Mapping from point observations

D G Rossiter

Nanjing Normal University, Geographic Sciences Department
南京师范大学地理学学院

March 18, 2020

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**
  - example: pH, organic C etc. in soil sample from an auger: 4 cm diameter (2D) + 10 cm length (3D)
  - example: biomass from vegetation plot 10 x 10 m (2D)
  - example: temperature, precipitation, relative humidity at a weather station (“point” instrument but variable is the same over some radius)

What we want to know

- The value of these attributes at **other** (unvisited) “point” locations, either ...
  - with the same **support** as the original observations
  - with some other support, usually larger (“block”)
- Often we want to know at a **grid** of other “point” locations $\rightarrow$ 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
- Both of these require **spatial prediction** based on the observed “point” observations

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".
  - 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.**

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 will be 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**

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

Spatial autocorrelation – formula

Semivariance of **one point-pair**:

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) \equiv \frac{1}{2}[z(\mathbf{s}_i) - z(\mathbf{s}_j)]^2 \tag{2}$$

So the same attribute z is correlated with itself, but in another observation point.

Average semivariance over some **separation range h**:

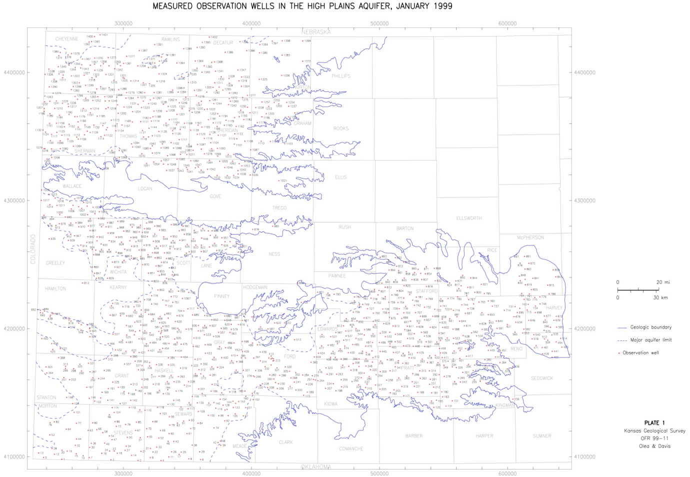
$$\overline{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} [z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h})]^2 \tag{3}$$

where $m($**h** is the number of point-pairs within the separation range.

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?

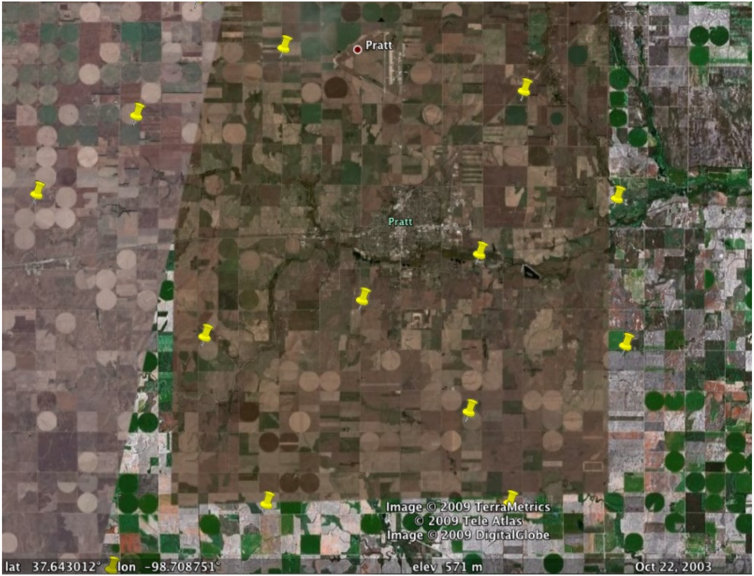
The observations



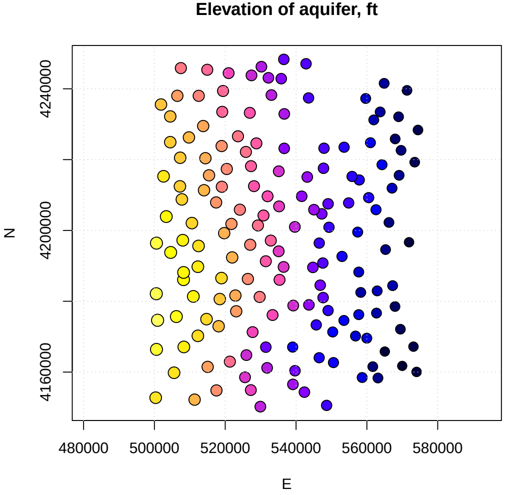
Model as: $Z(\mathbf{s}) = Z^*(\mathbf{s}) + \varepsilon(\mathbf{s}) + \varepsilon'(\mathbf{s})$

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

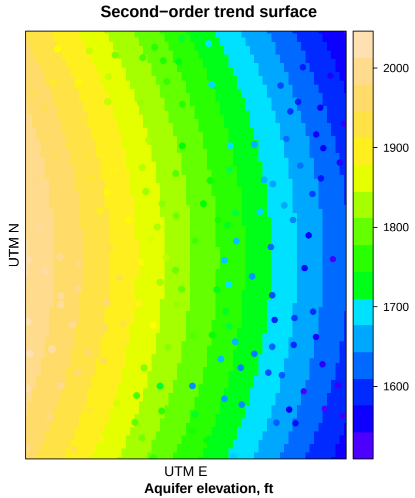


A study area



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

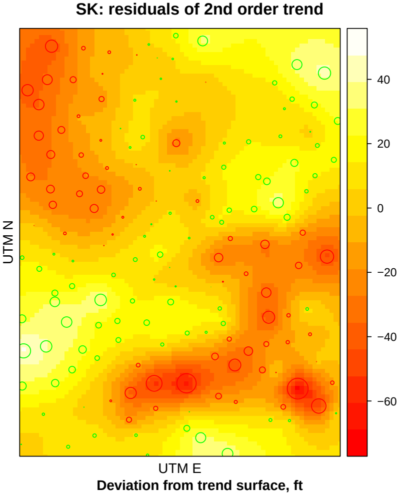
A **deterministic** trend:



process: dipping and slightly deformed sandstone rock: $Z^*(\mathbf{s})$

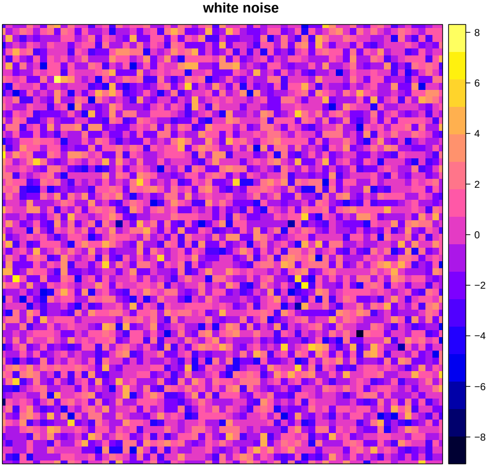
modelled with a 2nd-order polynomial **trend surface**

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**

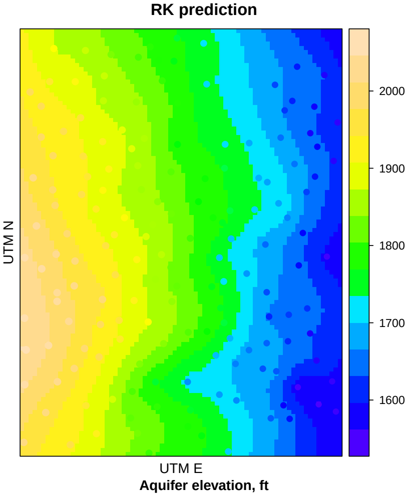
White noise: we do not know! but **assume** it looks like this:



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

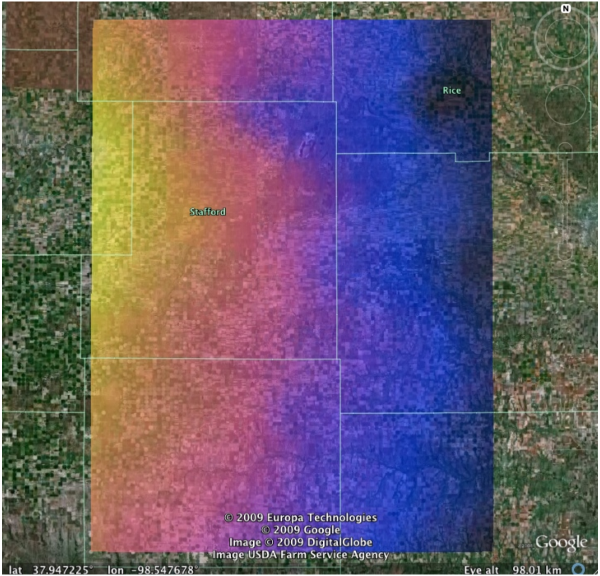
quantified as **uncertainty** of the other fits

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

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
- Cokriging (CK): as OK, but use the local spatial correlation of several variables together
- ***ad-hoc*** local interpolation, e.g., inverse distance
- **closest point**: Thiessen polygons

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

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

Choosing a mapping method

- Is **prediction** or **understanding** more important?
  - if prediction, may favour machine-learning or locally-adaptive methods
- What do you know or suspect about the spatial variability of the target attribute?
  - e.g., should there be local spatial dependence?
- For prediction, try various methods and compare **evaluation statistics** (see below)

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