

Phénomènes spatiaux continus – Interpolation 1

Echantillon

Variable mesurée en certains points

Interpolation

Estimation de la valeur en un point non mesuré



Introduction aux systèmes d'information géographique

Bienvenue à cette leçon, qui va porter sur l'interpolation. L'interpolation est une méthode qui va vous permettre de généraliser l'information discrète acquise sur le territoire. Un ensemble de points échantillonnés, selon l'une des procédures présentées à la leçon précédente, est en mesure de fournir des informations intéressantes sur le domaine auquel ils appartiennent. Et en effet, par l'intermédiaire d'indicateurs statistiques comme la médiane ou la moyenne on peut obtenir des informations utiles mais globales sur la zone étudiée. Et l'intérêt principal de l'échantillonnage est de permettre d'estimer la valeur de la variable en tout point du domaine par interpolation. Et quel que soit son mode opératoire, une interpolation est une inférence, c'est-à-dire une prédiction arbitraire ou justifiée d'une valeur de la variable en un point non mesuré.

Notes

Summary



0m 30s

Phénomènes spatiaux continus – Interpolation 1



Buts de la leçon

- Distinguer les méthodes d'interpolation déterministes globales des locales
- Expliquer le fonctionnement des principales méthodes d'interpolation déterministes locales

Après cette leçon vous serez capables

- De calculer des valeurs interpolées avec l'aide des différentes méthodes présentées

Introduction aux systèmes d'information géographique

Les buts de cette leçon sont d'expliquer le fonctionnement des principales méthodes d'interpolation déterministes et leurs déclinaisons en une approche globale et une approche locale. Les notions présentées, vous permettront d'acquérir la capacité de calculer des valeurs interpolées sur tout jeu de données qui contient des échantillons géoréférencés.

Notes

Summary



Méthodes déterministes

Méthodes d'interpolation déterministes

- Aucune étude statistique nécessaire sur le comportement de la variable
- Connaissance experte préalable du phénomène
- Choix de la méthode d'interpolation basé sur cette connaissance



information géographique

Les méthodes d'interpolation dites déterministes présentées dans cette leçon, sont des méthodes lors de l'application desquelles aucune étude statistique sur le comportement de la variable n'est effectuée. Et le choix de la méthode la plus appropriée va reposer sur l'expérience de l'analyste, et sur les connaissances qu'il possède du phénomène étudié.

Notes

Summary



1m 52s

Méthodes déterministes globales versus locales

Méthode globale

- Prend en compte tous les points mesurés

Régression
polynomiale

Introduction aux systèmes d'information géographique

Il existe deux familles de méthodes d'interpolation déterministes, ce sont les méthodes globales et les méthodes locales. Une méthode est dite globale, si le modèle spatial du phénomène est construit en prenant en compte tous les points de mesure existants sur le domaine étudié.

Notes

Summary



2m 12s

Méthodes déterministes globales versus locales

Méthode globale

- Prend en compte tous les points mesurés

Régression
polynomiale

Méthode locale

- Prend en compte uniquement un nombre limité de points mesurés, au voisinage du point à estimer
- Permet d'attribuer des poids particuliers, notamment en fonction de la distance

Plus proche
voisin

Réseau de
triangles

Fonction
B - spline

Moyenne
pondérée

Introduction aux systèmes d'information géographique

Les méthodes locales quant à elles ne prennent en compte qu'un nombre limité de points d'appui au voisinage du point à estimer. De plus, elles permettent d'attribuer des poids particuliers à chaque point de mesure notamment en fonction de la distance.

Notes

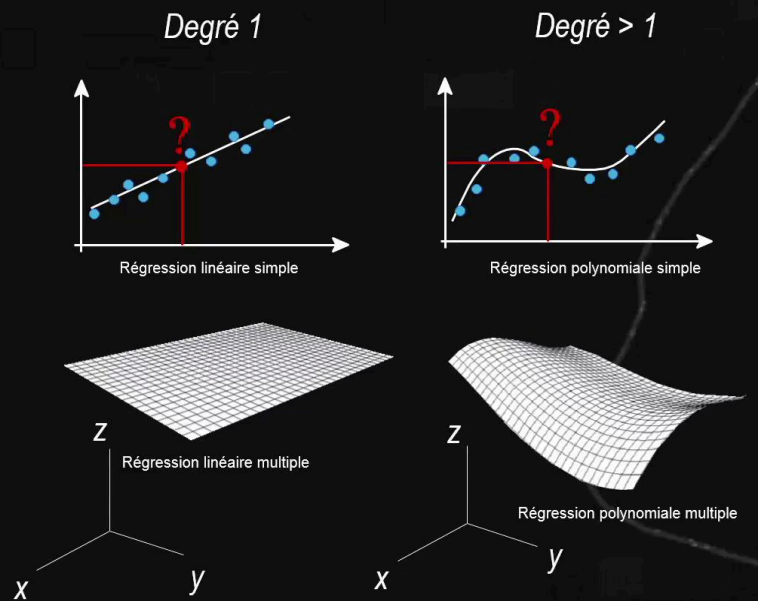
Summary



2m 28s

Méthode globale: régression polynômiale

- Polynôme qui minimise les écarts aux points de mesure (moindres carrés)
 - Création d'une surface de tendance
 - Degré inférieur à 4: modèle souvent trop simple pour fournir une approximation réaliste du phénomène
 - Degré supérieur à 5: difficile à calculer
- ➔ Méthodes locales souvent préférées



Introduction aux systèmes d'information géographique

Les méthodes globales sont utilisées pour analyser des tendances de distribution spatiale d'un phénomène dans un domaine donné. Et nous présentons ici, une approche globale qui consiste à calculer un polynôme de degré 1 ou de degré supérieur et qui minimise les écarts aux points de mesure selon la méthode des moindres carrés. Il en résulte une surface de tendance, représentative du comportement moyen du phénomène dans le plan des coordonnées X Y et Z représente ici la variable mesurée. Les modèles de référence qui peuvent être soit un plan soit une surface de degré 2 ou 3 sont trop simples dans la majorité des cas, pour fournir une approximation réaliste de la distribution spatiale de la variable. Mais les surfaces de degré supérieur à 5 ne sont pas faciles à calculer. C'est la première raison pour laquelle on préfère les méthodes d'interpolation locales. Une autre raison plus fondamentale est qu'il n'est pas très utile de faire intervenir des points de mesure au-delà d'une distance seuil d'auto-corrélation car la dépendance spatiale est par hypothèse locale et non globale.

Notes

Summary



2m 51s

Méthodes locales

- Nombre limité de points
- Pondération
- Plusieurs méthodes
- Leur choix dépend de:
 - Densité d'information
 - Précision souhaitée
 - Objectif de l'interpolation
 - Connaissance préalable du phénomène



information géographique

Les méthodes d'interpolation locales ne prennent en compte qu'un nombre limité de points de mesure au voisinage du point à estimer. De plus, elles permettent de pondérer ces points de mesure en fonction de leurs distances aux pixels à prédire ou encore éventuellement en fonction d'une information de qualité de mesure liée à l'utilisation d'un instrument spécifique. Plusieurs méthodes existent et leur choix dépend de la densité de l'information que l'on possède, de la précision souhaitée, de l'objectif principal de l'interpolation et comme évoqué plus tôt, de la connaissance préalable que l'on a du phénomène.

Notes

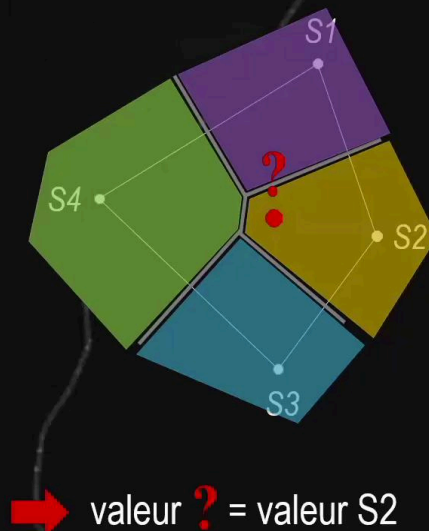
Summary



4m 03s

Plus proche voisin

- Valeur estimée = valeur du point mesuré le plus proche
- Polygones de Thiessen (médiatrices des segments reliant les points mesurés)
- Les points à estimer qui tombent à l'intérieur d'un même polygone obtiennent la même valeur



Introduction aux systèmes d'information géographique

La première méthode présentée est celle du plus proche voisin. La valeur attribuée en un point quelconque du domaine est celle de son plus proche voisin. La détermination du plus proche voisin est basée sur l'appartenance du point à interpoler, à la zone d'influence des points de mesure. Ces zones d'influence sont déterminées par les médiatrices des segments. Ces médiatrices délimitent ce que l'on appelle les polygones de Thiessen. Et tout point à prédire, situé à l'intérieur d'un polygone, est sensé prendre la même valeur que celle du point de mesure correspondant, soit ici la valeur S2. Cette méthode transforme en fait le phénomène continu en un modèle spatialement discret.

Notes

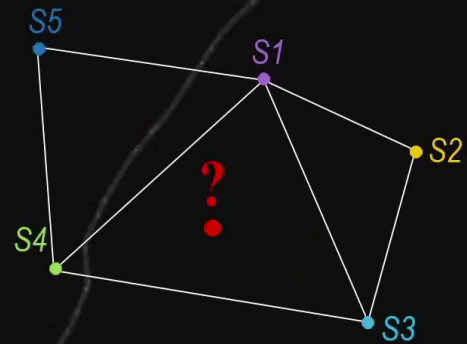
Summary



4m 40s

Réseau de triangles (TIN)

- Réseau de triangles formé en reliant entre eux les points mesurés (Delaunay)



Introduction aux systèmes d'information géographique

La deuxième méthode d'interpolation locale que nous décrivons est une méthode basée sur un réseau de triangles, abrégé TIN en anglais, pour Triangulated Irregular Network. Un TIN est formé en reliant les points d'appui entre eux, de manière à former un réseau de mailles triangulaires. Ce TIN est élaboré à partir de la méthode de Delaunay, dont la particularité est de créer des triangles dont les angles ne sont pas supérieurs à 90 degrés.

Notes

Summary



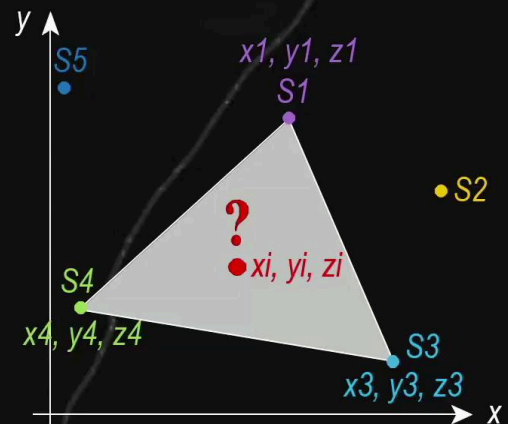
5m 20s

Réseau de triangles (TIN)

- Réseau de triangles formé en reliant entre eux les points mesurés (Delaunay)
- On suppose une variation linéaire à l'intérieur de chaque triangle
- Equation du plan formé par chacune des faces triangulaires

$$z = ax + by + c$$

➔ calcul de la valeur z_i en ? à partir de x_i, y_i



Introduction aux systèmes d'information géographique

La variation du paramètre à l'intérieur de chaque triangle est donc supposée linéaire. Et l'interpolation est réalisée en établissant l'équation à 3 inconnues du plan qui est formé par chacune des facettes triangulaires. Les trois inconnues sont a, b et c. Et elles sont déterminées par les valeurs connues des trois sommets du triangle. Pour mettre en œuvre cette méthode, on choisit en général des points de mesure caractéristiques, qui sont situés par exemple sur les crêtes, ou au fond des talwegs. À l'origine, l'interpolation par TIN a été développée en cartographie pour créer manuellement des lignes d'isovaleurs, comme les courbes de niveaux du relief.

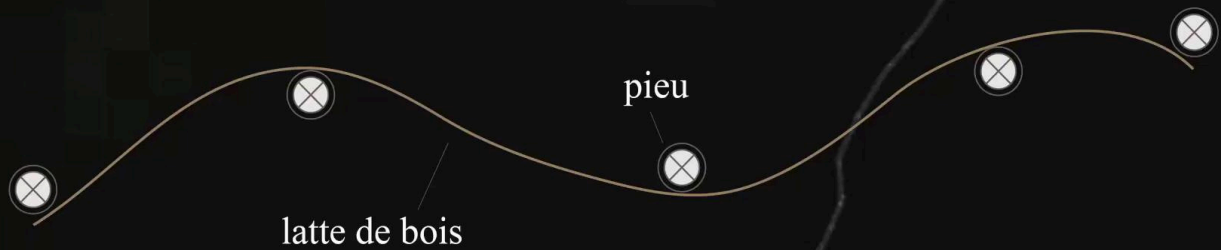
Notes

Summary



5m 45s

Fonction B-spline



Historiquement, le mot spline désigne en anglais, une latte de bois longue et mince, utilisée par les jardiniers pour dessiner une courbe passant par des pieux plantés en terre. Une application plus récente de la latte de bois est la règle flexible appelée Cobra, utilisée pour dessiner une courbe simulant au mieux la répartition linéaire d'une collection de points. La méthode B-spline est une mathématisation de cette pratique.

Notes

Summary



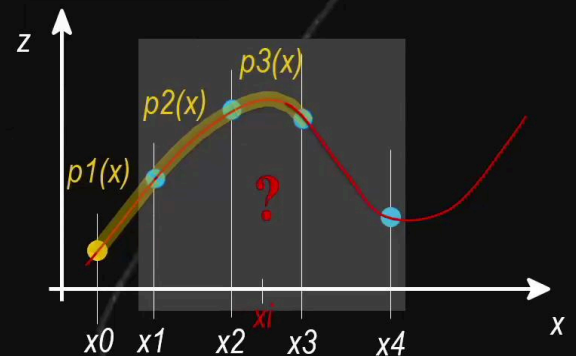
6m 23s

Fonction B-spline

- Fenêtre mobile autour de 4 points mesurés successifs
- Ajustement d'un polynôme de troisième degré
- On glisse ensuite la fenêtre d'un point et on estime à nouveau un polynôme
- Contraintes imposées à la jonction de deux polynômes consécutifs pour assurer une continuité

$$p_k(x_{k+1}) = p_{k+1}(x_{k+1})$$

$$p'_k(x_{k+1}) = p'_{k+1}(x_{k+1})$$



Introduction aux systèmes d'information géographique

Au lieu de rechercher une seule équation qui modélise une courbe passant par tous les points d'appui, un polynôme du troisième degré est calculé pour 4 points successifs, contenus dans une fenêtre mobile, comme l'illustre la figure à l'écran, pour le cas à une dimension. On suppose que pour chaque intervalle, l'évolution de la variable à prédire Z de X , est correctement simulée au moyen d'un polynôme P de X . On glisse ensuite la fenêtre mobile d'un point vers la droite et on estime à nouveau le polynôme, pour l'intervalle suivant. Le calcul du polynôme P de X est élaboré par l'introduction de contraintes à la jonction de deux polynômes consécutifs. La première contrainte est l'intervalle dans lequel X est compris, la deuxième concerne la continuité au changement de segment, en jaune ici. La troisième concerne la continuité au cas où un point d'inflexion se trouve au point de jonction de deux polynômes. Dans ce cas, la continuité est assurée par égalité approximative des dérivées secondes.

Notes

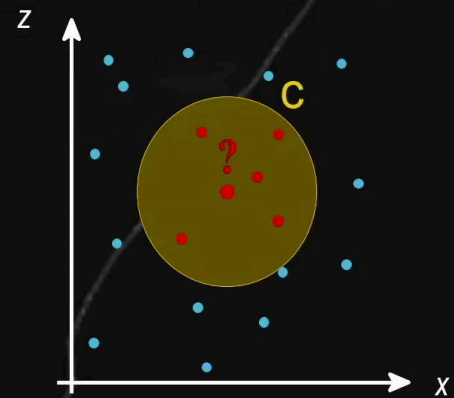
Summary



6m 56s

Moyenne mobile pondérée

- Estimation en s'appuyant sur les valeurs mesurées de quelques points du voisinage
- Paramètres à définir:
 - Taille du voisinage / nombre de points pris en compte
 - Poids: les points plus éloignés ont moins d'influence que les points proches
 - Tenir compte de l'orientation ? (anisotropie)



Introduction aux systèmes d'information géographique

Passons maintenant à l'interpolation par moyenne mobile pondérée, ou Inverse Distance Weighting en anglais, très souvent abrégée IDW. L'estimation de la variable est calculée en s'appuyant sur les valeurs de quelques points mesurés dans le voisinage, et en leur attribuant un poids, en fonction de la distance notamment. Cette approche correspond à l'intuition selon laquelle un point éloigné a moins d'influence qu'un point rapproché. Pour mettre en œuvre l'IDW, il faut répondre à certaines questions. Quelle taille de voisinage définir? Combien de points de mesure faut-il inclure? Quel poids respectif convient-il de leur accorder?

Notes

Summary



7m 55s

Moyenne mobile pondérée

- Estimation en s'appuyant sur les valeurs mesurées de quelques points du voisinage
- Paramètres à définir:
 - Taille du voisinage / nombre de points pris en compte
 - Poids: les points plus éloignés ont moins d'influence que les points proches
 - Tenir compte de l'orientation ? (anisotropie)



Faut-il tenir compte de l'orientation relative du phénomène? Et là, c'est la notion d'anisotropie que nous abordons et sur laquelle nous reviendrons plus en détail lors de la leçon suivante. La réponse à ces questions essentielles varie en fonction de la région d'étude. Elle varie aussi en fonction du phénomène considéré et de la configuration de l'échantillon des points de mesure. Dans le cadre déterministe, l'analyste doit se baser sur sa propre expérience, pour fixer des valeurs aux paramètres évoqués. En ce qui concerne la pondération par exemple, on considère que plus un point mesuré est éloigné du point interpolé, moins il a d'influence. On choisit dans ce cas une fonction inverse de la distance, ou inverse au carré de la distance.

Notes

Summary

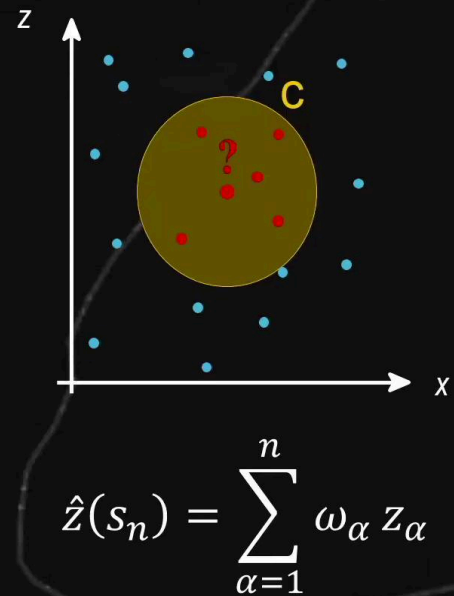


8m 36s

Moyenne mobile pondérée

- Estimation en s'appuyant sur les valeurs mesurées de quelques points du voisinage
- Paramètres à définir:
 - Taille du voisinage / nombre de points pris en compte
 - Poids: les points plus éloignés ont moins d'influence que les points proches
 - Tenir compte de l'orientation ? (anisotropie)

➔ valeur en ? = moyenne pondérée des voisins



Introduction aux systèmes d'information géographique

Une fois que le nombre de points d'appui est sélectionné, soit ici 5 points situés dans le cercle C autour de la valeur à prédire, cette dernière est égale à la moyenne pondérée de ces 5 points d'appui. Le poids accordé aux points de contrôle Z de α peut par exemple être proportionnel à 1 sur la distance, ou 1 sur la distance au carré, ou encore 1 sur la distance à une puissance plus élevée. Et plus la puissance est élevée, plus l'effet local est accentué.

Notes

Summary



9m 21s

Risque d'erreurs

Risque d'erreurs des méthodes déterministes

- Méthodes empiriques, les résultats dépendent de l'expérience de l'analyste et de sa subjectivité
- Choix de la méthode d'interpolation parfois arbitraire



Les méthodes locales d'interpolation que nous venons de passer en revue ne contiennent pas de moyens d'estimer l'incertitude liée aux valeurs estimées. Pour évaluer cette incertitude, on doit recourir à une méthode de validation croisée. Cela consiste à partager l'échantillon de manière aléatoire en deux parties. Les points de mesure de l'un des deux groupes servent à l'interpolation. Alors que les points de mesure du second servent de tests. En chaque site test, on va comparer la valeur estimée avec la valeur mesurée et en déduire l'incertitude. En général, en calculant la différence entre les moyennes quadratiques des deux distributions. La moyenne quadratique est la racine carrée de la moyenne arithmétique des carrés des valeurs. La méthode est répétée plusieurs fois avec à chaque reprise un nouveau jeu de sites d'appui et de sites tests aléatoirement déterminés, de manière à éviter tout biais dans la sélection de sites. Les résultats de cette validation croisée permettent d'évaluer l'incertitude globale et l'incertitude locale des valeurs interpolées. Les méthodes du plus proche voisin du réseau de triangles, de la fonction B-spline et de la moyenne mobile pondérée, sont très empiriques et les résultats sont fortement dépendants de l'expérience de l'analyste mais aussi de sa subjectivité.

Notes

Summary



9m 50s

Risque d'erreurs

Risque d'erreurs des méthodes déterministes

- Méthodes empiriques, les résultats dépendent de l'expérience de l'analyste et de sa subjectivité
- Choix de la méthode d'interpolation parfois arbitraire



Pour les phénomènes qui échappent à la perception visuelle, la situation peut être critique dans la mesure où, en l'absence d'information supplémentaire, rien n'indique si l'interpolation est possible. Il est important de noter qu'en effet, en l'absence de toute information, soit de tout modèle vérifié de comportement, le choix de la fonction d'interpolation est totalement arbitraire, comme le montre cette illustration.

Notes

Summary

11m 08s

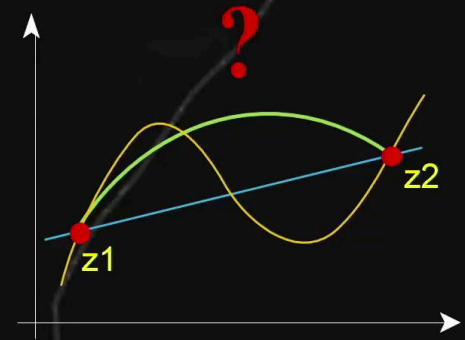


Risque d'erreurs

Risque d'erreurs des méthodes déterministes

- Méthodes empiriques, les résultats dépendent de l'expérience de l'analyste et de sa subjectivité
- Choix de la méthode d'interpolation parfois arbitraire

➔ Variables régionalisées et géostatistique



Introduction aux systèmes d'information géographique

En effet, quel est le comportement de la variable Z entre les points $Z1$ et $Z2$? Est-ce que c'est selon la fonction bleue ? Est-ce que c'est selon la fonction verte ? Ou encore selon la fonction orange ? Ce type d'incertitude peut entraîner des coûts inutiles et très élevés, dans le cas des sondages miniers par exemple. Et c'est pour les éviter que dès les années 1940, des travaux de recherche se sont développés pour proposer le concept de variables régionalisées, que nous allons étudier dans la prochaine leçon.

Notes

Summary



11m 32s