Applied geostatistics

Lecture 4 – Spatial prediction from point samples (Part 1)

D G Rossiter
University of Twente.
Faculty of Geo-information Science & Earth Observation (ITC)

October 16, 2010

Copyright © 2007–2010 University of Twente, Faculty ITC.
All rights reserved. Reproduction and dissemination of the work as a whole (not parts) freely permitted if this original copyright notice is included. Sale or placement on a web site where payment must be made to access this document is strictly prohibited. To adapt or translate please contact the author (`http://www.itc.nl/personal/rossiter`).

ITC

Topics for this lecture

1. A taxonomy of spatial prediction methods

2. Non-geostatistical prediction

3. Introduction to Ordinary Kriging

   Note: the derivation of the kriging equations is deferred to the next lecture.

Commentary

Spatial prediction from point samples is one of the main practical applications of geostatistics – we know the value of some attribute at some **observation points**, but we need to know it over an entire area – i.e. we want to **map** it.

Prior to the introduction of sound geostatistical methods, **contour** maps were drawn by hand, using the intuition / local knowledge of the mapper. These maps are often beautiful, but how realistic are they? With geostatistical methods we have a firm basis for both **prediction** and **assessing the quality** of the result.

Topic 1: A taxonomy of spatial prediction methods

Objective: to **predict** of the value of some attribute at an **unsampled point** based on the values of that attribute at **sampled points**.

Prediction can be at:

- **Selected points** of particular interest;
- **All points on a grid**; the result is a **map** of the spatial field at the grid resolution

In both cases the predictions can be of:

- the **points** themselves, always with some specified **support** ;
- **average values** in **blocks** centred on points.

By the way . . .

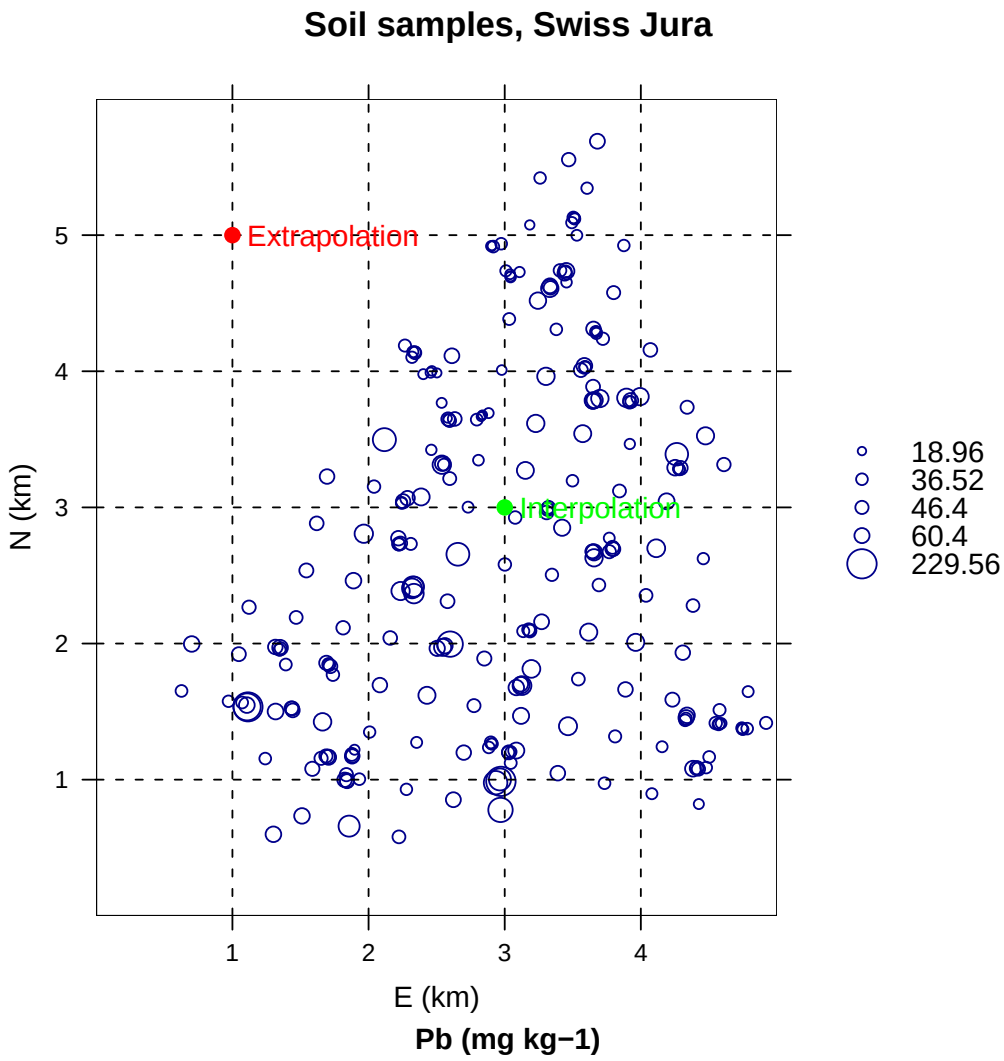
Sometimes it's enough to predict at some unknown **point** – we don't have to map an entire area. For example, consider the problem of a village that wants to deepen their village well to reach a more relaible groundwater supply. They only need to know the predicted depth of the groundwater table at that one point, not over the whole district – the village isn't about to move! So, we can use groundwater measurements at 'nearby' deep wells to predict the depth to which the village will have to dig.

Interpolation vs. Extrapolation

Spatial **prediction** is often referred to as **spatial interpolation**, but strictly speaking:

- **Interpolation**: prediction at points that are **geographically inside** the convex hull of the sample set;
- **Extrapolation**: prediction at points **outside** this geographic area.

Interpolation vs. Extrapolation from point samples



To check your understanding . . .

Q1 : *Suppose we have climate records for several stations in the western Dominican Republic (DR), but none that we can access for Haiti, adjacent on the same island of Hispaniola. Would it be* **interpolation** *or* **extrapolation** *to use the DR records to make a climate map of eastern Haiti, adjacent to the DR? Jump to A1* •

Q2 : *Would it be justified to use the DR records to map the climate of the easternmost 10 km of Haiti, immediately adjacent to the DR? Why or why not? Jump to A2* •

Q3 : *Would it be justified to use the DR records to map the climate of all of Haiti? Why or why not? Jump to A3* •

Commentary

So, we want to predict at unsampled locations. But how do we do this? There are many methods; the only thing they all have in common is that they use the available data in some way.

Before entering into a detailed description of the most common methods, we first **classify** them into a **taxonomy**, based on how they use the available data.

A taxomomy of spatial prediction methods

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

Global: predictors: use **all samples** to predict at **all points**; also called **regional** predictors;

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

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

We will see examples of all of these.

Commentary

The question that is always asked at this point is ...

Which method is best?

And the answer is, as for so many other things in the messy real world ...

It depends!

The key point is that we believe that there is some order in nature; there is some **reason** data values are as we observe them. We try to **model** this structure, then use this model to **predict**. If the model is correct, the prediction should be good.

Which prediction method is "best"?

- There is **no theoretical answer**
- Depends on how well the approach models the **'true' spatial structure**, and this is unknown (but we may have **prior evidence**)
- The method should correspond with what we know about the **process** that created the spatial structure

Which prediction method is "best"? (continued)

- Check against an independent **validation** dataset
  - **Mean squared error** ("precision") of **prediction** vs. **actual** (residuals)
  - **Bias** ("accuracy") of predicted vs. actual mean
- With large datasets, model with one part and hold out the rest for **validation**
- **Cross-validation** for small datasets with a modelled structure

We will see how to measure these later.

Commentary

We begin our exploration of prediction methods with a **non-geostatistical** approach: **prediction by stratification**. This take **no account of the location of the samples**, just their **classification** into strata as given in a **map legend**.

This model may be realistic and appropriate in some situations.

The **key assumption** with prediction by stratification is that there is **no spatial dependence**.

Approaches to prediction (1): Strata

Not really spatial analysis, since **spatial position is not used**, but it does predict in space.

- Example: Nutrient content in a field, since fields are treated as units in management

1. **Stratify** the landscape (e.g. by land use, geological formation . . .)
   - It is common to use an existing class map to identify the strata.
2. **Sample within strata** according to **non-spatial sampling theory**
3. **Analyze with non-spatial techniques**, e.g. ANOVA
4. **Each location** in stratum has the **same expected value and variance**, based on the sample from that stratum

By the way . . .

This is also called **design-based** prediction, which is opposed to geostatistical or **model-based** prediction, since there is **no model of spatial dependence**.

The **"design"** refers to the **probability sampling design** which is necessary to get correct inferences.

To check your understanding . . .

Q4 : *Give an example of a stratification in your application area. What attributes are expected to be related to the strata?*

Jump to A4 •

Commentary

Other approaches to spatial prediction do consider the **spatial location** of the sample and prediction points.

We begin with a prediction method that uses **all sample points** to calibrate a model of **regional trend**, which is then used to predict at unsampled points.

Approaches to prediction (2): Global (Regional) Predictors

- These are also called **trend surfaces**
- The **derivation of their mathematical form** was covered in a previous lecture; recall that the general **polynomial trend surface** of order p is:

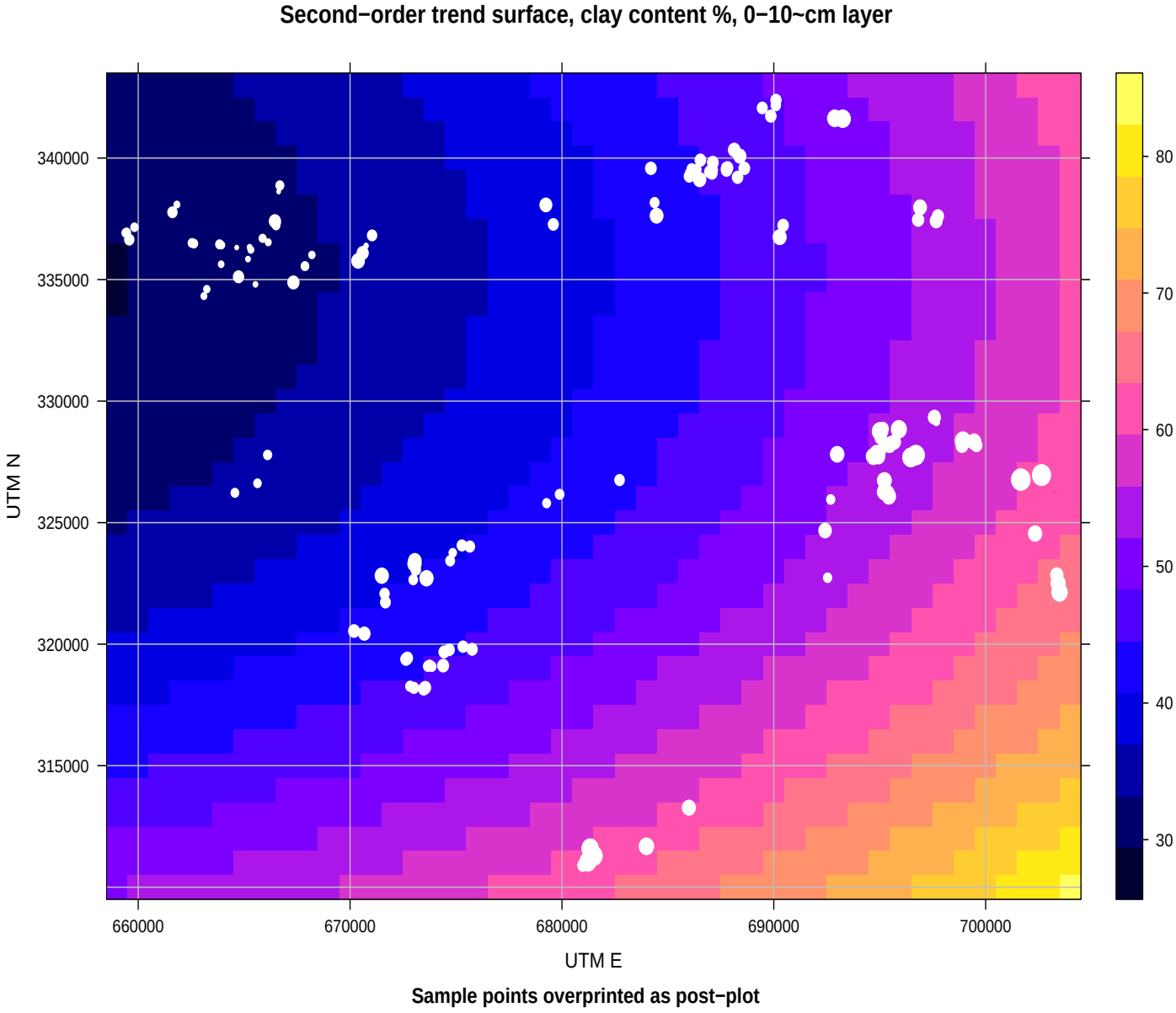
$$f(x,y) = \sum_{r+s \leq p} \beta_{r,s} x^r y^s$$

- The trend surface formula is a **function of the coördinates**; since every location as coördinates we can **predict at any and all locations** from the **formula**.
  - Example: $\text{clay35} = -0.0000251 - 0.651 \cdot \text{UTM_E} - 0.000045 \cdot \text{UTM_N}$
- That is, with any known x and y we can apply the formula to get z.
  - Example: $(\text{UTM_E} = 680000, \text{UTM_N} = 330000) \Longrightarrow \text{clay35} = 43.97\%$

To check your understanding . . .

Q5 : *Give an example of a regional trend in your application area. What attributes are expected to be related to the trend?* *Jump to A5* •

Prediction with a trend surface



White points are **observations**, everywhere else is **predictions**

Commentary

Strata are suitable to model processes that **depend on the stratifying classes**, and which have **no spatial structure**.

Trend surfaces are suitable to model **regional spatial processes**.

Another kind of process is **local**; that is, whatever is causing a certain attribute value at a location is also operating "nearby". We now investigate these.

Approaches to prediction (3): Local predictors

- **No strata**
- **No regional trend**
- Value of the attribute is predicted from **"nearby" samples**
  - Example: concentrations of soil constituents (e.g. salts, pollutants)
  - Example: vegetation density

To check your understanding . . .

Q6 : *Give an example of an attribute in your application area that you expect to have local spatial dependence.* *Jump to A6* •

Local predictors: Model-based or not?

- A predictor is called **model-based** or **geostatistical** if it requires a **model of spatial structure**.
  - The most common is some form of **kriging**; the geostatistical basis is the **variogram model**
- Otherwise it is based on untestable **assumptions** about spatial dependence
  - An example is inverse-distance weighted average.

We will study both of these in more detail in the next topic.

Commentary

We've seen stratified, regional and local **predictors**; these correspond to three classes of **processes**.

Of course, nature is never so simple! An attribute may owe its spatial distribution to a **combination of processes**; we then need a **mixed** predictor that somehow combines the predictor types.

Approaches to prediction (4): Mixed predictors

- For situations where there is **both** long-range structure (trend) or strata **and** local structure
  - Example: Particle size in the soil: strata (rock type), trend (distance from a river), **and** local variation in depositional or weathering processes
- One approach: model strata or global trend, subtract from each value, then model **residuals** → e.g. **Regression Kriging**.
- Another approach: model everything together → e.g. **Universal Kriging** or **Kriging with External Drift**

Topic 2: Non-geostatistical prediction

Before looking at so-called "optimal" weighting (⇒ **kriging**) we examine various **non-geostatistical** prediction methods.

These were widely-used before kriging was developed, and still are in some circumstances.

The **advantage** of these methods, compared to kriging, is that **no model of spatial dependence** is required; there is no need to compute or model variograms.

One **disadvantage** is that there is no theory behind them, only assumptions.

The **major disadvantage** is that they are often based on **invalid assumptions**, in particular **spatial independence** of the samples. So, the prediction may be **incorrect** even in the expected value.

Non-geostatistical stratified predictors

This was explained above; recall:

1. **Stratify** the landscape into "homogeneous" units; this is often on the basis of an existing class map;

2. **Sample within strata** according to **non-spatial sampling theory**; so each observation is identified with one stratum;

3. **Each location to be predicted** is in some stratum; it has the **same expected value and variance**, based on the **observations from that stratum**

4. No information from any other stratum is used, except that the variance may be pooled.

5. **The geographic locations** of the **prediction** and **observation** points are **irrelevant**.

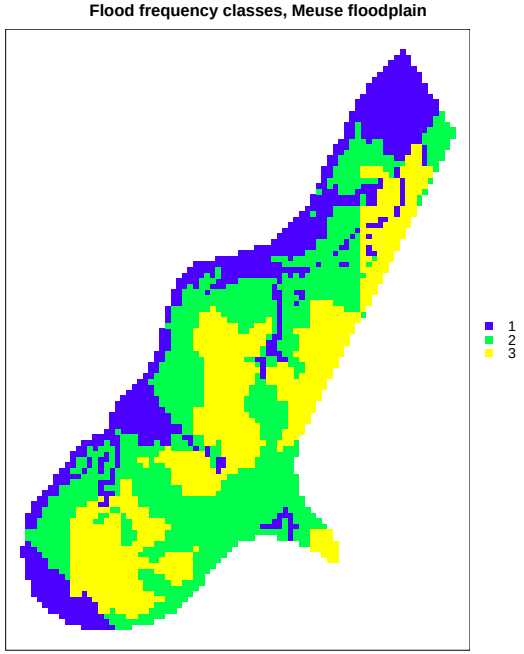
Commentary

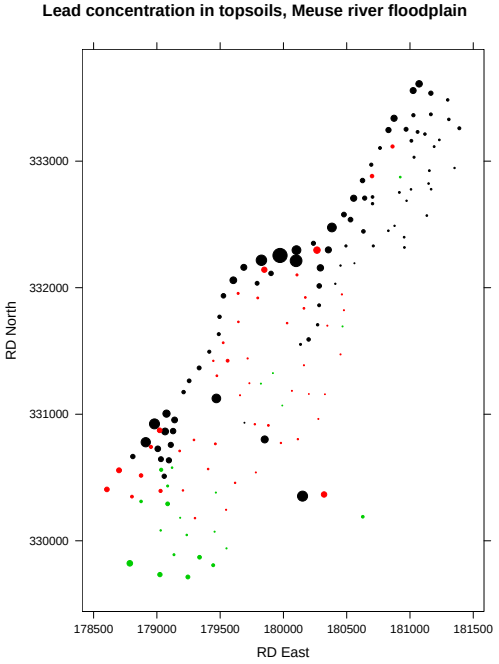
The following page shows a **stratification** of the Meuse floodplain by the three flood frequency classes, and then the **predicted value** at each point, based on the observations from that class:

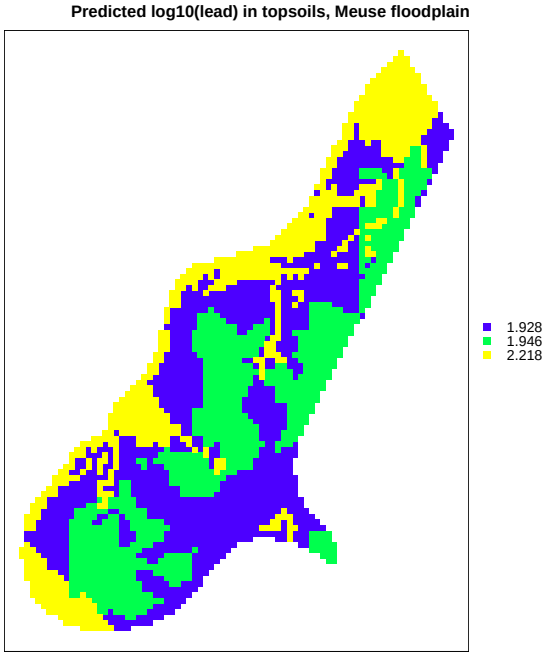
Class	N	Mean
"1"	84	2.218
"2"	48	1.983
"3"	23	1.946

Note that there is **no variability** of the predictions **within a stratum**. This is the best we can do with **design-based** methods.

Also, there is **no spatial dependence**; the computed means and variances assume this. This assumption is rarely met! which is why this method is rarely valid.


Flood frequency classes, Meuse floodplain
1
2
3


Lead concentration in topsoils, Meuse river floodplain
RD North
333000
332000
331000
330000
178500
179000
179500
180000
180500
181000
181500
RD East


Predicted log10(lead) in topsoils, Meuse floodplain
1.928
1.946
2.218

Non-geostatistical Local Predictors

- Nearest neighbour (Thiessen polygons)
- Average within a radius
- Average of the n nearest neighbours
- Distance-weighted average within a radius
- Distance-weighted average of n nearest neighbours

These all have an **implicit model of spatial structure**; these are **assumptions** which can not be tested.

Local predictor (1): Nearest neighbour (Thiessen polygons)

- also known as a **Voronoi mosaic**, computed by a **Delaunay triangulation**
- Predict each point from its **single nearest sample point**
- **Assumption**: process is the same within each polygon and changes abruptly at the borders
- Conceptually-simple, makes the minimal assumptions about spatial structure
- No way to estimate variance of the prediction error
- Ignores other 'nearby' information
- Maps show abrupt discontinuities at boundaries, so don't look very realistic
- But may be a more accurate predictor than poorly-modelled predictors

Commentary

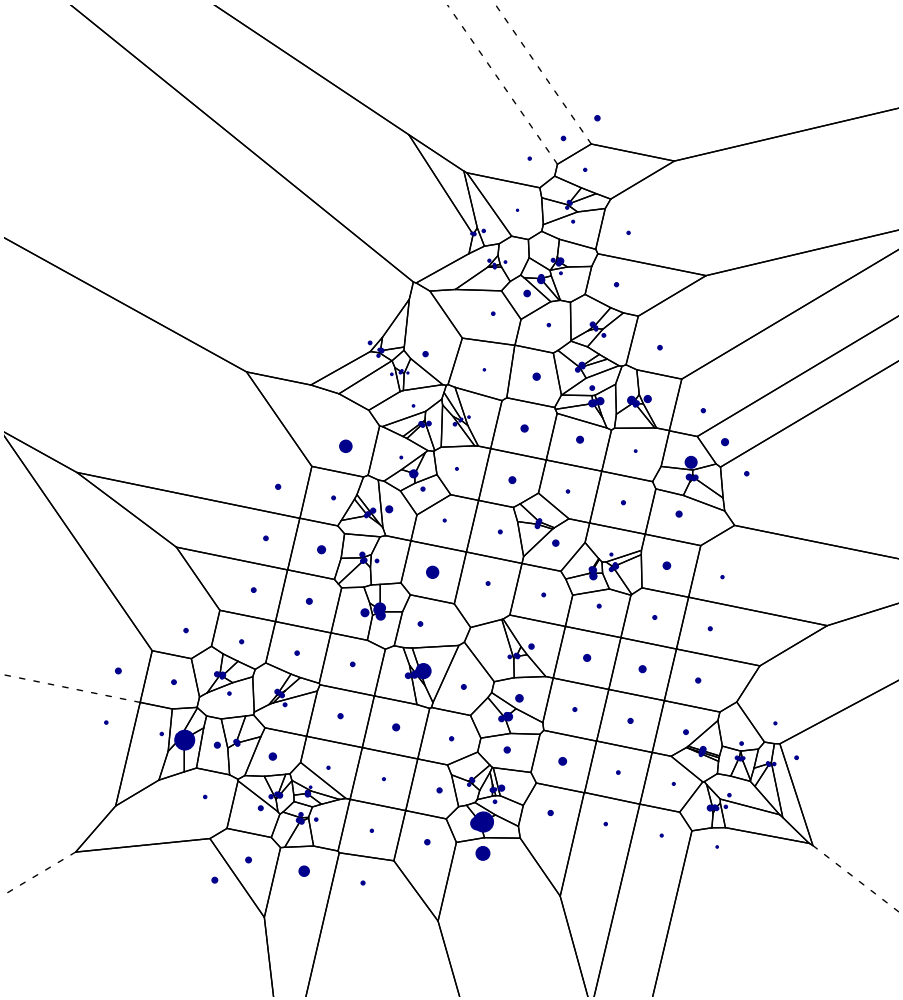
The following figure shows the Thiessen polygons for the Jura soil sample data set (259 calibration points).

Each point within a polygon is predicted by the **value of the nearest point**, i.e. the point within the polygon. These are shown as a **postplot** proportional to the lead content.

(Figure produced with the `tripack` package of the R environment for statistical computing.)

Nearest-neighbours

Thiessen polygons (Voronoi mosaic)



Jura soil samples (blue points)

Local predictor (2): Average within a radius

- Use the set of all neighbouring sample points within some radius r
- Predict by averaging :

$$\hat{\mathbf{x}}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \qquad d(\mathbf{x}_0, \mathbf{x}_i) \leq r$$

- **Assumption**: process is the same within the circle, but there is random variation due to a noisy process
  - This can't be true of overlapping circles!
- Although we can calculate variances from the neighbours, these assume no spatial structure closer than the radius
- Problem: How do we select a radius?

Local predictors (3): Distance-weighted average

- **Inverse of distance** of the point to be predicted to some set of n nearest-neighbours, to some **power** $k = 1, 2 \ldots$

$$\hat{\mathbf{x}_0} = \sum_{i=1}^{n} \frac{\mathbf{x}_i}{d(\mathbf{x}_0, \mathbf{x}_i)^k} / \sum_{i=1}^{k} \frac{1}{d(\mathbf{x}_0, \mathbf{x}_i)^k}$$

- $k = 1$: "inverse distance", $k = 2$: "inverse distance squared", etc.

- **Assumption**: process is a **power model** where the spatial correlation depends inversely on distance
  - This is like kriging with a power variogram model – except the spatial dependence **among** the neighbours (known points) is **not** accounted for!

- Can select all points within some limiting distance (radius), or some fixed number of nearest points, or . . . so, how to select radius or number and power objectively?

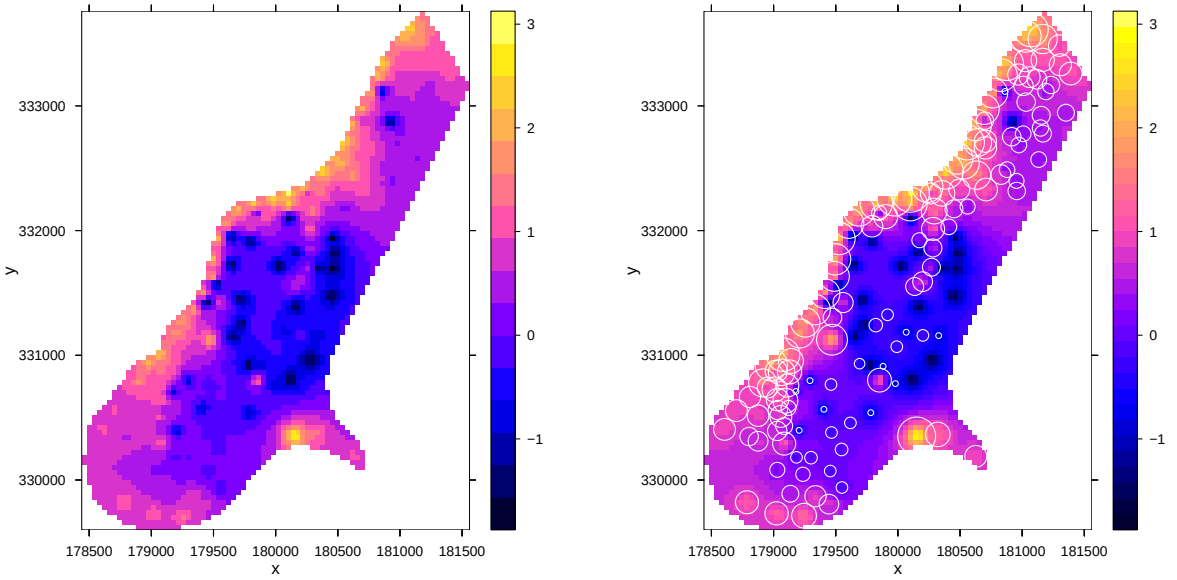
Inverse distance vs. Ordinary Kriging

In the following slide we compare inverse distance (linear) to Ordinary Kriging (OK) with a spherical model (range = 1150 m), to predict the base-10 log Cd concentration in soils in the Meuse river floodplain in the southern NL.

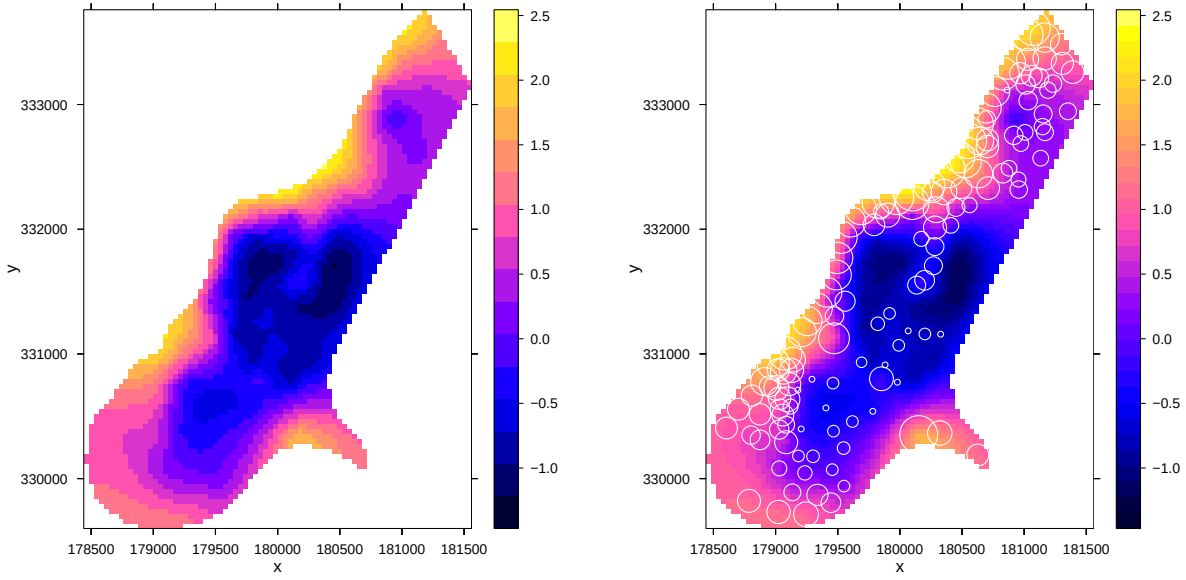
Notice:

- OK gives a smoother map;
- Inverse distance shows small "islands" or "spots"; the size of these is controlled by the **power** to which the inverse distance is raised.
- The "spots" are controlled by the observation points.

Inverse distance



Ordinary kriging



Exercise

At this point you should do the **first two sections** of **Exercise 4: Predicting from point samples (Part 1)** which is provided on the module CD:

- §2 **Trend surfaces**
- §3 **Design-based prediction**

These are short exercises and should take less than an hour.

As in all exercises there are **Tasks**, followed by R code on how to complete the task, then some **Questions** to test your understanding, and at the end of each section the **Answers**. Make sure you understand all of these.

Topic 3: Ordinary kriging

The theory of regionalised variables leads to an **"optimal"** prediction method, in the sense that the **kriging variance** is **minimized**.

This is based on the **theory of random fields** which was presented in a previous lecture.

Optimal local interpolation: motivation

- Problems with **Theissen polygons**:
  1. Abrupt changes at boundaries are an artifact of the sample spatial distribution
  2. Only uses one sample point for each prediction; inefficient use of information
- Problems with **average-in-circle** methods:
  1. No objective way to select radius of circle or number of points
  2. Obviously false underlying assumption
- Problems with **inverse-distance** methods:
  1. How to choose power (inverse, inverse squared ...)?
  2. How to choose limiting radius?
- In all cases:
  1. **uneven distribution of samples**: over– or under–emphasize some sample areas
  2. **kriging variance** must be estimated from a separate validation dataset

Commentary

These deficiencies in existing local interpolations were well-known.

The aim was to develop a **linear predictor** as a **weighted average** of the observations, with an objectively **optimal** method of assigning the weights.

The theory for this developed several times (Kolmogorov 1930's, Wiener 1949) but current practise dates back to Matheron (1963), formalizing the practical work of the mining engineer D G **Krige** (RSA).

In Krige's honour these methods are called **kriging** (now with a small "k"); it should really be written as "krigeing" (French **krigeage**) but it's too late for that.

Introduction to Ordinary Kriging (OK)

1. In what sense is OK "optimal"?
2. Derivation of the OK system of equations
3. Interpolation by kriging

An "optimal" local predictor would have these features:

- Prediction is made as a **linear** combination of known data values (a **weighted average**).
- Prediction is **unbiased** and **exact at known points**
- Points **closer** to the point to be predicted have **larger weights**
- **Clusters** of points **"reduce to"** single equivalent points, i.e., over-sampling in a small area can't bias result
- Closer sample points **"mask"** further ones in the same direction
- Error estimate is based only on the **spatial configuration of the sample**, not the data values
- The **kriging variance** should be **as small as possible**.

Kriging

- A **"Best Linear Unbiased Predictor"** (BLUP) that satisfies a certain **optimality criterion** (so it's "best" with respect to the criterion)
- It is only "optimal" with respect to the **chosen model** and the chosen **optimality criterion**
- Based on the **theory of random processes**, with **covariances depending only on separation** (i.e. a variogram model)

What is so special about kriging?

- Predicts at any point as the **weighted average** of the values at sampled points
  - as for inverse distance (to a power)
- Weights given to each sample point are **optimal**, given the **spatial covariance structure** as revealed by the **variogram model** (in this sense it is "best")
  - Spatial structure **between known points**, as well as **between known points and each prediction point**, is accounted for.
  - So, the prediction is only as good as the model of spatial structure.
- The **kriging variance** at each point is automatically generated as part of the process of computing the weights.
  - because this variance is used as an optimality criterion, it must be computed during the kriging process, and can be saved along with the BLUP.

How do we use Kriging in practice?

1. **Sample**, preferably at different resolutions

2. **Calculate** the **experimental variogram**

3. **Model** the variogram with one or more **authorized functions**

4. **Apply** the **kriging system of equations**, with the variogram model of spatial dependence, at each point to be predicted

   - Predictions are often at each point on a **regular grid** (e.g. a raster map)
   - These 'points' are actually blocks the size of the sampling support
   - Can also predict in **blocks** larger than the original support

5. As part of the solution of the kriging system, calculate the **variance** of each prediction; this is based only on the **sample point locations**, *not* their data values.

6. **Display maps** of both the predictions and their variances.

Commentary

Kriging makes strong **assumptions** about the process that produced the attribute values; namely, a theory of **random fields**, which was discussed in a previous lecture.

Each variety of kriging has different assumptions, but they all require a **spatially-correlated random field** that can be modelled by a variogram.

Major differences with inverse-distance weighted prediction are:

1. the model can be estimated by variogram analysis and thus is semi-objective, based on the evidence of the samples;
2. the **inter-relation** between sample points is modelled.

Prediction with Ordinary Kriging (OK)

The most common form of kriging is usually called "Ordinary". In OK, we model the value of variable z at location $\mathbf{x}_i$ as the sum $z(\mathbf{x}_i) = m + e(\mathbf{x}_i)$ of:

1. a **regional mean** m and

2. a **spatially-correlated random component** $e(\mathbf{x}_i)$

The **regional mean** m is estimated from the sample, but not as the simple average, because there is spatial dependence. It is **implicit** in the OK system. This mean is **constant** across the field, i.e. the **expected value** is the same and **unknown**; this is the **"Ordinary"** situation.

The **spatially-correlated random component** $e(\mathbf{x}_i)$ is estimated from the **spatial covariance structure** as revealed by the **variogram model**.

Ordinary Kriging (OK)

- The estimated value $\hat{z}$ at a point $\mathbf{x}_0$ is predicted as the **weighted average** of the values at *all* sample points $\mathbf{x}_i$:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i)$$

- The weights λ_i assigned to the sample points **sum to 1**:

$$\sum_{i=1}^{N} \lambda_i = 1$$

- Therefore, the prediction is **unbiased**:

$$E[\hat{z}(\mathbf{x}_0) - z(\mathbf{x}_0)] = 0$$

What makes it "Ordinary" Kriging?

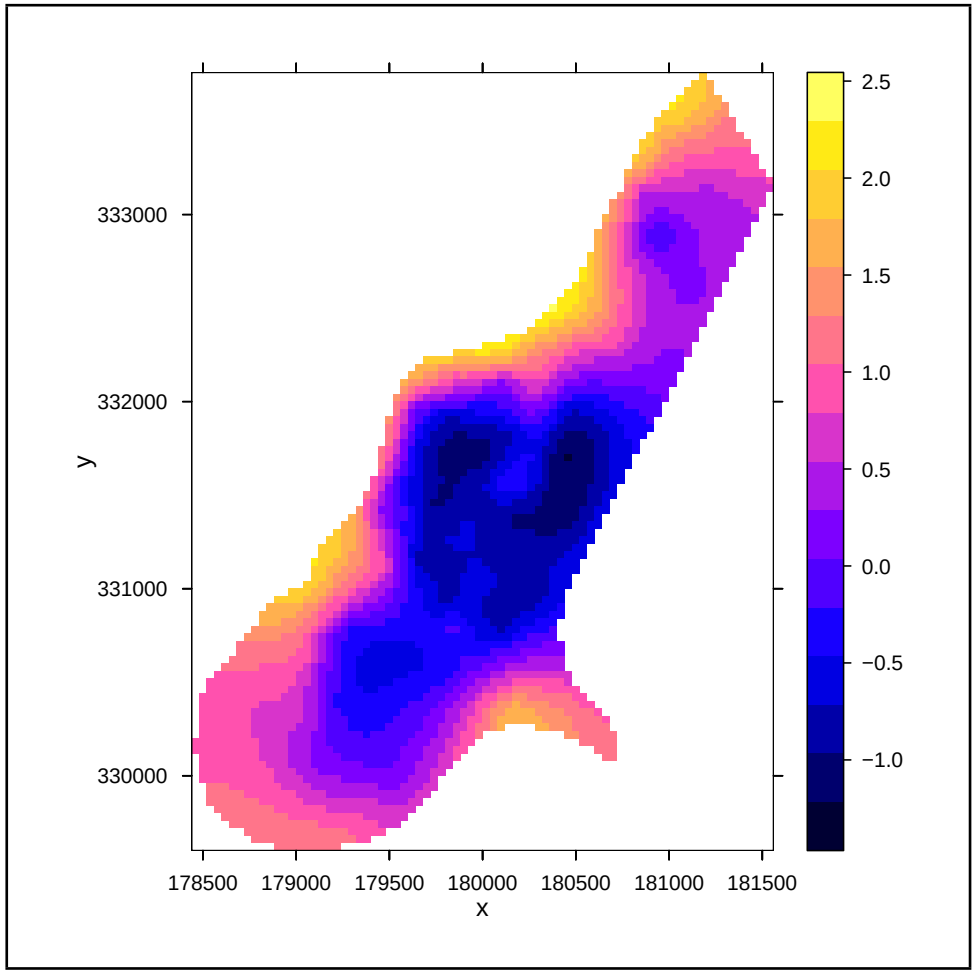
- The expected value (mean) is **unknown**, and must be estimated from the sample
  - If the mean is **known** we have **Simple Kriging** (SK)
  - We will see this in Regression Kriging (known mean of residuals is zero)
- There is **no regional trend**
  - If so we use **Universal Kriging** (UK)
- There is **no feature-space predictor**, i.e. another attribute that helps explain the attribute of interest
  - If so we use **Kriging with External Drift** (KED) or **Regression Kriging** (RK)

Commentary

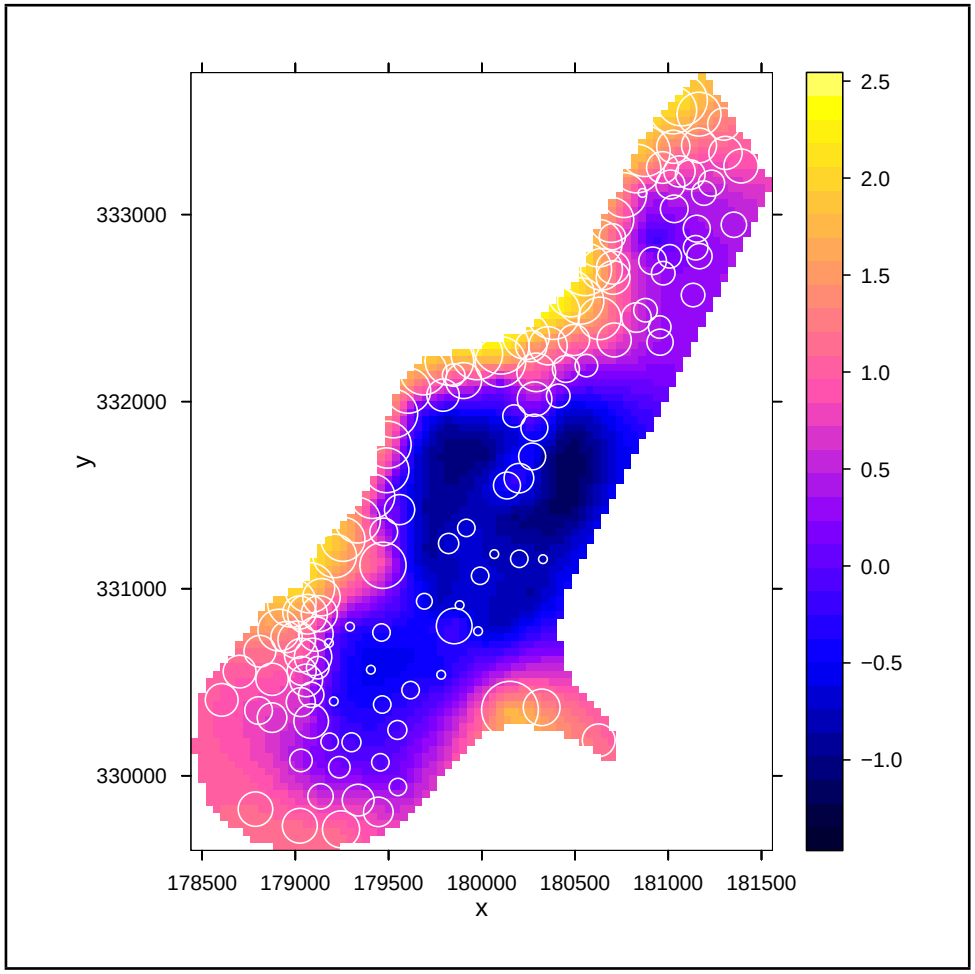
We defer the **derivation of the OK variance**, and from that the **kriging equations**, to the next lecture.

The important point here is that the **kriging equations minimize the kriging variance** at each point to be predicted, so that OK is in that sense **optimal**, of course if the **variogram model is correct**.

Ordinary kriging (OK) predictions for Meuse log(Cd)

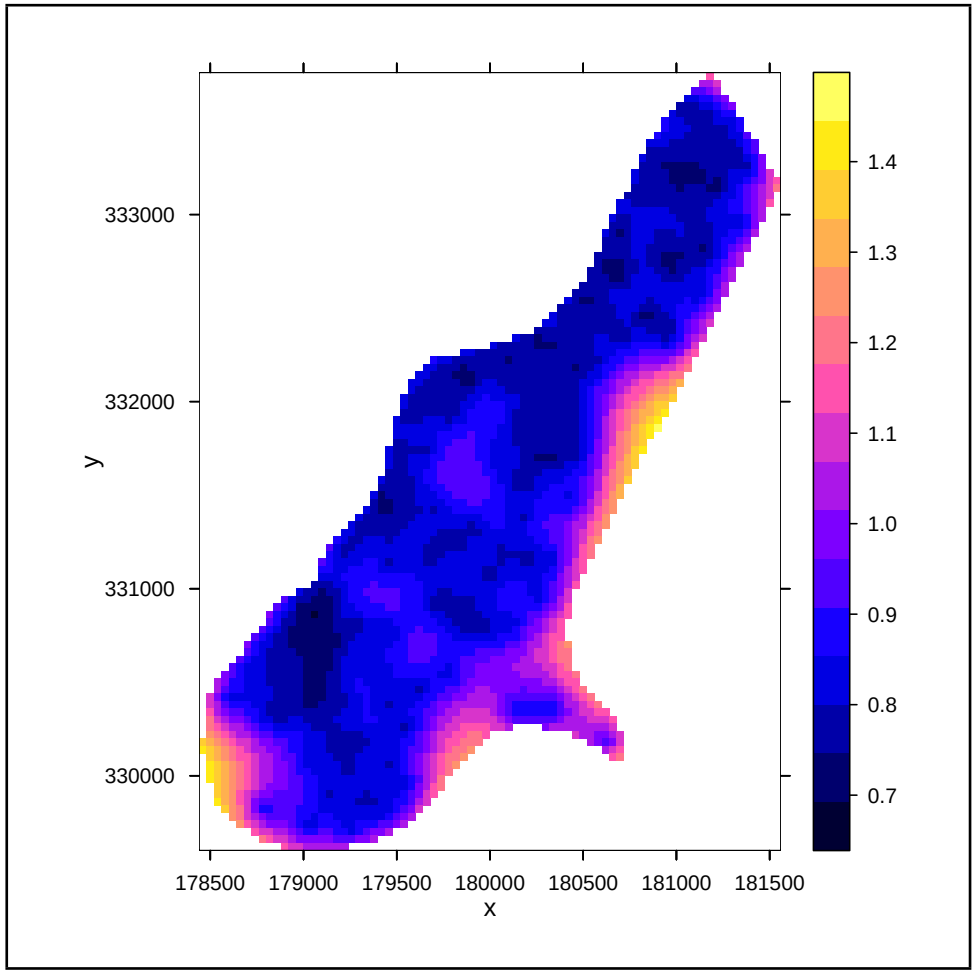


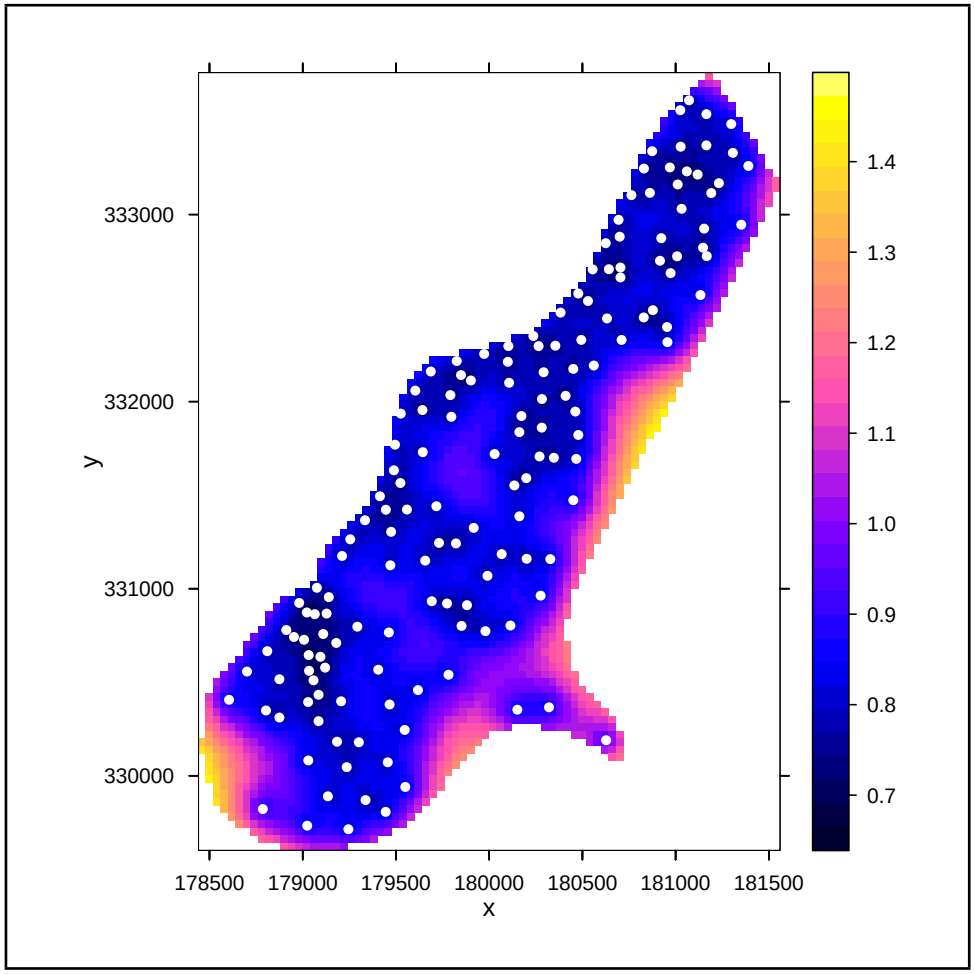
Predictions



With postplot superimposed

Variance of the OK prediction for Meuse log(Cd)





Variance

With sample points superimposed

Note that the variance depends only on the **configuation of the sample points**

The variance does **not** depend on the **data values**!

Use of the kriging variance

One of the major advantages of kriging is that it produces both a **prediction** and its **variance**. This can be used to:

- construct **confidence intervals** around the predicted value, and to
- compute the **probability** of **exceeding** any given **threshold**

These are particularly useful in **risk assessment**.

Confidence intervals

The **two-sided interval** which has **probability** $(1-\alpha)$ of containing the **true value** z is:

$$(\hat{z} - t_{\alpha/2,n-1} \cdot s_{\hat{z}}) \leq z \leq (\hat{z} + t_{\alpha/2,n-1} \cdot s_{\hat{z}})$$

where:

- $\hat{z}$ is the **estimated value** from OK;
- $t_{\alpha/2,n-1}$ is Student's t with $n-1$ degrees of freedom at **confidence level** $\alpha/2$;
- $s_{\hat{z}}$ is the sample standard deviation, i.e. the **square root of the kriging variance** from OK;
- n is the sample size.

How realistic are maps made by Ordinary Kriging?

- The resulting surface is **smooth** and shows **no noise**, no matter if there is a nugget effect in the variogram model
- So the field is the **best at each point taken separately**, but taken as a whole is **not a realistic map**
  - See topic 5 "Spatial simulation" in lecture 6
- The **sample points are predicted exactly**; the observations are assumed to be without error, again even if there is a nugget effect in the variogram model
  - Predicting at a grid point near to, but not exactly identical to, a sample point, will indeed result in smoothing and a positive kriging variance.
  - Block kriging does not have this problem, even if the block is centred on a sample point.

OK in a local neighbourhood

- In practice, **the nearest few points** contribute most of the **weight** ...
- ...so we can set up the kriging system **locally** with only a few points; then the solution is rapid.
- Furthermore, this allows a **local 1st-order stationarity** rather than a **global** one; a much weaker assumption
- Note that the same **covariance structure** (i.e. variogram) is used, so we still assume **global 2nd-order stationarity**.

This is advocated by Goovaerts:

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, Oxford and New York.

Implementing OK in a local neighbourhood

- With modern computers there is no problem with fairly large kriging systems (several 100's of points)
- But we want to avoid giving negative weights to distant points
- Rule of thumb: use points out to the variogram **range**.
- But use a sufficient **number of points**.

Commentary

This concludes the taxonomy of spatial prediction methods. In the next lecture we will see:

1. how the kriging equations are derived from optimality conditions, and
2. mixed predictors that use kriging for residual spatial dependence after accounting for a trend or feature-space predictor

Exercise

At this point you should do the **last sections** of **Exercise 4: Predicting from point samples (Part 1)** which is provided on the module CD:

- §4 **Ordinary kriging**

This should take about an hour.

As in all exercises there are **Tasks**, followed by R code on how to complete the task, then some **Questions** to test your understanding, and at the end of each section the **Answers**. Make sure you understand all of these.

Then do the **self-test** at the end of Exercise 4.

Answers

Q1 : *Suppose we have climate records for several stations in the western Dominican Republic (DR), but none that we can access for Haiti, adjacent on the same island of Hispaniola. Would it be* **interpolation** *or* **extrapolation** *to use the DR records to make a climate map of eastern Haiti, adjacent to the DR?* •

A1 : *This is extrapolation, because we don't have any points in the area to be predicted.* *Return to Q1* •

Q2 : *Would it be justified to use the DR records to map the climate of the easternmost 10 km of Haiti, immediately adjacent to the DR? Why or why not?* •

A2 : *Yes, because we expect that climate does not change much in 10 km. However this is not true if we reach a radically-different climate zone because of topographic factors.* *Return to Q2* •

Q3 : *Would it be justified to use the DR records to map the climate of all of Haiti? Why or why not?* •

A3 : *No, because we don't expect climate to be consistent over 100's of km.* *Return to Q3* •

Answers

Q4 : *Give an example of a stratification in your application area. What attributes are expected to be related to the strata?* •

A4 : *(Depends on application). An example from soil survey: a stratification by landscape position along a hillslope (summit, shoulder, backslope, footslope, toeslope) may be related the attribute "soil depth", since the stable positions (summit) and positions with accumulation from erosion (toeslope) should have deeper soils.*

Return to Q4 •

Answers

Q5 : *Give an example of a regional trend in your application area. What attributes are expected to be related to the trend?* •

A5 : *(Depends on application). An example from soil survey: the trend from E to W and N to S in the Great Plains of the USA and Canada; soil organic carbon (SOC) decreases along this gradient in both directions: E to W because of decreasing rainfall (less vegetative matter to contribute to the SOC), N to S because of increasing temperatures (faster decomposition of organic matter).* *Return to Q5* •

Answers

Q6 : *Give an example of an attribute in your application area that you expect to have local spatial dependence.* •

A6 : *(Depends on application). An example from soil survey: most soil physical and chemical properties.*

Return to Q6 •