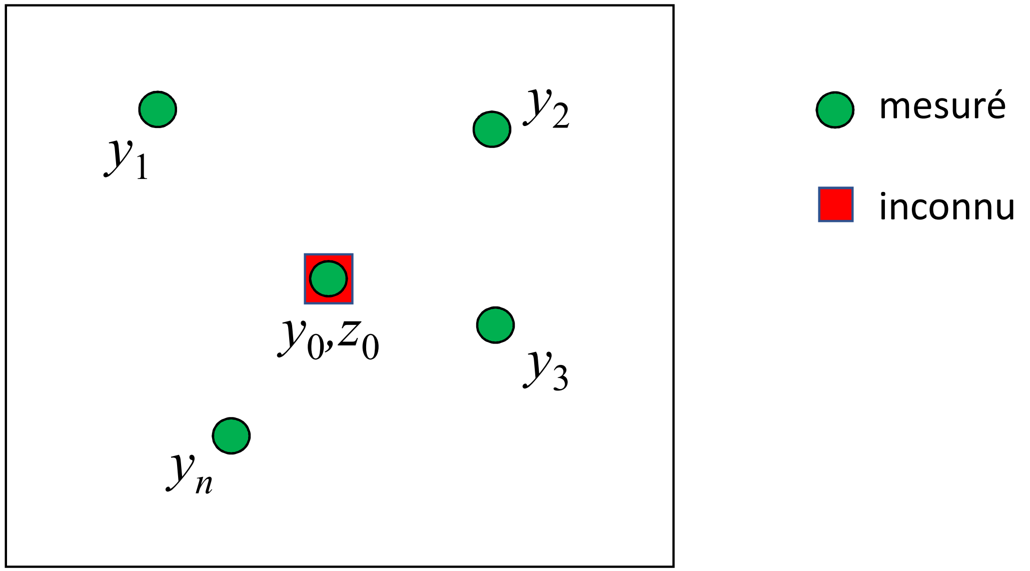
Prédiction spatiale : l'approche *Bayesian Data Fusion* (BDF)

P. Bogaert – UCL
Janvier 2018

Prédiction spatiale : définition

Fournir une information de la meilleure qualité possible sur une variable Z en un endroit x_0 à partir d'informations sur une (des) variable(s) Y mesurées en des endroits $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$.



Prédiction spatiale : cas de figure

Ce qu'il faut prédire :

- Prédiction de variables continues vs variables catégorielles
- Prédictions ponctuelles vs agrégées sur des supports

Ce dont ont dispose :

- Mesures spatiales vs spatio-temporelles
- Mesures ponctuelles vs agrégées
- Mesures univariées vs multivariées
- Mesures exactes vs incertaines

Prédiction spatiale: ce qui est souhaitable

Pour les input (informations):

- Souplesse pour intégrer des informations multiples
- Prise en compte d’une qualité variable des informations

Pour les output (prédictions) :

- Qualité optimale des prédictions
- Mesure de l’incertitude sur les prédictions
- Temps de calcul & complexité raisonnables

Prédiction spatiale : types de méthode

Méthodes linéraires :

- On accorde des poids au mesures
- On prédit sur base d'une combinaison linéaire pondérées

- *Exemples* : distance inverse, krigeage, geographically weighted regression, splines, …

Méthodes non linéraires :

- Tout le reste (neural networks, some kriging variations, BME/BDF, …)

Prédiction spatiale : ce qu'on cherche

- Une expression pour la loi conditionnelle

$$f(\overset{\text{inconnu}}{z_0} \mid \overbrace{y_0, y_1, \ldots, y_n}^{\text{input } (y_0,\mathbf{y})})$$

- Un prédicteur ayant des propriétés « optimales »

<u>ex</u> : $E[z_0 \mid y_0, y_1, \ldots, y_n]$

- Une mesure de la qualité de prédiction

<u>ex</u> : $Var[z_0 \mid y_0, y_1, \ldots, y_n]$

Prédiction spatiale : ce que nous faisons (BDF)

- Choisir une distribution conjointe de départ

$$f(\mathbf{z}) = f(z_0, z_1, \ldots, z_n)$$

- Recoder l'information apportée par les y_i 's

$$(\hat{y}_i, \hat{z}_i) \rightarrow f(z_i | y_i)$$

- Intégrer l'information pour la distribution conditionnelle

$$f(z_0|\mathbf{y}) = \frac{1}{B}\frac{f(z_0|y_0)}{f(z_0)} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\mathbf{z}) \prod_{i=1}^{n} \frac{f(z_i|y_i)}{f(z_i)} \mathrm{d}z_1 \cdots \mathrm{d}z_n$$

Prédiction spatiale : ce que nous faisons

Avantages :

- Permet d'incorporer des y_i's arbitraires via les $f(z_i|y_i)$'s
- Peu d'hypothèses de modélisation pour ces $f(z_i|y_i)$'s
- Prise en compte de l'incertitude des input via les $f(z_i|y_i)$'s
- Plusieurs input $y_{i,1}, \ldots, \mathrm{y}_{i,m}$ aux mêmes endroits sont possibles
- Compatible avec des choix classiques pour $f(\mathbf{z})$

Désavantages :

- Basé sur une hypothèse d'indépendance conditionnelle (pas forte)
- Coûteux dans sa formulation la plus générale (intégrales multiples)

Prédiction spatiale : ce que nous faisons

Simplifications :

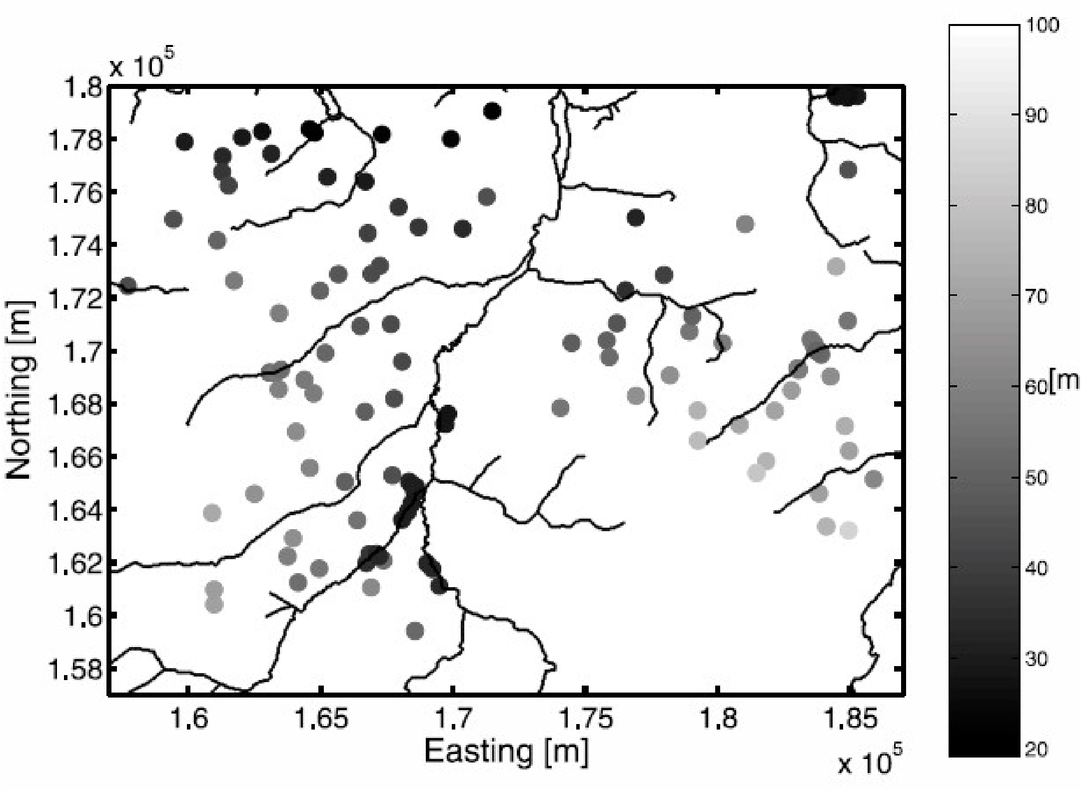
- Se réduit à une formule simple si les inputs sont $y_0, z_1, \ldots, z_n$

distribution conditionnelle
« classique » (krigeage)

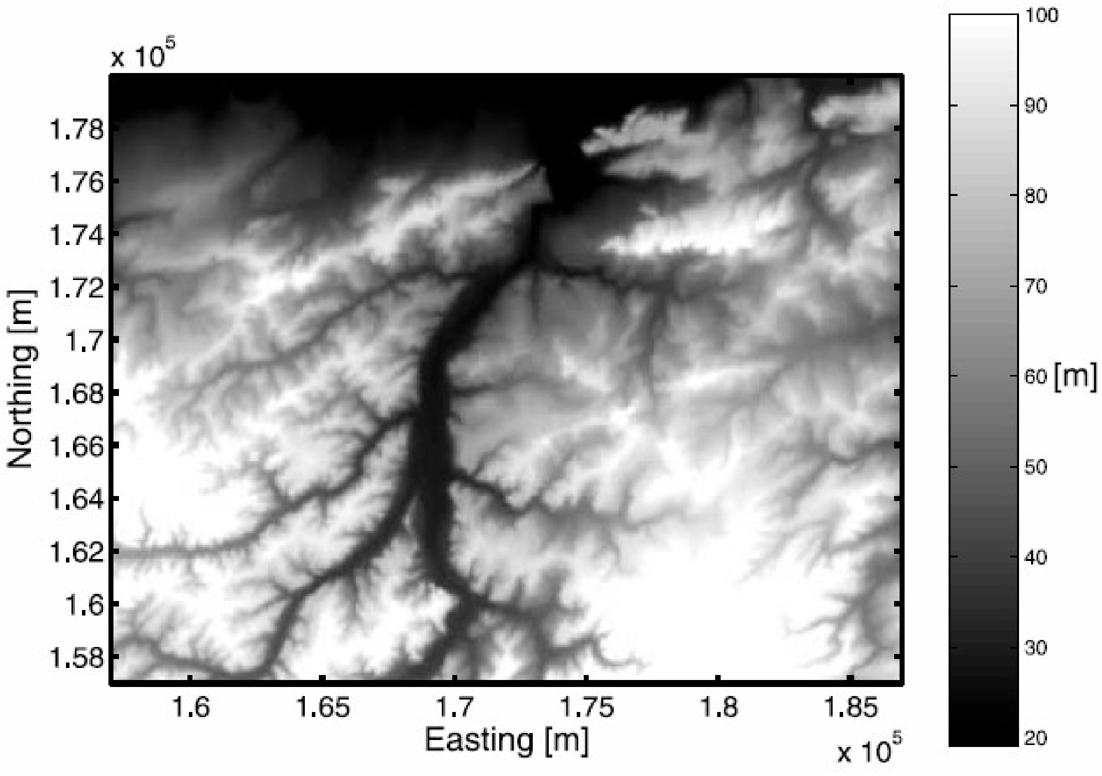
$$f(z_0 \mid \underbrace{y_0, z_1, \ldots, z_n}_{\text{input } (y_0, \mathbf{z})}) = \frac{1}{B} \frac{f(z_0 | y_0)}{f(z_0)} \, f(z_0 | \mathbf{z})$$

- Se réduit à une prédiction linéaire si $f(\mathbf{z})$ et $f(z_i|y_i)$'s sont gaussiens (kriging with measurement errors)

Prédiction spatiale : une illustration (water table depth)

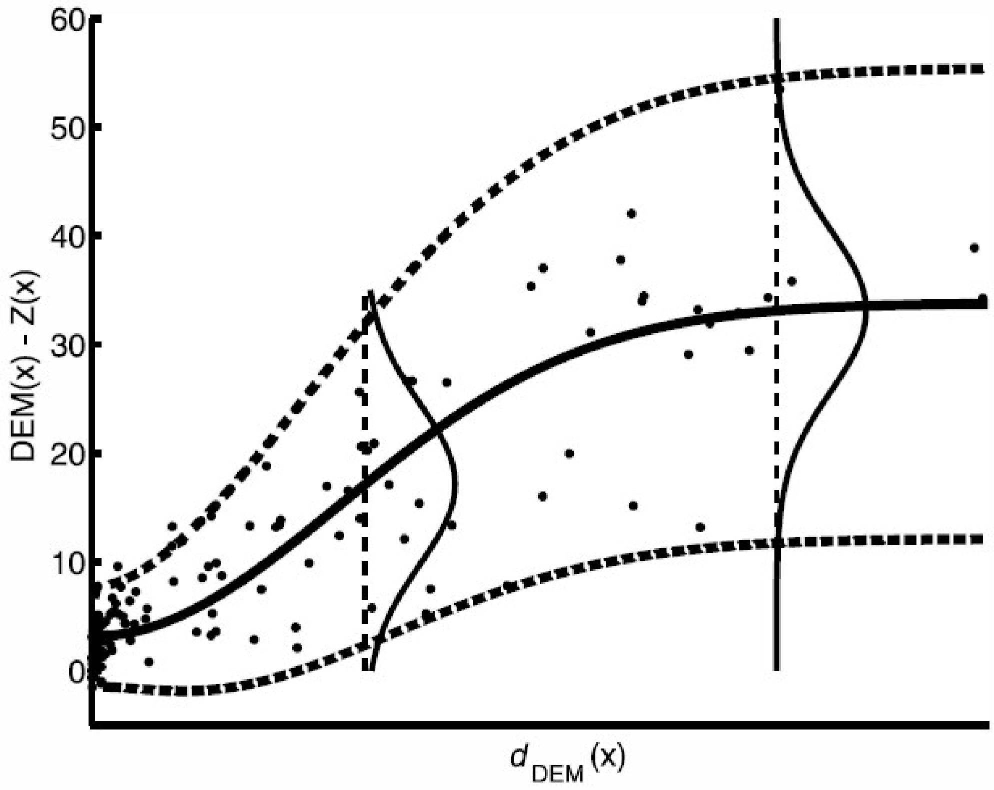


Rivers network and sampled water table depth



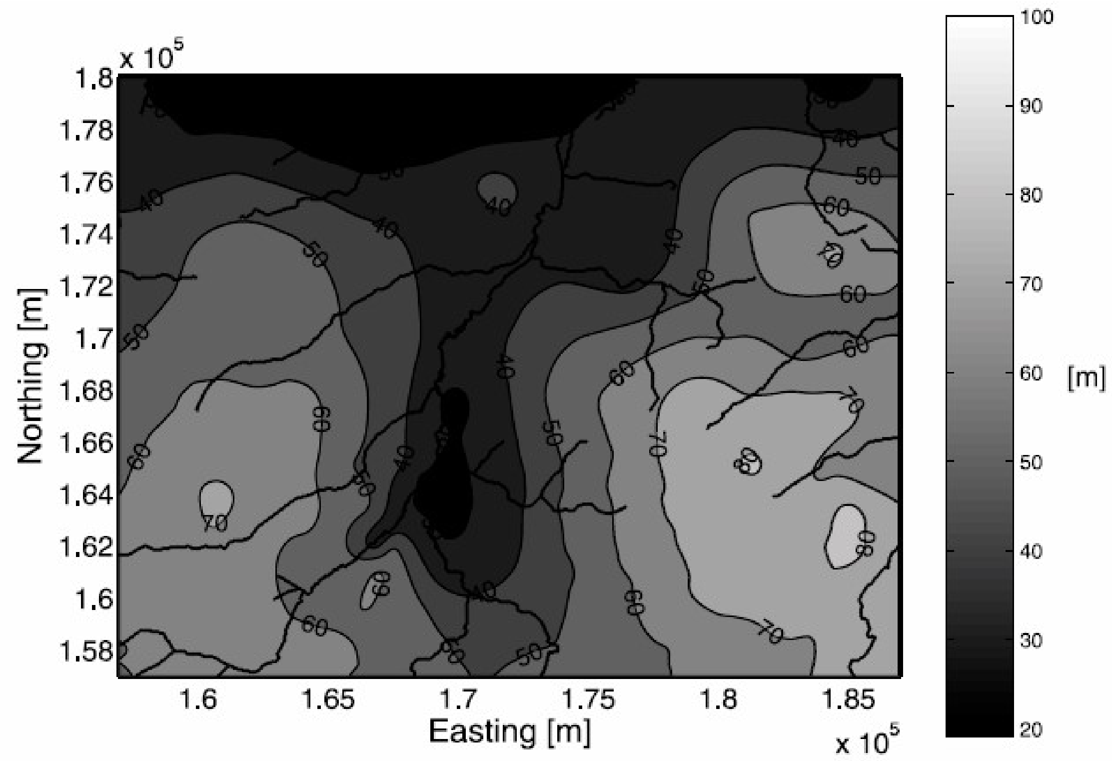
Digital Elevation Model

Prédiction spatiale : une illustration (water table depth)

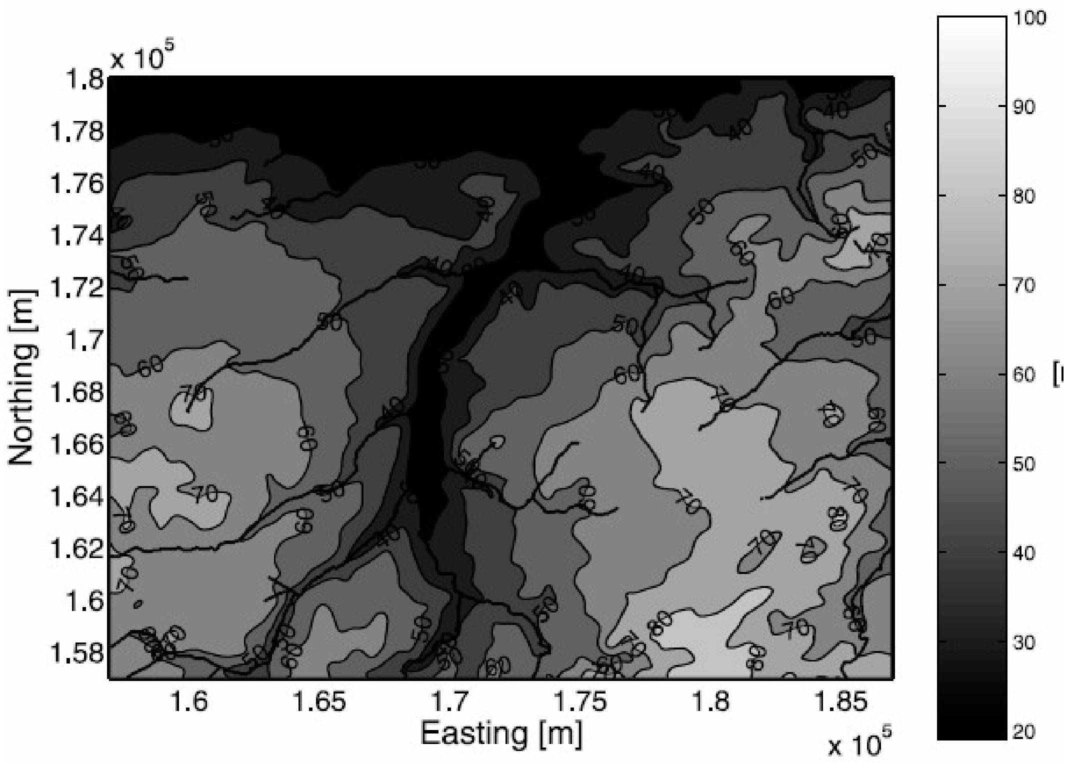


Différences entre profondeur de la nappe $Z(\mathbf{x})$ et $DEM(\mathbf{x})$
en fonction de la distance à la rivière $d_{DEM}(\mathbf{x})$

Prédiction spatiale : une illustration (water table depth)



Carte de la nappe par cokrigeage



Carte de la nappe par BDF