

Exercice 9 : Interpolation spatiale

Objectifs :

- Comprendre les concepts de l'interpolation
- Interpolation spatiale avec Inverse Distance Weighting (IDW)
- Estimation de densité (Kernel Density Estimation)

Introduction

Mise en contexte

Dans cet exercice, vous allez interpoler la quantité de phosphore dans le canton de Fribourg, et également la densité de séismes en Suisse.

Données :

- **communes:** fichier vectoriel (ESRI Shapefile) représentant les limites administratives des communes suisses.
- **seisms_ch:** fichier vectoriel (ESRI Shapefile) représentant les séismes ayant eu lieu entre 2001 et 2008 en Suisse et dans ses environs. (Source : [Service Sismologique Suisse](#))
- **soil_sample:** fichier vectoriel (ESRI Shapefile) correspondant aux échantillons de sols (245 points d'échantillonnage) effectués dans le canton de Fribourg pour relever divers indicateurs (pH, matière organique OM, phosphore, etc.) sur différents types de sol agricoles (134 échantillons dans les terres assolées (culture), 67 dans les prairies et 44 dans les pâturages d'alpage).

Préparation du dossier de travail

1. Ouvrez un nouveau projet dans QGIS
2. Importez la couche des communes suisse (**communes.shp**)
3. Vérifiez que le système de projection du projet est adapté pour votre étude

Exercice

Partie 1 : Interpolation de la quantité de phosphore dans les sols

Dans cette première partie, nous allons interpoler la quantité de phosphore dans le canton de Fribourg à partir d'échantillons relevés dans les sols.

Visualisation des données

Nous allons commencer par visualiser la variable que nous souhaitons interpoler, à savoir la quantité de phosphore relevée lors de la campagne d'échantillonnage.

1. Importez dans QGIS la localisation des points d'échantillonnage du sol (**soil_sample.shp**)
2. Attribuez une palette de couleur graduée à la couche **soil_sample** en fonction de la quantité de phosphore (attribut **Pt_HF**). Dans **Properties** > **Symbology** de la couche, spécifiez les paramètres suivants :
 - **Style** : Graduated
 - **Column** : Pt_HF
 - **Color ramp** : YlOrRd
 - **Mode** : Natural Breaks (Jenks)
3. (optionnel) Vous pouvez augmenter la taille des points dans l'option **Symbols** et en modifiant l'attribut *Size*.
4. Cliquez sur **Classify** et **OK**

À partir de ces points, le processus d'interpolation va nous permettre d'estimer la quantité de phosphore dans les sols pour l'ensemble de la zone d'étude (pas seulement aux points d'échantillonnage).

Division du jeu de données

Avant de commencer l'interpolation, nous allons diviser l'échantillon en deux parties aléatoires. Une partie des points sera utilisée pour générer l'interpolation ("train sample") et la seconde pour vérifier le résultat de l'interpolation ("test sample").


1. Sélection aléatoire de 80 points d'échantillonnage (pour le test sample) avec l'outil **Random Selection** (**Vector** > **Research Tools** > **Random Selection**)
 - **Input layer** : **soil_sample**
 - **Method** : si nous voulons sélectionner un certain nombre de points ou alors un pourcentage du nombre total (ici, *Number of selected features*)
 - **Number/percentage of selected features** : le nombre/pourcentage voulu (ici, 80 points)
2. Exportez les entités sélectionnées (cf. Série 1, p12) dans une nouvelle couche **soil_test_sample.shp**.
3. Inversez la sélection dans la table des attributs de la couche **soil_sample** (cf. Série 1, p5) pour exporter dans une nouvelle couche, que vous nommerez **soil_train_sample.shp**, l'autre partie des points.


Interpolation avec la méthode Inverse Distance Weighting (IDW)

Les méthodes d'interpolation vont permettre d'estimer la valeur d'un attribut en chaque cellule d'une grille (raster) à partir d'un certain nombre de points. En utilisant une approche IDW, la valeur d'une cellule sera estimée en calculant la moyenne pondérée (en fonction de la distance) des points d'échantillons voisins. Dans notre cas, nous allons interpoler l'attribut représentant le phosphore total présent dans les sols.

1. Sélectionnez l'outil **IDW interpolation** dans la **Processing Toolbox** (vous pouvez directement le trouver avec la barre de recherche).
2. Spécifiez les paramètres suivants :

- **Vector layer:** couche contenant les points que nous souhaitons interpoler (ici, **soil_train_sample**)
- **Interpolation attribute:** attribut à interpoler (ici, la quantité de phosphore total *Pt_{HF}*)

Appuyez ensuite sur 

- (si nécessaire) **Distance coefficient P:** plus le coefficient est petit, plus le poids donné aux observations voisines diminuera rapidement avec la distance (ici, choisissez une valeur de 5).
 - **Extent:** étendue de la grille à interpoler selon des coordonnées, votre canevas QGIS, ou l'étendue d'une couche spécifique en cliquant sur  . Dans notre cas, nous souhaitons interpoler sur l'ensemble de la couche **soil_sample** même si nous n'utilisons qu'un échantillon de points pour faire l'interpolation et il faut donc sélectionner cette couche sous « *Use layer extent* ».
 - **Output raster size:** résolution de votre raster final (dans notre cas, vous pouvez utiliser une résolution de 50m et la diminuer en augmentant la taille des pixels si l'algorithme prend trop de temps).
 - **Interpolated:** fichier de sortie (ici, **Pt_idw.tif**)
3. Cliquez sur Run.

Évaluation de la performance de l'interpolation

Maintenant que nous avons la couche interpolée est chargée dans QGIS, nous allons tester la validité de l'interpolation à l'aide de la couche **soil_test_sample**. Autrement dit, nous voulons voir si les valeurs calculées par interpolation aux positions de **soil_test_sample** correspondent à celles qui sont indiquées dans la colonne *Pt_{HF}*.

Pour retourner pour chaque point d'une couche, la valeur de la cellule raster dans laquelle le point se situe :

1. Installez le plugin **Point sampling tools** dans **Plugins > Manage and Install Plugins** puis ouvrez-le dans **Plugins > Analyses > Point Sampling Tools**.

2. Spécifiez les paramètres suivants :

- **Layer containing sampling points** : sélection de la couche contenant les points (ici, **soil_test_sample**).
- **Layers with fields/bands to get values from** : la liste des couches:attributs/bandes qui se superposent aux points et dans lesquelles nous voulons extraire les valeurs (ici, nous voulons comparer l'attribut *Pt_HF* de la couche **soil_test_sample** et la valeur interpolée par IDW contenue dans **Pt_HF_idw.tif**). Pour sélectionner plusieurs couches, maintenez *Ctrl* appuyé.
- **Outpoint point vector layer** : fichier de sortie (par exemple **soil_test_val.shp**).

3. Cliquez sur OK.

Dans la couche résultante, chaque ligne permet de comparer à un même point la valeur échantillonnée dans les sols (*Pt_HF*) et la valeur interpolée avec Inverse Distance Weighting (*Pt_HF_idw*).

Pour évaluer la qualité de l'interpolation, nous pouvons calculer la moyenne des erreurs (ME) et l'erreur quadratique moyenne (RMSE) avec les formules suivantes :

$$ME = \frac{1}{v} \sum_{j=1}^v [z^*(s_j) - z(s_j)], \quad RMSE = \left\{ \frac{1}{v} \sum_{j=1}^v [z^*(s_j) - z(s_j)]^2 \right\}^{0.5}$$

où $z^*(s_j)$ est la valeur interpolée, $z(s_j)$ la valeur réelle et v le nombre de valeurs. La moyenne devrait être proche de 0 pour une méthode non-biaisée et le RMSE le plus petit possible pour une prédiction précise et non-biaisée.

Question 1: Commentez les résultats de l'interpolation.

Partie 2 : Estimation de la densité de séismes en Suisse

Dans cette deuxième partie, nous allons aborder l'approche du Kernel Density Estimation à l'aide d'un nouveau jeu de données représentant les séismes ayant eu lieu en Suisse et alentours entre 2001 et 2008.

Importez la couche **seims_ch** dans QGIS

La méthode KDE permet de produire un raster où la valeur de chaque cellule représente la densité de points compris dans un certain rayon. Cela va nous permettre de produire une carte indiquant la densité de séisme attendue en tout point du territoire.

Pour créer une heatmap avec KDE :

1. Sélectionnez l'outil **Heatmap (Kernel Density Estimation)** dans la **Processing Toolbox**.
2. Spécifiez les paramètres suivants :
 - **Point layer** : la couche contenant les points (ici, **seims_ch**)
 - **Radius** : rayon définissant les points que nous allons considérer pour le calcul de la density (essayez avec une valeur de 1km, 10km et 50km)
 - **Output raster size** : résolution du raster (ici, choisissez une résolution de 50m pour le rayon 1km et une résolution 500m pour 10km et 50km)
 - **Heatmap** : fichier de sortie (nous nommerons les couches **seisms_heatmap_1km.tif**, **seisms_heatmap_10km.tif**, **seisms_heatmap_50km.tif**)

Note : Dans *Advanced parameters*, vous pouvez spécifier un rayon ou une pondération en fonction d'un attribut. Par défaut, tous les points ont un poids de 1.

3. Cliquez sur Run.
4. Visualisez les couches avec une palette de couleur (**Properties > Symbology**) :
 - **Render Type** : Singleband pseudocolor
 - **Color ramp** : une palette divergente (p.ex. Spectral, Magma), et inversez l'échelle de couleur.
5. Cliquez sur Classify, Apply et OK.

Une fois les 3 couches créées, zoomer sur une région contenant un point isolé et observer les valeurs de densité obtenues dans les pixels aux alentours. Est-ce que ces valeurs correspondent à vos attentes ?

Le canton du Valais vous a demandé d'évaluer le risque de séisme pour ses communes. Évaluez les différentes cartes produites.

Question 2 : *Quel est l'effet d'un rayon plus grand/plus petit ? Quel rayon vous semble plus adapté pour cette analyse ?*