

Programme d'acquisition de connaissances sur les eaux souterraines du Québec

Protocole pour la préparation des livrables

- 15 - Estimation de l'épaisseur des formations superficielles**
- 16 - Topographie du roc**

Par l'Institut national de la recherche scientifique, Centre - Eau Terre Environnement
(INRS-ETE) en collaboration avec la Commission géologique du Canada (CGC)

20 avril 2012

Auteurs

CGC

Nicolas Benoit
Christine Rivard

INRS

Marc-André Carrier
René Lefebvre

TABLE DES MATIÈRES

1	INTRODUCTION.....	1
2	SOURCES DE DONNÉES.....	1
3	SÉLECTION ET VALIDATION DES DONNÉES.....	1
4	TRAITEMENT DES DONNÉES.....	5
5	ÉLÉMENTS À PRÉSENTER	6
6	FORMAT ÉLECTRONIQUE	6
7	ÉLÉMENTS A DEFINIR DANS UNE LEGENDE	7
8	BIBLIOGRAPHIE.....	7

LISTE DES TABLEAUX

Tableau 1 : Statistiques des épaisseurs de formations superficielles en Montérégie Est	2
--	---

LISTE DES FIGURES

Figure 1 : Exemple de cokrigage ordinaire vs cokrigage simple pour les Appalaches internes en Montérégie Est	3
Figure 2 : Exemple du modèle de variogramme directionnel pour les Appalaches externes en Montérégie Est	4

1 INTRODUCTION

La définition de l'épaisseur totale des dépôts meubles, ou encore d'unités aquifères ou aquitard, est d'intérêt pour la compréhension du contexte hydrogéologique d'une région d'étude. De plus, la définition d'un modèle géologique numérique en trois dimensions (3D) peut bénéficier de l'interpolation des épaisseurs des unités géologiques d'intérêt. Le résultat d'une interpolation est largement influencé par le choix des sources de données, de leur sélection, de leur validation, de leur traitement et de la méthode d'interpolation retenue. La sélection finale de la méthode d'interpolation devrait être basée sur une validation objective (ex. : validation croisée entre les données utilisées et les résultats interpolés), une validation qualitative (ex. : jugement professionnel relatif aux contextes et conditions hydrogéologiques du secteur) et sur l'utilisation du résultat (ex. : analyse des tendances régionales, modélisation numérique, calcul de volumes, probabilité que l'estimation soit inférieure ou supérieure à un seuil donné, etc.).

L'objectif du livrable 15 est de représenter spatialement l'épaisseur totale des formations superficielles (dépôts meubles), tandis que celui du livrable 16 est de définir la topographie du socle rocheux. Ces deux livrables sont donc intimement liés. Ce travail demande l'intégration de données de différentes sources et de qualité variable. En fonction du territoire étudié et des données disponibles, il peut être nécessaire de morceler la zone d'étude en différents contextes pour représenter adéquatement ces deux paramètres.

2 SOURCES DE DONNÉES

Les données acquises dans le cadre des projets PACES et qui peuvent être utilisées pour l'élaboration de ces livrables proviennent de différentes sources :

- Bases de données provinciales contenant des données de forages (ex. : SIH, SIGEOM, SIGPEG, bases de données du MTQ, ...);
- Études antérieures, mémoires et thèses universitaires;
- Cartes géologiques indiquant la présence d'affleurements locaux ou les zones de roc sans couverture de dépôts meubles (ministères provinciaux et fédéraux);
- Travaux de terrain réalisés dans le cadre des projets PACES (ex. : forages, sondages, levés géophysiques, ...).

L'ensemble des données tirées des différentes sources devrait être compilé dans une base de données unique avec des champs normalisés.

3 SÉLECTION ET VALIDATION DES DONNÉES

La première étape du processus d'interpolation consiste à valider les données du projet de manière à pouvoir sélectionner les données à retenir. Le processus de validation des données devrait s'appuyer sur le protocole de validation développé par l'UQAC (2010). Certains des critères de validation utilisés dans ce protocole ont été adaptés pour le projet PACES en Montérégie Est et sont documentés en annexe du protocole de l'UQAC.

Une fois cette validation effectuée, la sélection des données peut être basée sur différents critères : fiabilité et validation des données, zonage et jugement professionnel. Chacun de ces critères est décrit ci-dessous.

- A. Fiabilité : critère basé sur les cotes de fiabilité du protocole de validation. Dans le cadre du projet en Montérégie Est, seule la cote de fiabilité FS1, basée essentiellement sur la source de la donnée, a été considérée.
- B. Validation : critère basé sur les différents critères de validation inclus dans le protocole de validation, portant entre autres sur la localisation, l'identification des doublons et des valeurs erronées.
- C. Zonage : critère basé sur la définition de contextes hydrogéologiques et différentes statistiques sur l'épaisseur des dépôts meubles, notamment les deux premiers moments statistiques des données (i.e. la moyenne et la variance), ainsi que la covariance spatiale entre celles-ci. Ce critère peut être utilisé pour définir des zones où les données retenues sont utilisées pour modéliser des variogrammes spécifiques à chaque zone.
- D. Jugement professionnel : critère basé sur les connaissances et la compréhension du système hydrogéologique étudié par le(s) responsable(s) de l'interpolation des données. L'étude des résidus de la validation croisée peut être un moyen objectif et efficace pour identifier les données douteuses.

Bien que ces critères aient été développés pour les données de forages, ils s'appliquent également pour d'autres types de données, incluant les données relatives aux affleurements et aux levés géophysiques.

Étapes proposées pour la validation et la sélection des données de forages :

1. Établir les cotes de fiabilité des données, notamment pour la cote FS1 (critère A) :

Cotes FS1 (UQAC, 2010) :

- 1 - Données extraites d'études ou rapports antérieurs, de mémoires ou de thèses
- 2 - Forages réalisés dans le cadre du PACES
- 3 - Données extraites d'une compilation de données de même source
- 4 - Données extraites d'une compilation de sources multiples
- 5 - Banque de données hydrogéologiques provinciale SIH
- 6 - Source de données douteuses

Il est proposé d'attribuer une plus grande fiabilité aux données collectées lors des travaux de terrain des projets PACES, ainsi que celles extraites des études antérieures ou de thèses et celles provenant des bases de données jugées fiables.

Dans le cadre du projet PACES en Montérégie Est, les valeurs 1 à 4 ont été jugées acceptables et les valeurs 5 et 6 ont été jugées moins fiables. Dans le cas d'une interpolation par cokrigage, il est recommandé de bien étudier la distribution statistique des différentes sources. Le tableau 1 présente les statistiques préliminaires des épaisseurs des formations superficielles en Montérégie Est pour différentes zones définies sur la base de la géologie et de la physiographie.

Tableau 1 : Statistiques des épaisseurs de formations superficielles en Montérégie Est

Régions	Source	Quantité	Moyenne (m)	Médiane (m)	Écart-type (m)
Appalaches externes	fiable (1 à 4)	141	11.2	9.8	8.6
	moins fiable (5 et 6)	8053	7.1	5.2	6.5
Appalaches internes	fiable (1 à 4)	73	13.0	12.0	8.2
	moins fiable (5 et 6)	3352	7.0	4.3	8.1
Basses-terres nord	fiable (1 à 4)	342	24.1	22.5	15.3
	moins fiable (5 et 6)	3315	16.8	15.8	10.3
Basses-terres sud	fiable (1 à 4)	288	12.2	10.0	10.9
	moins fiable (5 et 6)	5413	8.7	7.5	6.1

Les résultats illustrés au tableau 1 montre que l'échantillonnage des données selon leur source peut générer des statistiques différentes. Une telle situation peut nécessiter une révision de la classification des données et/ou une adaptation de la méthode d'interpolation. À titre d'exemple, une interpolation par cokrigage simple pourrait être préférée à une interpolation par cokrigage ordinaire pour la zone des Appalaches internes en Montérégie Est où les forages fiables sont en généralement localisés dans des vallées et la représentation de la moyenne de ces données n'est donc pas représentative de l'ensemble de la zone. La figure 1 illustre l'effet de la diminution de la moyenne des épaisseurs en utilisant le cokrigage simple plutôt que cokrigage ordinaire pour cette région.

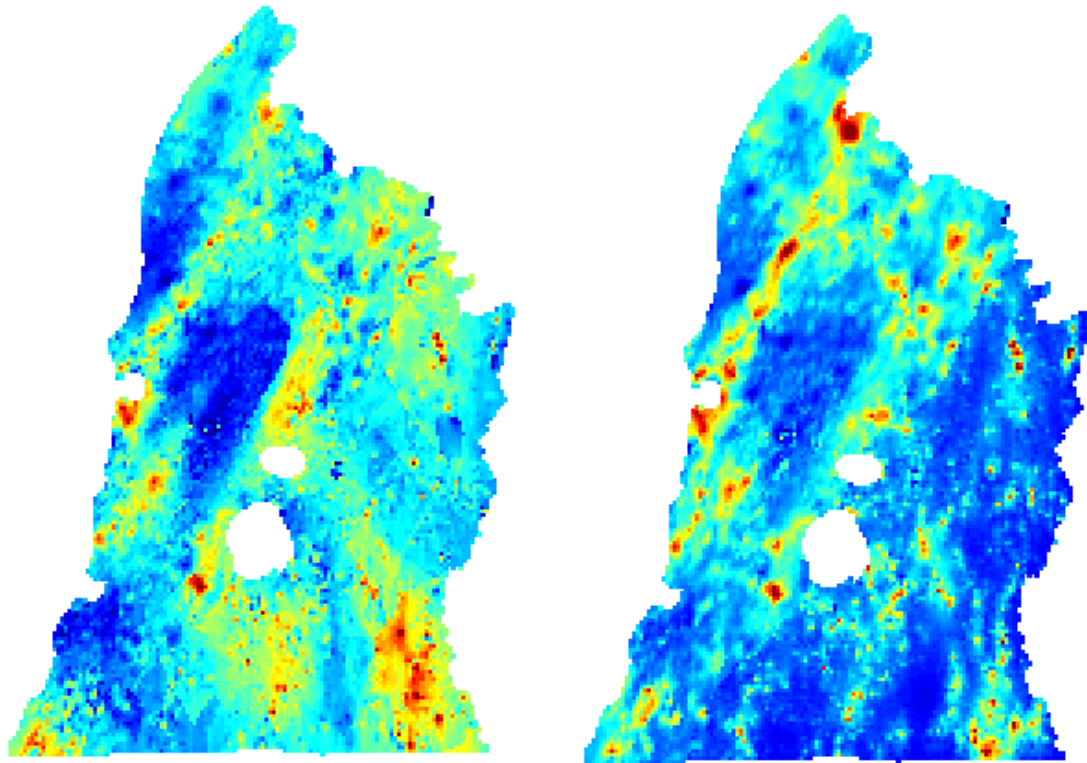


Figure 1 : Exemple de cokrigage ordinaire vs cokrigage simple pour les Appalaches internes en Montérégie Est

2. Valider les données à l'aide des critères de validation (critère B)

3. Explorer les données de façon préliminaire. Il est recommandé de procéder à une analyse spatiale des données (distribution et magnitude) et à la recherche d'erreurs qui n'auraient pas été filtrées par la validation (ex. : valeurs extrêmes impossibles).
4. Établir des zones (critère C) dont les deux premiers moments statistiques des données diffèrent (si pertinent). Pour les données sur les épaisseurs des dépôts meubles, les limites des unités géologiques, contextes hydrogéologiques et/ou physiographiques peuvent être un bon point de départ. Ces zones peuvent être validées ou modifiées par la suite lors de la modélisation des variogrammes et d'un krigeage ordinaire préliminaire.
5. Définir un variogramme théorique en fonction du variogramme expérimental pour chacune des zones identifiées. Il est recommandé de modéliser les variogrammes directionnels pour au moins 4 directions (ex. : 0, 45, 90 et 135 degrés) (voir exemple à la figure 2).

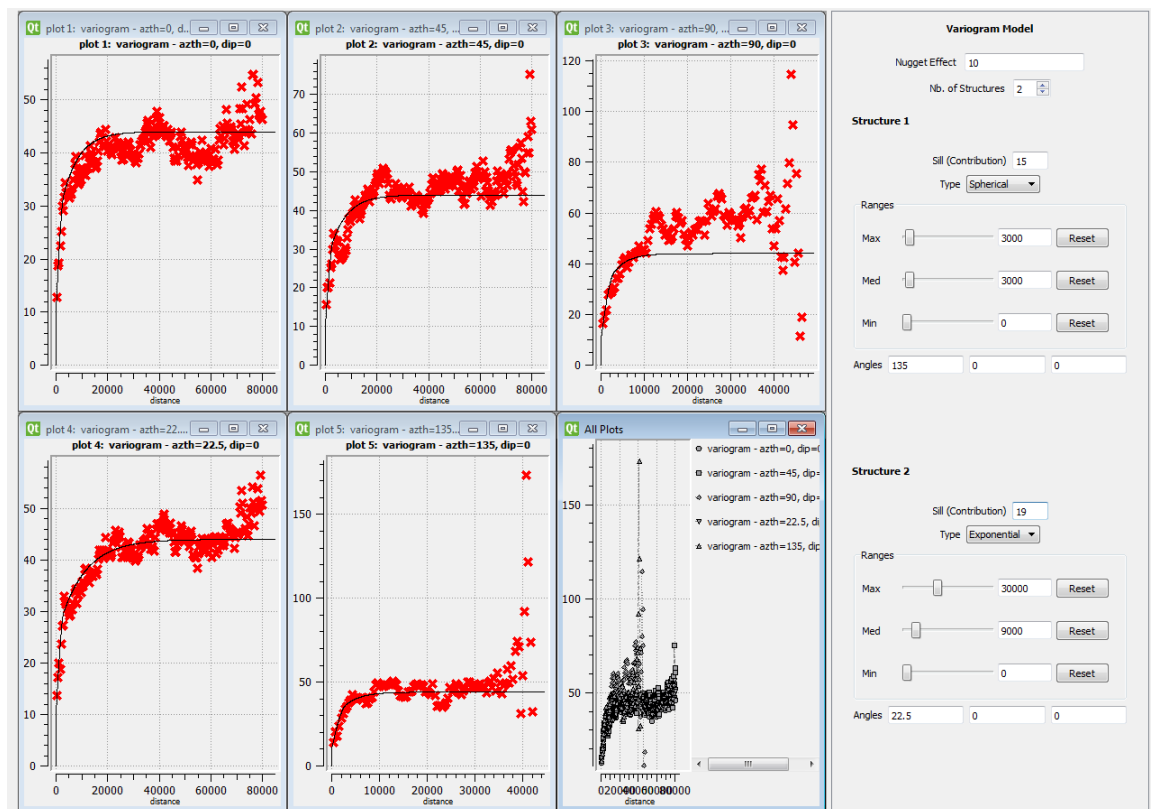


Figure 2 : Exemple du modèle de variogramme directionnel pour les Appalaches externes en Montérégie Est

6. Effectuer une interpolation préliminaire par krigeage ordinaire sur l'ensemble des forages (voir la section suivante). Par la suite, il est recommandé de faire une validation croisée et une étude spatiale des résidus. Lors de cette dernière étape, le jugement professionnel (critère D) des personnes impliquées peut également être utilisé pour identifier les anomalies issues de l'interpolation préliminaire ou pour exclure ou corriger des données.
7. Au besoin, redéfinir les zones.

4 TRAITEMENT DES DONNÉES

Le traitement des données peut être divisé en 8 étapes :

1. Ajuster les variogrammes suite à la sélection et la validation des données et ré-interpoler les données par krigeage ordinaire avec toutes les données de forages sélectionnées.
2. Effectuer une interpolation par cokrigeage ordinaire en séparant les données selon leur fiabilité (ex. : variable primaire correspondant aux données fiables et variable secondaire correspondant aux données moins fiables). Il est possible que le nombre de données fiables soit insuffisant pour modéliser le variogramme de la variable primaire; il est alors proposé d'utiliser le même variogramme que pour les données secondaires mais d'ajuster l'effet de pépité. Il est recommandé d'utiliser seulement les données de forages sélectionnées (sans les données sur les affleurements ou les levés géophysiques) pour déterminer les variogrammes.
3. Effectuer la validation croisée pour le krigeage et le cokrigeage afin d'optimiser les variogrammes et les paramètres d'interpolation. La validation croisée est basée sur la différence entre la valeur mesurée (Z) et l'estimation de cette valeur (Z^*) par interpolation, qui s'appelle le résidu. Les quatre étapes suivantes sont proposées pour l'analyse des résidus (erreurs d'estimation):
 - a. Mesure du biais : $\frac{1}{n} \sum e_i \cong 0$ où le résidu $e_i = Z - Z^*$
 - b. Mesure de la magnitude : $\frac{1}{n} \sum e_i^2$; Valeur à minimiser
 - c. Mesure de validité du modèle d'estimation : $\frac{1}{n} \sum \frac{e_i^2}{\sigma^2} \cong 1$; σ^2 représente la variance de krigeage. La division des résidus de la validation croisée par σ^2 permet de comparer la magnitude de l'erreur réelle et prédite. Ce paramètre peut fournir une appréciation de la validité du modèle et des paramètres utilisés. Plus la valeur est proche de 1, mieux le modèle est ajusté. Par rapport aux deux mesures précédentes, on doit accorder moins d'importance à cette mesure.
 - d. Vérifier si les statistiques générées à partir des résidus sont influencées par des données extrêmes en analysant les histogrammes des résidus e_i ainsi que leur distribution spatiale (N.B. : la distribution devrait être assez homogène).
4. Comparer les mesures des étapes précédentes de la validation croisée et choisir les paramètres et la méthode d'estimation à retenir.
5. Intégrer des données supplémentaires, telles que les affleurements et les interprétations des levés géophysiques. Dans le cas du cokrigeage, ces informations doivent être classées comme fiable ou moins fiable. Il est ici recommandé de reprendre la validation croisée pour mesurer l'effet de l'ajout de ces informations. Il est également suggéré de porter une attention particulière à leur densité en faisant une présélection ou en adaptant le voisinage de l'estimation.
6. Analyser la distribution spatiale des résultats de l'interpolation (N.B. : il est recommandé de travailler avec une grille grossière (ex. : 500x500m) lors des premières étapes du traitement afin de minimiser le temps de calcul).
7. Ajouter si nécessaire des points de contrôle afin d'obtenir une représentation plus « réaliste », notamment dans les secteurs caractérisés par un manque d'information. Il

faut toutefois garder en tête que cette opération se rapproche plus de la modélisation que de l'estimation et qu'elle peut entraîner un biais statistique de l'estimation.

8. Réaliser une interpolation finale sur la grille désirée. Une grille de 250x250m est proposée comme grille de base. Pour toute utilisation des résultats d'interpolation sur une grille différente, il est recommandé de refaire l'interpolation sur la nouvelle grille pour conserver l'effet du modèle de variogramme.

Les interpolations devraient être effectuées sur l'épaisseur des sédiments. La topographie du roc devrait, quant à elle, être générée en soustrayant l'épaisseur interpolée des sédiments à l'élévation du modèle numérique de terrain (MNT) l'estimation de l'épaisseur. Le résultat de cette opération devrait être vérifié afin de s'assurer que les valeurs négatives pouvant découler de la soustraction soient remplacées par une valeur de 0 (i.e. épaisseur nulle).

Le traitement des données d'épaisseur ne devrait pas se limiter pas à ces étapes. Il est notamment souhaitable de valider les interpolations réalisées à l'aide de coupes. Aussi, en fonction des besoins de chaque projet, des données disponibles et des contextes identifiés, le choix de la méthode d'interpolation pourrait différer (ex. : krigeage avec dérive externe ou krigeage d'indicateur). Toutefois, il est recommandé de choisir une méthode géostatistique qui tienne compte de la covariance spatiale et qui permet de donner plus d'importance aux données fiables comme dans le cas du cokrigeage. Parmi ces méthodes, seul le krigeage d'indicateur permet une mesure de l'incertitude dans les méthodes d'estimation. Les simulations géostatistiques, qui n'ont pas été abordées, permettent de mesurer l'incertitude en produisant plusieurs résultats équiprobables. Les méthodes de simulations permettent la représentation de la structure du variogramme, contrairement aux méthodes d'estimation (ex. : krigeage) qui cherche le « meilleur » estimé en un point (moyenne). Ainsi les cartes de simulations présentent une cohérence spatiale entre les points, ce qui n'est pas le cas des estimations. Les cartes de simulations permettent également de reproduire une estimation par la moyenne de plusieurs simulations. Ces méthodes ont également l'avantage de reproduire le variogramme.

5 ÉLÉMENTS À PRÉSENTER

La représentation retenue pour les livrables 15 et 16 consiste en deux cartes qui illustrent par des isocontours les épaisseurs des formations superficielles et l'élévation du toit du roc sur l'ensemble de la zone d'étude. Les éléments à présenter sont :

- Distribution spatiale des épaisseurs en format matriciel (N.B. : des isocontours en mètres peuvent également être ajoutés au besoin) (Livable 15).
- Distribution spatiale de la topographie du roc en format matriciel (N.B. : des isocontours en mètres par rapport au niveau moyen des mers peuvent également être ajoutés au besoin) (Livable 16).
- Localisation des données (i.e. points de contrôle en format vectoriel) utilisées pour l'interpolation
- Données de base (ex. : hydrographie, routes, limites municipales et de MRC et toponymie).

6 FORMAT ÉLECTRONIQUE

Pour le format électronique, les couches matricielles et vectorielles devraient être stockées dans une geodatabase. Les cartes doivent être compatibles avec le logiciel ArcGIS.

7 ÉLÉMENTS A DEFINIR DANS UNE LEGENDE

Les éléments proposés pour la légende sont :

- Échelle de couleur pour les valeurs interpolées (mince ou peu élevé = pâle; épais ou élevé = foncé)
- Isocontours (s'ils sont utilisés) espacés de manière à faciliter la lecture de la carte
- Données ayant servies à l'interpolation représentées par des points noirs ou gris

8 BIBLIOGRAPHIE

Université du Québec à Chicoutimi (UQAC) (2010) Protocole sur la fiabilité et la validation des données de forage - Version 02. Université du Québec à Chicoutimi, janvier 2010, 24 p..