfondie, et une recherche du voisinage plus fine à considérer.

# Annexes

# Préparation des données

Les données nécessaires afin de répondre aux problématiques nous ont été partiellement fournies.

Les polygones du Canada, au format shapefile, détermine les contours des provinces. Les polygones de la province du Québec seront extraits et conservés dans un nouveau shapefile.

Les semis de points représentant la liste des principales stations météorologiques du Canada, au format shapefile, contiennent l’élévation, la latitude et la longitude de la station. Un jeu de données plus récent sera utilisé à la place de celui fourni.

Afin de compléter le jeu de données, nous avons récupéré un fichier Excel provenant du site des données météorologiques du Gouvernement du Canada contenant les températures moyennes, en degré Celsius, et les précipitations totales, en millimètre, pour Décembre 2013 pour la province du Québec (Environnement Canada, 2016). Les données seront filtrées dans Excel afin de ne conserver que celles dont la température moyenne et la précipitation totale sont disponibles. Une fois les données préparées, elles ont été importées dans une table dans ArcMap et joint avec la table des stations et leur élévation.

Un autre fichier CSV provenant du même site contient la liste la plus à jour des stations météorologiques du Canada, avec leur élévation en mètre, leur latitude et longitude en degrés décimaux. Ce fichier sera celui utilisé pour la jointure avec les températures moyennes et les précipitations totales pour Décembre 2013. Les données seront filtrées à partir d’Excel afin de ne conserver que celles associées à la province du Québec.

A noter que le ID de la station et utilisé pour la jointure contient des caractères alphanumériques et doit donc être de type ‘text’.

# Système de coordonnées

Le système de coordonnées des polygones du jeu de données est NAS 1983.

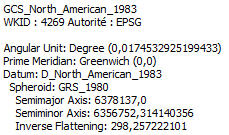


Figure 1 : Système de coordonnées

# Symbologie de l’élévation des stations météorologiques du québec

Voici la symbologie représentant l’élévation de chacune des 203 stations météorologique du Québec. Nous pouvons observer à l’œil nu une tendance de faible élévation qui sépare le sud et le nord du Québec, et qui correspond à l’étude géologique du Québec.

Nous pouvons aussi observer le faible semis de points disponible par rapport à la superficie du Québec, et déterminer ainsi les zones d’interpolation qui ne seront pas optimales.

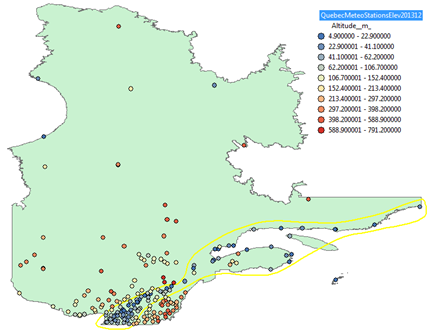


Figure 2 : Symbologie des élévations

# Symbologie des températures moyennes des stations météorologique du québec

En utilisant les données de températures moyennes pour le mois de Décembre 2013 pour les 203 stations du Québec à notre disposition nous pouvons observer quatre zones virtuellement bien définies.

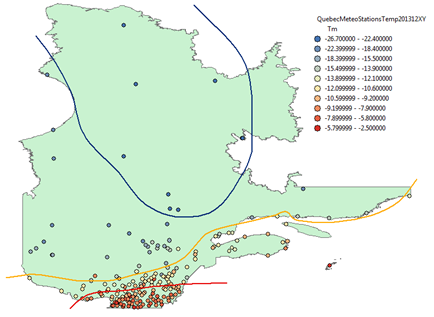


Figure 3 : Symbologie des températures moyennes

# Symbologie des précipitation totales des stations météorologiques du Québec

En utilisant les données de précipitation totales pour le mois de décembre 2013 pour les 203 stations du Québec à notre disposition, il est assez difficile d’observer un découpage ou une tendance bien définie. Les précipitations sont assez bien arrangées dans le sud du Québec. Subjectivement, plus on monte vers le nord, moins il semble y avoir de précipitation.

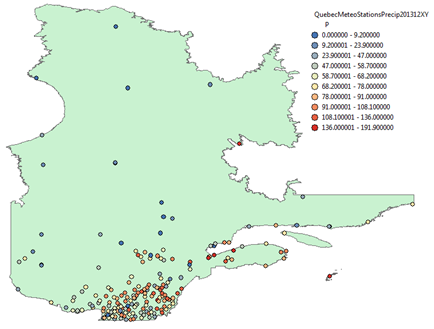


Figure 4 : Symbologie des précipitations totales

# Bassins hydrographiques du Québec

Voici des images des principaux bassins hydrographiques du Québec.

|  |  |
| --- | --- |
| Figure 5 : Vue générale des bassins hydrographiques (UQAM, 1999) | Figure 6 : Vue détaillée bassins hydrographiques (MDDELCC, 2003) |

# Zones Géologiques de l’amérique du nord

|  |  |
| --- | --- |
| Figure 7 : Zones Géologiques vue en plan (Pierre-André Bourque, 2010) | Figure 8 : Zones Géologiques vue en profil (Pierre-André Bourque, 2010) |

# Comparaison des méthodes déterministes pour l’interpolation de l’élévation

Voici la comparaison des méthodes d’interpolation déterministes.

|  |  |  |  |
| --- | --- | --- | --- |
| Méthode et paramètre utilisé | Étape 2 de 3 | Étape 3 de 3 | Valeurs extrêmes Cross-Validation et RMS |
| IDW avec le *Neighborhood type = Standard* a donnée le meilleur *Root Mean Square* |  |  | Erreur varie de -571.67 à 403.21  Écart de 974.88  RMS = 105,9935 |
| Global Polynomial- Order of polynomial = 4 a donné le meilleur *Root Mean Square* |  |  | Erreur varie de -632.41 à 736.42  Écart de 1368.83  RMS = 141,6618 |
| RBF – avec *Kernel function = Spline* with tension donne le meilleur *Root Mean Square* |  |  | Erreur varie de  -544.49 à 396.56  Écart de 941.05  RMS = 98,4619 |
| Local polynomial avec *kernel function = constant*. A donné le meilleur *Root Mean Square* |  |  | Erreur varie de -629.52 à 431.33  Écart de 1060.85  RMS = 115,5081 |

Tableau 1 : Question 2 – Comparaison des modèles déterministes

Afin de valider qualitativement les résultats, nous comparons visuellement les modèles interpolés.

|  |  |
| --- | --- |
| Modèle | Résultat en image |
| IDW |  |
| Global Polynomial |  |
| RBF  avec la fonction « *Spline Tension* » |  |
| Local polynomial |  |
| MNA 250k (RNCAN, 2016) |  |

Tableau 2 : Question 2 – Comparaison des surfaces des modèles déterministes

# Question 3 – Objectif 1 : krigeage Tm et Co-krigeage Elevation

Ce paragraphe détaille la comparaison entre le modèle d’interpolation de type krigeage avec les paramètres par défaut pour les températures moyennes, et celui du cokrigeage qui tient compte de l’élévation en plus des températures moyennes.

Nous commençons à analyser le modèle d’interpolation de type krigeage avec les paramètres par défaut pour les températures moyennes.

Nous observons que, malgré la transformation « *Normal Score* » de l’assistant, la distribution n’est pas asymétrique, il y a un étalement vers la gauche. Le QQPlot nous indique qu’il n’y a pas d’anormalités dans la distribution, mis à part pour quelques stations. L’étape de la modélisation du semivariogramme nous permet de quantifier les autocorrélations de nos données empiriques et déterminer si le modèle de la fonction « *Stable* » sélectionné par défaut est adéquat. La courbe représentant la fonction du modèle passe proche des valeurs moyennes de chaque décalage, jusqu’au 12ème, ce qui est correcte. Elle passe au milieu du nuage des points représentant les groupements des pairs selon leurs distances (ESRI, 2016c), qui devient normalement clairsemé autour d’elle plus les distances entre les stations augmentent. Le nombre de décalage (*lags*) du modèle est de 12, chacun ayant un pas (*lap*) de 0.299 degrés. La distance totale, qui est la différence de température la plus large entre tous les points, est de 3.6 degrés, et correspond au nombre de décalage multiplié par le pas « 12 x 0.299 » (ESRI, 2016c). La pépite (*nugget*) a une valeur de 0.286 et coïncide presque avec la valeur du pas. Par manque de temps et de connaissance, nous ne pouvons pas déterminer si la pépite est due à une erreur de mesure ou à un phénomène naturel (ESRI, 2016b). La portée (*range*) est difficilement identifiable, la courbe monte très progressivement et ne semble jamais atteindre le seuil (*sill*), tel que nous le confirme le logiciel qui indique qu’elle mesure 3.599 degrés, c'est-à-dire la distance totale du décalage. Cela nous indique que toutes les stations de l’échantillonnage auront une influence sur leurs voisins, même la plus éloignée (ESRI, 2016b). La forme de la courbe de la fonction « *Stable* » ressemble à celle de la fonction gaussienne. L’assistant du logiciel ne sélectionne pas d’anisotropie par défaut, ce qui donnera une fonction non réaliste pour les températures qui peuvent varier selon la distance mais aussi selon un angle polaire dans plusieurs directions, surtout selon la latitude (sud-nord) dans le cas de la province du Québec, mais aussi selon l’élévation (altitude), qui est tout de même moins prononcée dans cette région du monde. La recherche des voisins les plus proches est de type « *Standard* » par défaut. Elle est influencée par la fonction du semivariogramme et applique un poids plus ou moins important aux points voisins selon leur distance. Le nombre de points maximums à rechercher dans chacun des quatre secteurs de 45 degrés est de cinq (5), et le nombre minimum est de deux (2).

La dernière étape de l’assistant nous offre de valider les erreurs d’interpolation pour chacun de nos échantillons avec une validation croisée.

Lorsque l’interpolation de type krigeage implique l’élévation, plusieurs paramètres ont évolués. Le pas étant plus petit (0.229 degrés), la différence maximale de température prise en compte par la fonction « Stable » est dorénavant de 2.71 degrés. La portée de la courbe est de 1.78 degrés, ce qui implique une diminution de l’influence des stations dont la température est située au delà de cette mesure (ESRI. 2016b).

|  |  |
| --- | --- |
| krigeage pour les températures moyennes avec les paramètres par défauts | Cokrigeage pour les températures moyennes et l’altitude, avec les paramètres par défauts |

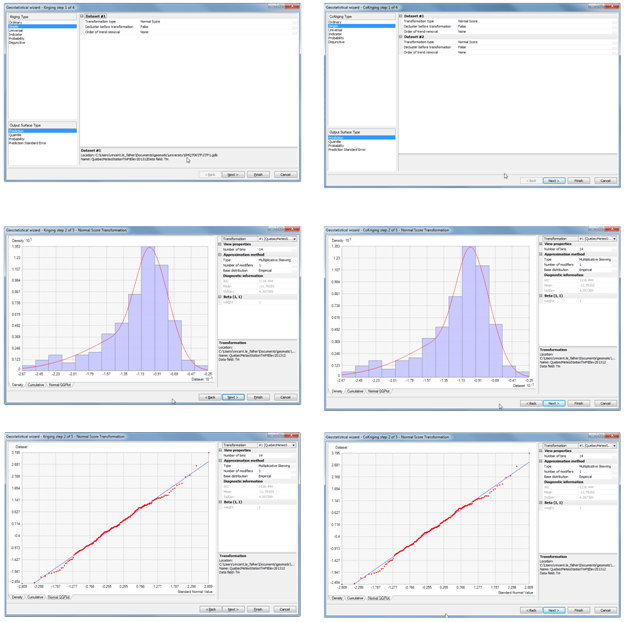


Figure 9 : Question 3 – Objectif 1 – Comparaison des étapes 1-2 – Températures moyennes

|  |  |
| --- | --- |
| krigeage pour les températures moyennes avec les paramètres par défauts | Cokrigeage pour les températures moyennes et l’altitude, avec les paramètres par défauts |

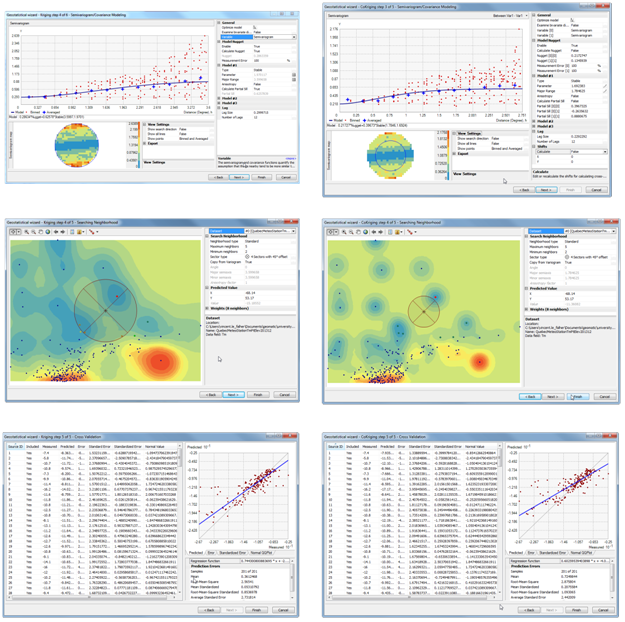


Figure 10 : Question 3 – Objectif 1 – Comparaison des étapes 3-5 – Températures moyennes

|  |  |
| --- | --- |
| krigeage pour les températures moyennes et avec les paramètres par défauts | cokrigeage pour les températures moyennes et l’altitude, avec les paramètres par défauts |

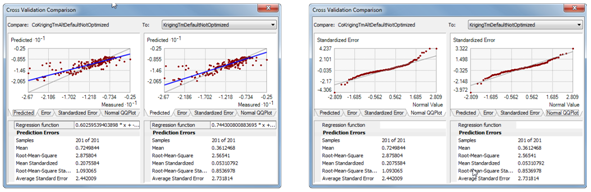


Figure 11 : Question 3 – Objectif 1 – Comparaison des indicateurs d'erreurs – Températures moyennes

|  |  |
| --- | --- |
| krigeage pour les températures moyennes et avec les paramètres par défauts | cokrigeage pour les températures moyennes et l’altitude, avec les paramètres par défauts |

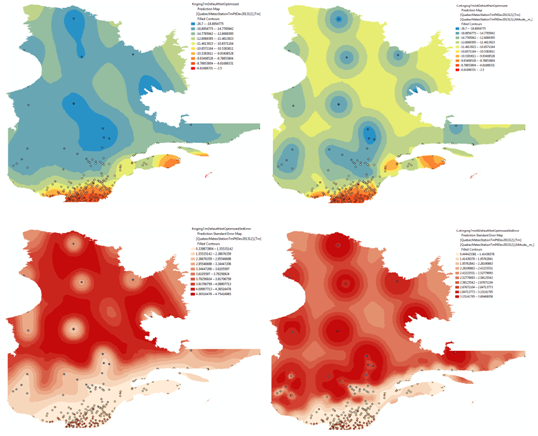


Figure 12 : Question 3 – Objectif 1 – Comparaison des surfaces des modèles d’interpolation – Températures moyennes

# Question 3 – Objectif 2 – Pré-Analyse

Avant lancer le krigeage, nous devons analyser nos variables régionalisées et s’assurer qu’elles répondent au trois prérequis du krigeage :

1. Les données sont t’elles normalement distribuées ?
2. Les données sont t’elles stationnaires ?
3. Y a-t-il une ou des tendances ?

# Températures Moyennes

Voici une synthèse de la pré-analyse. Les détails se trouvent plus bas dans chacune des sections suivantes.

1. grâce à l’histogramme et au QQPlot, nous observons que cette variable est pratiquement normalement distribuée, elle suit une courbe proche de celle de la loi normale : la moyenne et la médiane sont pratiquement identique, l’asymétrie est proche de 0, et le kurtosis proche de 3; l’outil ne nous propose d’ailleurs aucune fonction de transformation.
2. la carte de Voronoi de type « Standard Deviation » nous indique que la stationnarité n’est pas optimale pour l’ensemble de la province, certaines zones sont trop homogènes pour répondre à l’hypothèse de la stationnarité pour un krigeage optimal. Nous suggérons d’utiliser le krigeage empirique EBK afin de tenir compte de la mauvaise stationnarité de ce jeu de données.
3. enfin, l’outil « Trend Analysis » nous confirme que les températures moyennes suivent une tendance de polynôme de 1er ordre: logiquement les températures sont progressivement plus froides en direction du nord pour le mois de Décembre pour la province du Québec. Cette tendance sera indiquée lors de l’exécution du krigeage afin de l’enlever. Les détails se retrouvent en annexe.
   1. Les températures moyennes sont elles normalement distribuées ?

|  |
| --- |
|  |
|  |

Figure 13: Question 3 – Objectif 2 – Pré-analyse de la distribution pour les températures moyennes

L’histogramme et le QQPlot nous confirme que les données des températures moyennes sont normalement distribuées. En effet, l’outil ne nous propose aucune transformation, le coefficient d’asymétrie (*Skewness*) étant de -1.2, la valeur optimale étant 0.

Le coefficient d’aplatissement (*Kurtosis*) vaut 4.3, la valeur optimale étant 3. On en déduit que la distribution est un peu pointue.

La médiane de 10.8 est peu distancée de la moyenne qui est de -11.8. L’outil ne nous propose aucune autre transformation. On en déduit que l’histogramme respecte pratiquement la Loi normale.

Cela permet de dire que les données proviennent d’une distribution qui a pratiquement la même variabilité, ce qui permet de répondre au requis nécessaire pour appliquer l’hypothèse de la stationnarité.

Pour les températures, il n’y a donc pas besoin d’appliquer de transformation pour avoir une variance égale, ce qui est un prérequis pour répondre à l’hypothèse de la stationnarité (ESRI, 2016c).

* 1. Les données sont elles stationnaires ?

|  |  |
| --- | --- |
| La non-stationnarité s’exprime par la présence de « clusters ». On obtient une non-stationnarité pour la plupart du Québec, le bouclier canadien, lorsqu’on utilise le type **« simple ».** |  |
| À l’aide de la carte de Voronoi, nous avons sélectionné le type **« Entropy »** pour obtenir ce qui nous semble le meilleur résultat visuel.  À partir de la partie sud du bouclier canadien, on observe une stationnarité. Au dessus, c’est clairement non stationnaire. Cette dernière partie risque de donner de moins bons résultats. |  |
| D’après la carte de Voronoi avec le type **« Standard Deviation »,** la partie basse Côte Nord et le centre du bouclier n’est pas stationnaire. |  |

Figure 14 : Question 3 – Objectif 2 – Pré-analyse de la stationnarité pour les températures moyennes

On peut en conclure que la stationnarité n’est pas optimale pour un krigeage normale. Pour cette étape, il est difficile de trancher puisqu’il s’agit d’une appréciation qualitative et que les cartes ne sont pas homogènes d’une région à l’autre.

En conséquence, on propose d’utiliser le krigeage empirique EBK.

* 1. Ya t’il une / des tendances ?

L’outil « Trend Analysis » permet d’observer, avec un polynôme de 1er ordre, que les températures d’ouest vers l’est sont semblables, pour un angle de location de 0°; Alors que dans l’axe sud-nord, on observe une tendance des températures vers les négatifs. Ce qui est tout à fait logique.

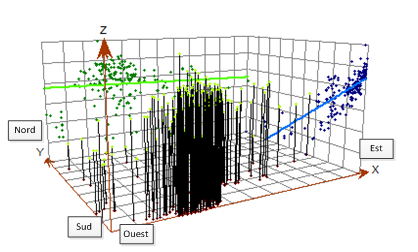


Figure 15 : Question 3 – Objectif 2 – Pré-analyse des tendances pour les températures moyennes

# Précipitations Totales

La même analyse est faite pour étudier les précipitations totales.

Voici une synthèse de la pré-analyse. Les détails se trouvent plus bas dans le tableau.

1) grâce à l’histogramme et au QQPlot, nous observons que cette variable est pratiquement normalement distribuée.

2) la carte de Voronoi de type « InterQuantile Range  IQR» nous indique que le jeu de données pour les précipitions totales peut être considéré stationnaire pour l’ensemble de la province.

3) enfin, l’outil « Trend Analysis » nous confirme qu’il existe une tendance de polynôme de 1er ordre: plus nous allons vers le nord-est, moins il semble y avoir de précipitation pour le mois de Décembre 2013 pour la province de Québec, alors qu’au sud-ouest elles semblent importantes. Nous pensons que notre jeu d’échantillon est responsable de cette « tendance ». Elle sera indiquée lors de l’exécution du krigeage afin de l’enlever.

|  |  |  |
| --- | --- | --- |
| Histogram | On peut sélectionner la colonne à l’extrême gauche qui est à l’écart de la moyenne pour repérer les échantillons qui pourraient être des aberrations. Ces données devraient faire l’objet de vérifications. Au besoin, elles pourraient être retirées. |  |
| QQPlot |  | |
| Carte Voronoi |  | |
| Les tendances | Pas de tendance avec un polynôme de 1er ordre et avec un angle d’observation vers le nord-est. |  |

Tableau 3 : Question 3 – Objectif 2 – Pré-analyse pour les précipitations totales

# Synthèse de la pré-analyse

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Données normalement distribuées | Stationnarité | Pas de tendance | Conclusion krigeage ou non |
| Températures moyennes | . Pratiquement normalement distribuée  . Skewness : -1.78  . Kurtosis : 4.26  . Médiane : -10.8  . Moyenne : -11.8 | On peut en conclure que la stationnarité n’est pas optimale pour un krigeage normale. On est mitigé. | . Pas de tendance ouest-est 0°  . Tendance sud-nord | Comparer le krigeage avec la méthode EKB en raison d’une mauvaise stationnarité dans la partie nord. |
| Précipitations totales | . Pratiquement normalement distribuée. Petite partie déplacée à gauche.  . Skewness : 0,14  . Kurtosis : 3,37  . Moyenne : 69.33  . Médiane : 70.6 | Stationnarité avec le type IQR | . Tendance vers le nord-est. | Oui |
| Modèle  standard | . Bell shape (Loi normale)  . No outliers (aberration)  . Skewness (Asymétrie) = 0  . Kurtosis = 3  . Moyenne = médiane  . Appliquer une transformation pour avoir une variance égale, et qui est un prérequis pour répondre à l’hypothèse de la stationnarité (ESRI, 2016c). | . Voronoi map symbolized by Entropy  . Voronoi map symbolized by Entropy Standard Deviation  . Transformation  . EBK | If the curve through the projected points is flat, no trend exists. (ESRI, 2016c) | Oui |

Tableau 4 : Question 3 – Objectif 2 – Synthèse de la pré-analyse – Températures moyennes et précipitations totales

# Question 3 – Objectif 2-3 : Analyse krigeage Températures Moyennes

Nous observons que la transformation « *Normal Score* » optimisée par le logiciel et tenant compte d’une tendance polynômiale de premier ordre à supprimer, tel que convenue lors de notre pré-analyse, donne une belle distribution asymétrique de type loi normale. Le QQPlot nous indique qu’il n’y a pas d’anormalités dans la distribution, mis à part pour deux points aux extrémités du graphique. Le semivariogramme permet de confirmer que le modèle de la fonction « *Stable* » sélectionné par l’optimisation est adéquat. En effet, la courbe passe proche des valeurs moyennes de chaque décalage, jusqu’au 12ème. Elle passe au milieu du nuage des points, qui devient normalement clairsemée autour d’elle plus les distances entre les stations augmentent. Le nombre de décalage (*lags*) du modèle est de 12, chacun ayant un pas (*lap*) de 0.361 degrés. La différence de température la plus large entre tous les points est de 4.332 degrés. La pépite (*nugget*) est nulle, ce qui indique qu’à une distance de séparation nulle le semivariogramme est nulle, et qu’il ne semble donc pas y avoir d’erreur de mesure ni de variation à de petites échelles inférieures à la distance d’échantillonnage (ESRI, 2016b). La portée (*range*) vaut 2.88 degrés, minimisant l’influence des stations au delà de cette mesure (ESRI, 2016b). La forme de la courbe de la fonction « *Stable* » ressemble à celle de la fonction exponentielle. Elle est très raide proche de l’origine, ce qui nous indique que les voisins les plus proches auront une plus grande influence sur la prédiction (ESRI, 2014). L’assistant du logiciel ne sélectionne pas d’anisotropie par défaut. La recherche des voisins les plus proches est de type « *Standard* » par défaut. Elle est influencée par la fonction du semivariogramme et applique un poids plus ou moins important aux points voisins selon leur distance. Le nombre de points maximums à rechercher dans chacun des 4 secteurs de 45 degrés est de cinq, et le nombre minimum est de deux.

|  |
| --- |
| Étapes 1-2 du krigeage |

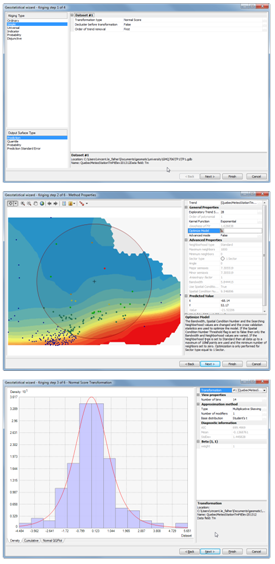


Figure 16 : Question 3 – Objectif 2-3 – Étapes similaires du krigeage – Températures moyennes

|  |  |  |
| --- | --- | --- |
| krigeage avec anisotropie | krigeage sans anisotropie | EBK |

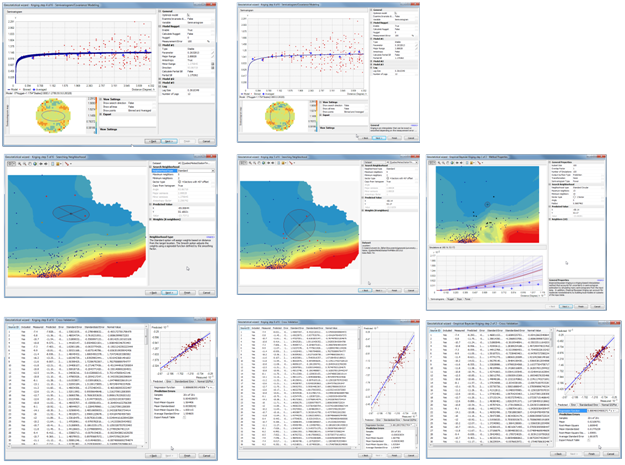


Figure 17 : Question 3 – Objectif 2-3 – semivariogramme et validation croisée – Températures moyennes

|  |  |
| --- | --- |
| Comparaison entre le krigeage sans anisotropie (gauche) et avec anisotropie (droite) | Comparaison entre l’EBK (gauche) et le krigeage avec anisotropie (droite) |

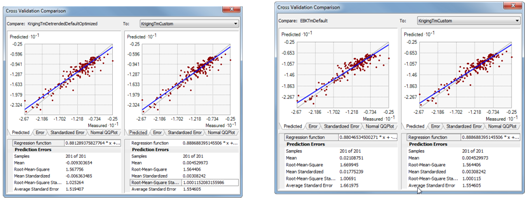


Figure 18 : Question 3 – Objectif 2-3 – Indicateurs d'erreurs – Températures moyennes

|  |  |  |
| --- | --- | --- |
| krigeage avec anisotropie | krigeage sans anisotropie | EBK |

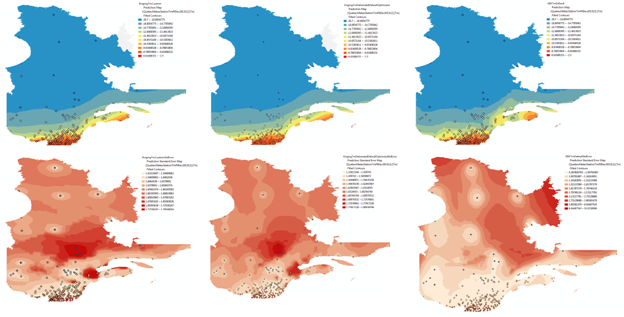


Figure 19 : Question 3 – Objectif 2-3 – Surfaces des modèles d’interpolation – Températures moyennes

# Question 3 – Objectif 2-3 : Analyse krigeage Précipitations

Nous observons que la transformation « *Normal Score* » optimisée par le logiciel et tenant compte d’une tendance polynômiale de premier ordre à supprimer, tel que convenue lors de notre pré-analyse, donne une distribution étalée vers la droite. Le QQPlot nous indique qu’il n’y a pas d’anormalités dans la distribution, mis à part pour deux points aux extrémités du graphique, comme pour la variable des températures moyennes. La comparaison des fonctions disponibles du semivariogramme nous dirige vers la fonction « *Exponential* ». En effet, la courbe passe assez proche de la plupart des valeurs moyennes de chaque décalage, jamais parfaitement, et même assez éloignée pour certaines, tel que pour la première et la dixième. Elle passe au milieu du nuage des points représentant les écarts regroupés (*binned*), qui devient dispersé assez rapidement autour d’elle plus les distances entre les stations augmentent. Le nombre de décalage (*lags*) du modèle est de 12, chacun ayant un pas (*lap*) de 0.252 millimètres. La différence de précipitation la plus large entre tous les points est de 3.027 millimètres. La pépite (*nugget*) vaut 0.689, ce qui indique qu’à une distance de séparation nulle le semivariogramme est 0.689, et qu’il semble donc y avoir des erreurs de mesure ou des variations à de petites échelles inférieures à la distance d’échantillonnage (ESRI, 2016b). La portée (*range*) vaut 2.017 millimètres, minimisant l’influence des stations au delà de cette mesure (ESRI, 2016b). L’assistant du logiciel ne sélectionne pas d’anisotropie par défaut, mais une recherche manuelle sera effectuée pour un autre modèle. Nous laissons la recherche des voisins les plus proches de type « *Standard* » par défaut. Elle est influencée par la fonction du semivariogramme et applique un poids plus ou moins important aux points voisins selon leur distance. Le nombre de points maximums à rechercher dans chacun des 4 secteurs de 45 degrés est de cinq, et le nombre minimum est de deux.

|  |  |
| --- | --- |
| Étapes 1-3 du krigeage par défaut | Étapes 1-2 du krigeage avec et sans anisotropie |

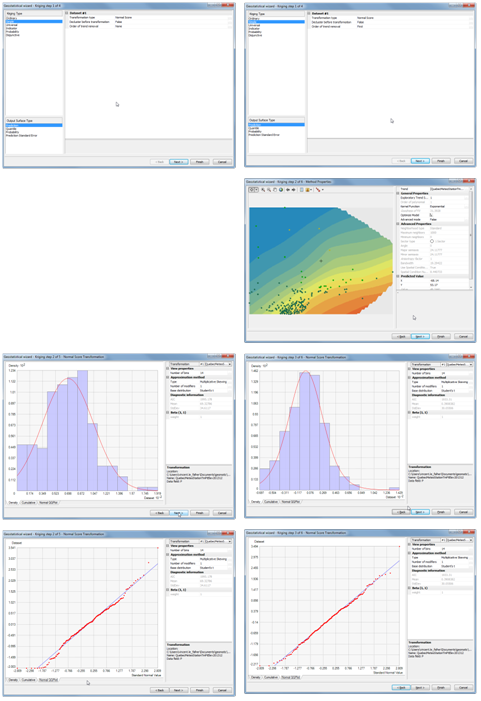
****

Figure 20 : Question 3 – Objectif 2-3 – Étapes 2-3 – Précipitations totales

|  |  |  |  |
| --- | --- | --- | --- |
| krigeage par défaut | krigeage sans anisotropie | krigeage avec anisotropie | EBK |

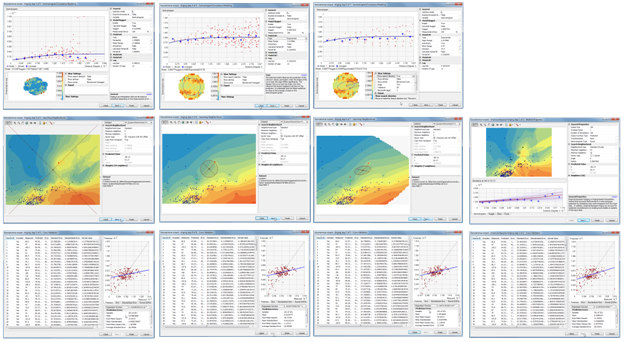


Figure 21 : Question 3 – Objectif 2-3 – Semivariogramme et validation croisée – Précipitations totales

|  |  |  |
| --- | --- | --- |
| Comparaison entre le krigeage avec anisotropie (gauche) et par défaut (droite) | Comparaison entre le krigeage avec anisotropie (gauche) et sans anisotropie (droite) | Comparaison entre le krigeage avec anisotropie (gauche) et EBK (droite) |

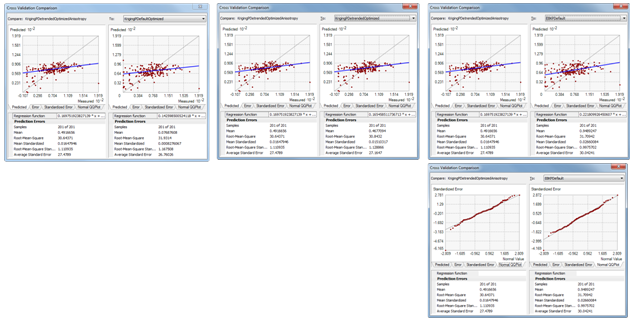


Figure 22 : Question 3 – Objectif 2-3 – Indicateurs d'erreurs – Précipitation totales

|  |  |  |  |
| --- | --- | --- | --- |
| krigeage par défaut | krigeage sans anisotropie | krigeage avec anisotropie | EBK |

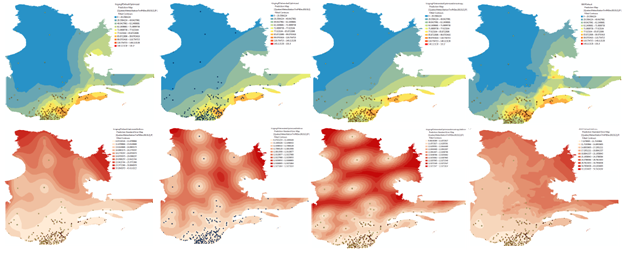


Figure 23 : Question 3 – Objectif 2-3 – Surfaces des modèles d’interpolation – Précipitations totales

# Références bibliographiques

{Auteur. Année. *Titre*. [Source au besoin - ministère, université, etc.]. [Si accessible par lien Internet : Consulté le Date. Hyperlien]}

UQAM. 1999. *Le territoire*. Consulté le 03 Décembre 2016. <https://unites.uqam.ca/hypera/Quebec/territoire.htm>

MDDELCC. 2003. *Suivi hydrologique de différentes stations hydrométriques*. Consulté le 03 Décembre 2016. <https://www.cehq.gouv.qc.ca/suivihydro/default.asp>

ESRI. 2003. *ArcGIS 9 Using ArcGIS® Geostatistical Analyst*. Consulté le 10 Décembre 2016. <http://dusk2.geo.orst.edu/gis/geostat_analyst.pdf>

ESRI. 2010. *Geostatistical Analyst Tutorial*. Consulté le 10 Décembre 2016. <http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/geostatistical-analyst-tutorial.pdf>

ESRI. 2014. *Aide du logiciel ArcGIS 10.2.2*.

ESRI. 2016a. *An overview of the Geostatistical Analyst toolbar and toolbox*. Consulté le 03 Décembre 2016. <http://desktop.arcgis.com/fr/arcmap/10.3/tools/geostatistical-analyst-toolbox>

ESRI. 2016b. *GeoStatistical Analyst*. Consulté le 11 Décembre 2016. <http://pro.arcgis.com/fr/pro-app/help/analysis/geostatistical-analyst>

ESRI. 2016c. *GeoStatistical Analyst*. Consulté le 11 Décembre 2016. <http://desktop.arcgis.com/fr/arcmap/10.3/guide-books/extensions/geostatistical-analyst>

R.Fournier. 2005. *Cours d’analyse et de modélisation spatiale.* Département de géomatique appliquée. Université de Sherbrooke.

S.Baillargeon. 2005. *Le krigeage : revue de la théorie et application à l'interpolation spatiale de*

*données de précipitations*. Faculté des sciences et de génie, Université de Laval.

RNCAN. 2016. *Données Gratuites Géogratis*. Ressources naturelles Canada. Gouvernement du Canada. <http://ftp.geogratis.gc.ca/pub/nrcan_rncan/vector/canvec/shp/Elevation/>

Pierre-André Bourque. 2010. *Planète Terre*. Département de Géologie et de Génie géologique de l’Université Laval. Consulté le 3 Décembre 2016. <http://www2.ggl.ulaval.ca/personnel/bourque/intro.pt/planete_terre.html>

Environnement Canada. 2016. *Données climatiques historiques*. Consulté le 1er Décembre 2016.

<http://climat.meteo.gc.ca/prods_servs/cdn_climate_summary_f.html>