Mapping from point observations

D G Rossiter

Problem

Spatial Prediction

Universal model

Mapping methods

Stratification

Global methods

Local methods

Locally-adaptive methods

Choosing a mapping method

Mapping from point observations

D G Rossiter

Section of Soil & Crop Sciences
College of Agriculture & Life Sciences
Cornell University

November 30, 2021

1. Problem
2. Spatial Prediction
3. Universal model
4. Mapping methods
   - Stratification
   - Global methods
   - Local methods
   - Locally-adaptive methods
5. Choosing a mapping method

What we know

- One or more **attributes** have been measured at a set of **“point” locations** with defined **coördinates** in geographic space.
  - 0-dimensional “point” actually has some spatial extent, its **support**
    - pH, organic C etc. in soil sample from an auger: 4 cm diameter (2D) + 10 cm length (3D)
    - biomass from vegetation plot 10 x 10 m (2D)
    - temperature, precipitation, relative humidity at a weather station (“point” instrument but variable is the same over some radius)
- The coördinates could also be or include **time.**

Example set of "point" observations

318 O. Atteia, J.-P. Dubois, R. Webster

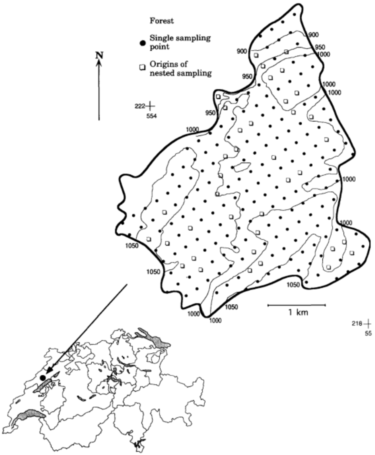
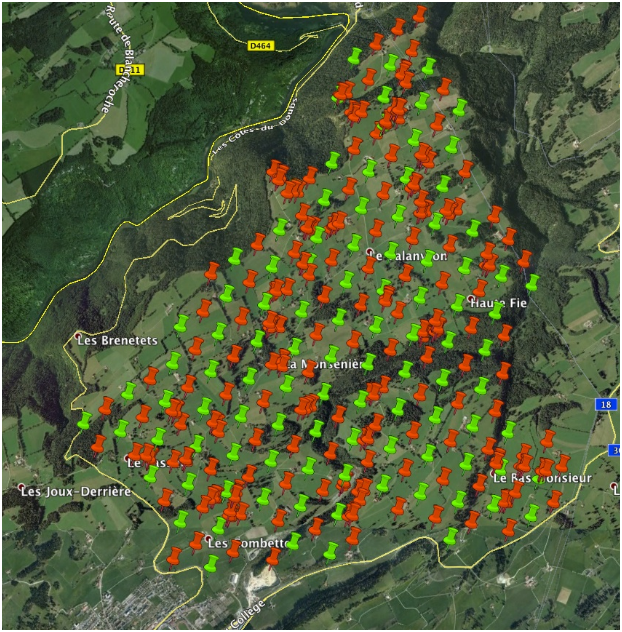


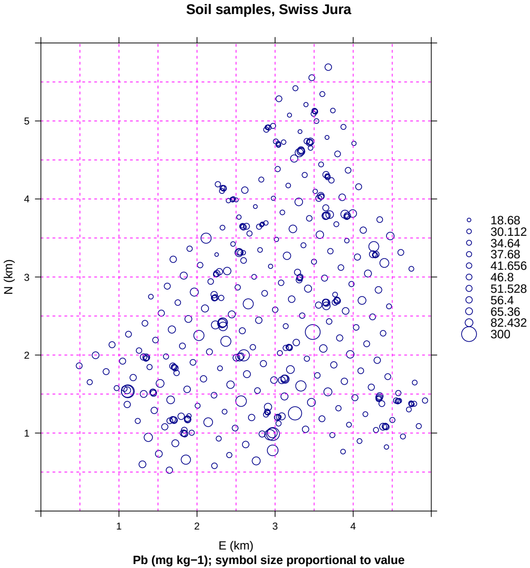
Fig. 1. Location of the region studied and of the sampling points.

Atteia, O., Dubois, J. P., & Webster, R. (1994). Geostatistical analysis of **soil contamination** in the Swiss Jura. Environmental Pollution, 86(3), 315–327. `https://doi.org/10.1016/0269-7491(94)90172-4`

KML file in Google Earth



Observed values of an attribute at "points"



Support: 5 cm diameter tube, 0-25 cm depth

What we want to know (1)

- The value of the attributes at **other** (unvisited) “point” locations, either . . .
  - with the same **support** as the original observations, or . . .
  - . . . with some other support, usually larger (“block”)
- This requires **spatial prediction** based on the observed “point” observations

What is the value at this “point”?

With the same support as the observations.

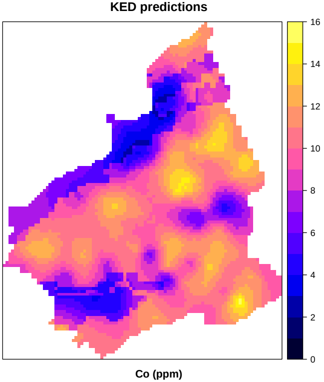
What is the value over this block?

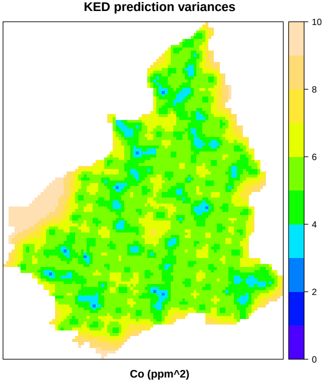
Mean, maximum, standard deviation . . .

What we want to know (2)

- The values of the attributes over a **grid** of other “point” locations → a **map** of the attributes
  - predict at **point** support at centre of grid, or . . .
  - . . . predict as **grid** support, i.e., average over the grid cell, or . . .
  - . . . predict as a **continous surface** which can be queried at any location
- Again, **spatial prediction** based on the observed “point” observations

“Point” predictions at centres of $(50m)^2$ grid cells





This prediction method also provides a **prediction variance** (uncertainty measure)

Discrete vs. Continuous predictions

- **Points**: prediction at single locations (with some support)
  - **discrete**; can be any set of points, also on a regular **grid**
  - identified by coördinates in as many dimensions as the object (can be in space, time or both)
- **Surfaces**: conceptually, a **continuous**, **smooth** prediction; can be examined anywhere
  - often presented as a regular grid, but must be able to compute at any location given by coördinates
  - 1-D: lines, 2-D: surfaces; 3-D volumes or 2-D+time, 3-D+time

Spatial prediction

- Two objectives: (1) practical and (2) scientific
- **Objective 1** (*practical*): Given a set of **attribute values** at **known points**, **predict** the value of that attribute:
  - at other “points”,
  - over a block (area),
  - or over a surface.
  - Preferably with the **uncertainty** of the prediction.
- **Objective 2** (*scientific*): **Understand** why the attribute has its spatial distribution.
- These may require different methods

Spatial modelling vs. spatial mapping

Modelling A **conceptual** and **statistical** representation of the **geographic distribution of the observations**

- **conceptual**: what geographic factors determine the geographic distribution?
- **statistical**: how are these represented in computation?

Mapping Using the statistical model to **predict** at unknown locations, typically regular-spaced across the study area

A taxonomy of spatial prediction methods

Strata: divide area to be mapped into 'homogeneous' **strata**; predict **within each stratum** from all observations in that stratum.

Global: (or "regional") predictors: use **all observations** to build a model that allows to predict at **all points** or over a surface.

Local: predictors: use only **'nearby'** observations to predict at each point.

Mixed: predictors: some of structure is explained by strata or globally, the **residuals** from this are explained **locally**

These are discussed in detail, below.

Universal model of spatial variation

$$Z(\mathbf{s}) = Z^*(\mathbf{s}) + \varepsilon(\mathbf{s}) + \varepsilon'(\mathbf{s}) \tag{1}$$

- $(\mathbf{s})$ a location in space, designated by a **vector** of coördinates(1D, 2D, 3D)
- $Z(\mathbf{s})$ **true** (unknown) value of some property at the location
- $Z^*(\mathbf{s})$ **deterministic** component, due to a **non-stochastic** process
- $\varepsilon(\mathbf{s})$ **spatially-autocorrelated stochastic** component
- $\varepsilon'(\mathbf{s})$ pure (“white”) **noise**, no structure

These components each require a **model**

adapted from: Matheron, G. (1969). Le krigeage universel. École nationale supérieure des mines de Paris; Cahiers du Centre de morphologie mathématique de Fontainebleau, fasc. 1.

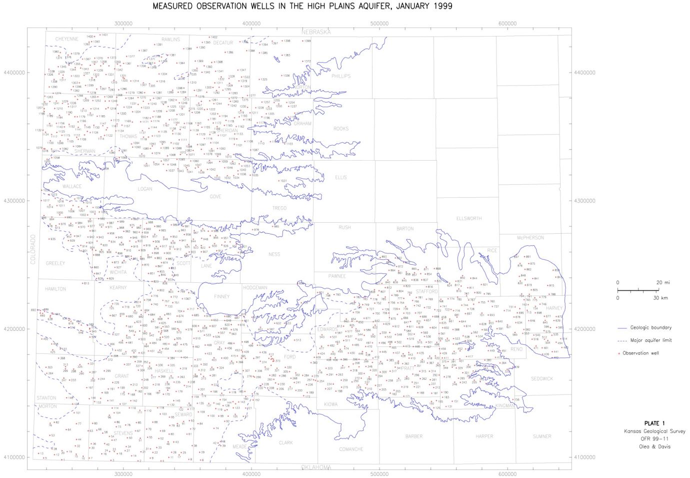
Spatial autocorrelation – concept

- "Auto" = "self", i.e., an attribute correlated to itself
- "Spatial": the correlation depends on the **spatial relation** between points.
- Key idea: observations have a relation in both **geographic** and **feature** (attribute) spaces.
- Can be applied to an attribute (observation) $Z(\mathbf{s})$ or the **residuals** $\varepsilon(\mathbf{s})$ from some deterministic model

A 2D geographic example

- **attribute to map**: elevation above sea level of the top of an aquifer in Kansas (USA)
- **observed** at a large number of wells ("points")
- Q: What determines the spatial variation? (the **physical process**)
- Q: How can we **model** this from the observations? (using the **universal model** of spatial variation)
- Q: How can we **map** over a regular grid covering the region, using the model?

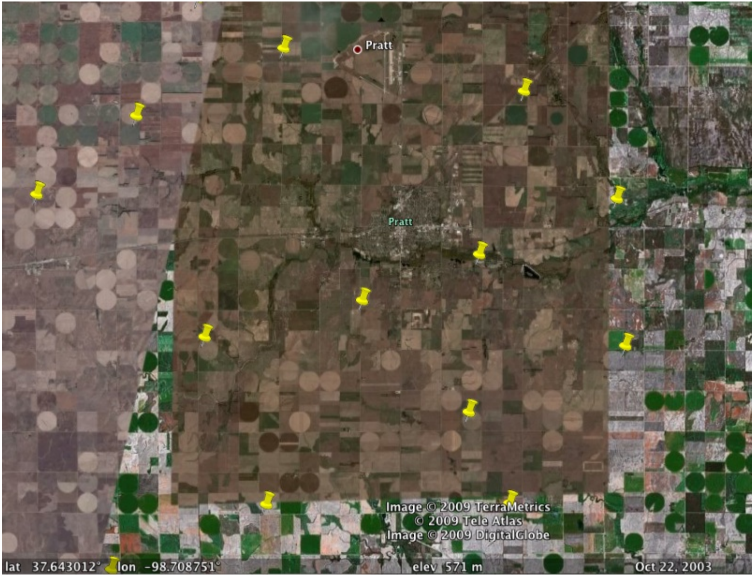
The observations



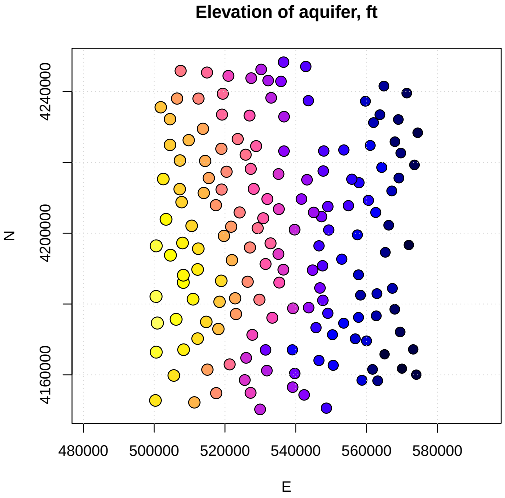
Model as: $Z(\mathbf{s}) = Z^*(\mathbf{s}) + \varepsilon(\mathbf{s}) + \varepsilon'(\mathbf{s})$

Olea, R. A., & Davis, J. C. (1999). *Sampling analysis and mapping of water levels in the High Plains aquifer of Kansas* (KGS Open File Report No. 1999-11). Lawrence, Kansas: Kansas Geological Survey. Retrieved from `http://www.kgs.ku.edu/Hydro/Levels/OFR99_11/`

Observation wells

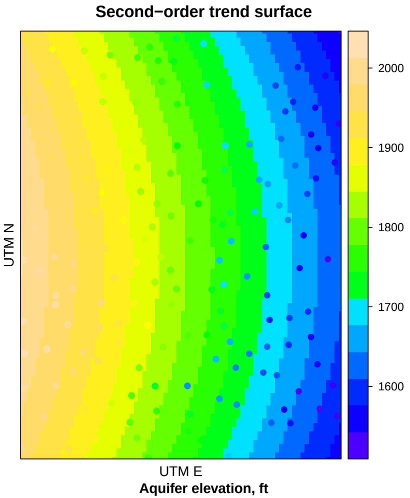


A study area



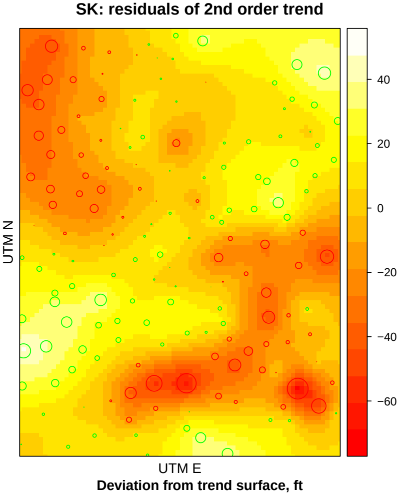
Model these observations $Z(\mathbf{s})$ by $Z^*(\mathbf{s})$, $\varepsilon(\mathbf{s})$, and $\varepsilon'(\mathbf{s})$?

A **deterministic** trend:



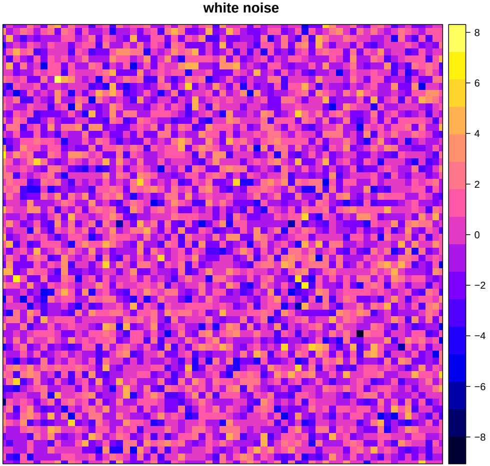
process: dipping and slightly deformed sandstone rock: $Z^*(\mathbf{s})$
modelled with a 2nd-order polynomial **trend surface**

A **spatially-correlated** random field



process: local variations from trend: $\varepsilon(\mathbf{s})$ (**model residuals**) **modelled** by **variogram modelling** of the random field and simple **kriging**

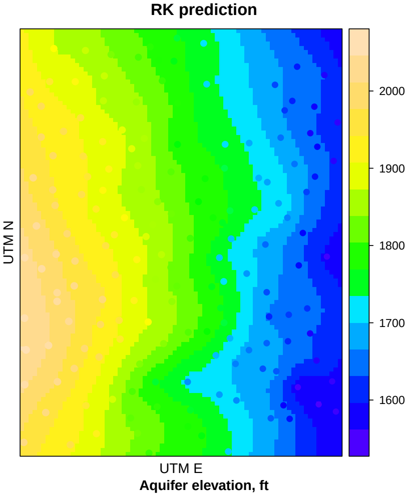
White noise: we do not know! but **assume** it looks like this:



$$\varepsilon'(\mathbf{s})$$

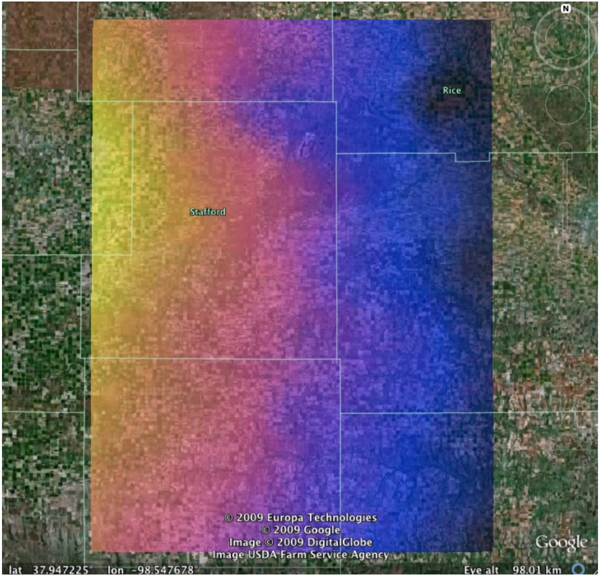
quantified as **uncertainty** of the other fits

Model with both trend and local variations



$Z^*(\mathbf{s}) + \varepsilon(\mathbf{s})$; prediction uncertainty $\varepsilon'(\mathbf{s})$

Predictions shown on the landscape

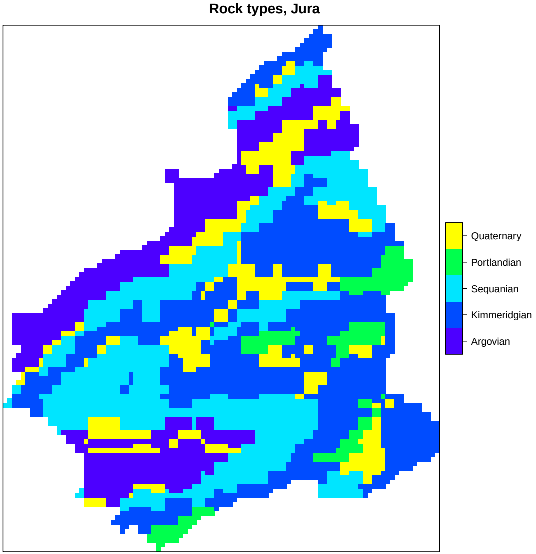


Mapping methods – stratification

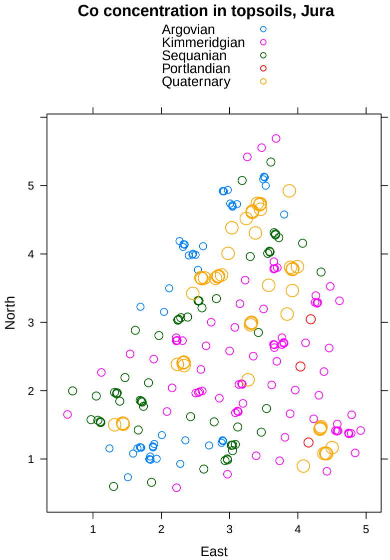
This models only $Z^*(\mathbf{s}) + \varepsilon'(\mathbf{s})$

- Divide the prediction area into **strata** based on objectives or pre-defined, e.g., political divisions
  - The stratum defines the deterministic $Z^*(\mathbf{s})$, each location $\mathbf{s}$ is in exactly one stratum
- Divide the point set, each point into its stratum
- Compute appropriate statistics per-stratum based on its points, e.g., mean, total, standard deviation . . .
  - The s.d. is one measure of $\varepsilon'(\mathbf{s})$
- Present as a polygon map

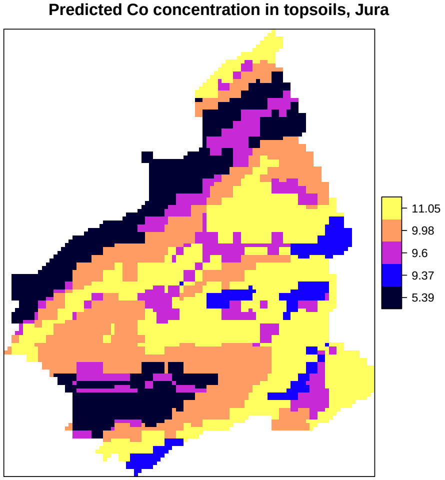
Strata


Rock types, Jura
Quaternary
Portlandian
Sequanian
Kimmeridgian
Argovian

Observations in strata



Prediction by strata



Can compute standard errors from the linear model (one-way ANOVA).

Mapping methods – global methods

These model only $Z^*(\mathbf{s}) + \varepsilon'(\mathbf{s})$

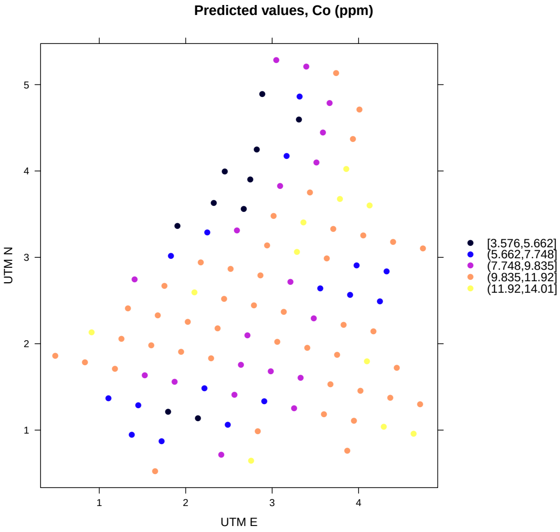
- **Trend surface**: one equation (linear model) using the coördinates of all the observations as predictors
  - The model can be used to map because the coördinates are also known at each prediction location
- **Multiple regression from covariates** one equation (linear model) using the attribute values of **environmental covariates** as predictors
  - These must be known at each prediction location, so covariate maps must cover the prediction area
- **Data-driven**: machine learning, e.g., random forests, using the atribute values of **environmental covariates** and/or coördinates as predictors

Mapping methods – local methods

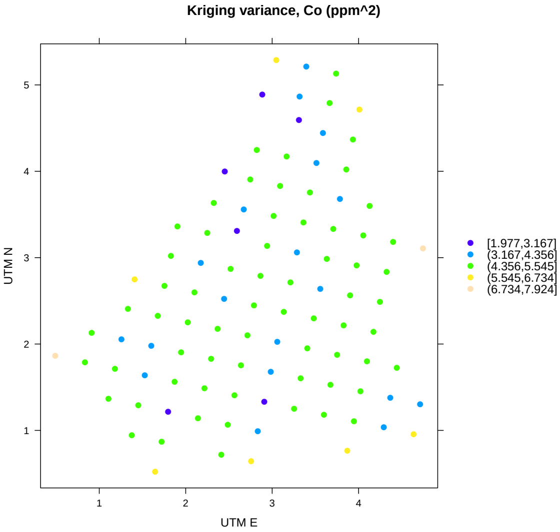
These model only $\varepsilon(\mathbf{s}) + \varepsilon'(\mathbf{s})$

- **model-based** (“geostatistical”) local interpolation, e.g., Ordinary Kriging
  - requires a **model** of local spatial correlation
- ***ad-hoc*** local interpolation, e.g., inverse distance
  - Note: **no theory**, just intuition
- **closest point**: Thiessen polygons

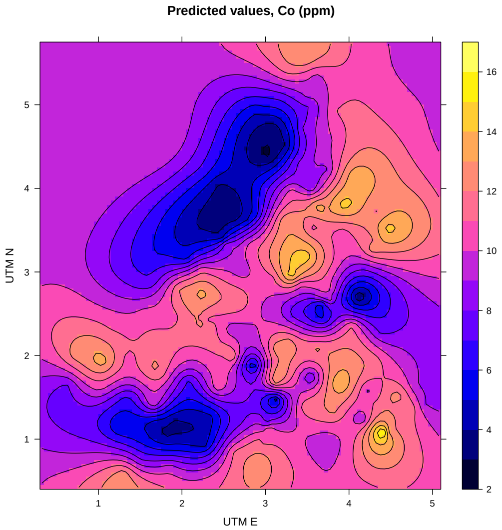
Prediction by Ordinary Kriging – at points



Prediction variances by Ordinary Kriging – at points

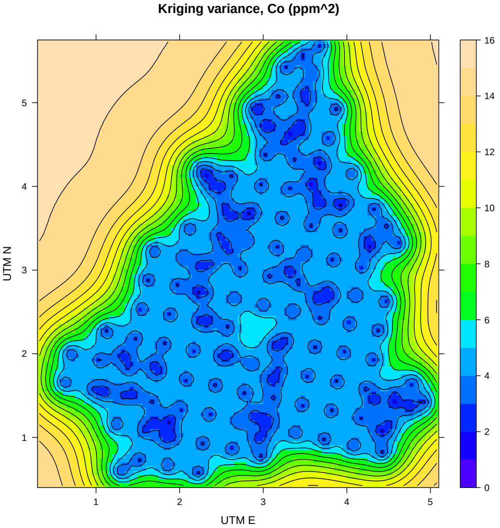


Prediction by Ordinary Kriging – at grid centres



Contours calculated after surface, for visualization

Prediction variances by Ordinary Kriging – at grid centres

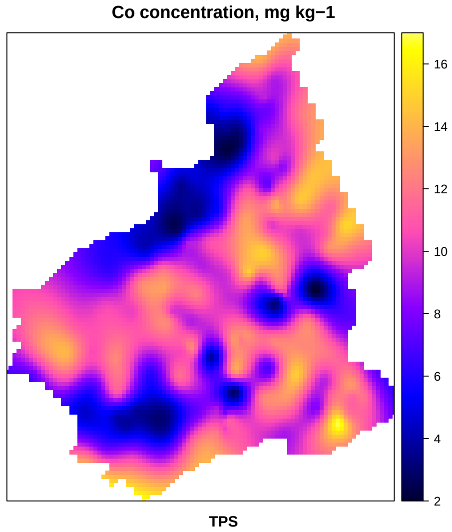


Mapping methods – locally-adaptive methods

These model only $Z^*(\mathbf{s}) + \varepsilon'(\mathbf{s})$ but use **locally-adaptive** functions for $Z^*(\mathbf{s})$

- **Thin-plate splines** (“minimum curvature”) warped surfaces (local fitting of a trend surface)
- **Geographically-weighted regression** (GWR): multiple regression from covariates, with locally-adapted coefficients
- **Generalized additive models** (GAM): like multivariate regression, but allow smooth functions of covariates as predictors

Prediction by thin plate splines – continuous surface

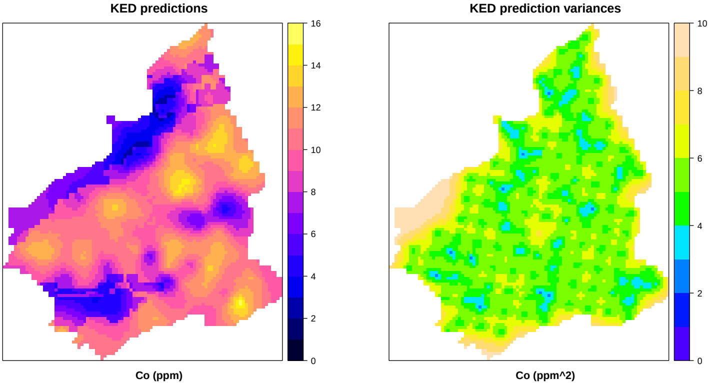


Mapping methods – mixed methods

These model $Z^*(\mathbf{s}) + \varepsilon'(\mathbf{s})$ first and then $\varepsilon(\mathbf{s}) + \varepsilon'(\mathbf{s})$ from the **residuals** of the global model

- **Regression Kriging** (RK) with any of the global predictors for $Z^*(\mathbf{s})$
- **Kriging with External Drift** (KED), one-step method of RK
- **Stratified Kriging** (StK): separate geostatistical model per stratum

Prediction by Kriging with External Drift, rock type as covariate



The pattern of rock types is modified by kriging the residuals from the linear model.

Choosing a mapping method

- Is **prediction** or **understanding** more important?
    - if **prediction**, may favour machine-learning or locally-adaptive methods
    - if **understanding**, may favour explicit models (local, global or mixed)
- What do you know or suspect about the spatial variability of the target attribute?
    - e.g., should there be local spatial dependence?
    - e.g., do we suspect a regional trend? of what form?
    - e.g., are there covariates related to the target variable? do we have maps of these?
- For prediction, try various methods and compare **evaluation statistics**

Which prediction method is "best"?

- There is **no theoretical answer.**
- It depends on how well the approach models the **'true' spatial structure**, which is unknown (but we may have **prior evidence**).
- The method should correspond with what we know about the **process** that created the spatial structure.
  - e.g., relation with environmental covariates or stratifying factor
- It should also be achievable with the **available data.**
  - e.g., for OK need "closely-"spaced observations, closer than the range of spatial dependence, to take advantage of local spatial structure $\varepsilon(\mathbf{s})$
  - e.g., for RF or MLR need observations covering the feature-space range

(continued . . .)

Which prediction method is “best”? (continued)

- Check against an independent **evaluation** (“validation”) dataset
  - **Mean squared error** (“precision”) of **prediction** vs. **actual** (residuals)
  - **Bias** (“accuracy”) of predicted vs. actual mean
- External vs. internal evaluation
  - With **large** datasets, model with one part and hold out the rest for **validation**
  - For **small** datasets use **cross-validation**
- How well it reproduces the **spatial variability** (pattern) of the calibration dataset
  - Difficult statistical problem

End