Lecture 3

Spatial prediction

1. Inferential questions
2. Geostatistical prediction
3. The connection to Kriging
4. A universal algorithm
5. Lead pollution in Galicia, northern Spain
6. Onchocerciasis in Liberia
7. Design

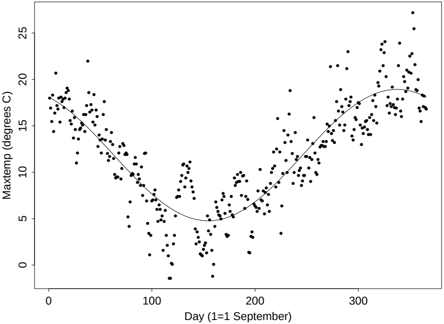
Inferential questions

- Testing: to what extent do our data support a pre-specified hypothesis about the process that generated the data?
- Estimation: what can our data tell us about particular properties of the process that generated the data?
- Prediction: what can our data tell us about particular properties of the realisation of the process that generated the data

A digression into time series revisited

Maximum daily temperatures, September 1995 to August 1996

Data



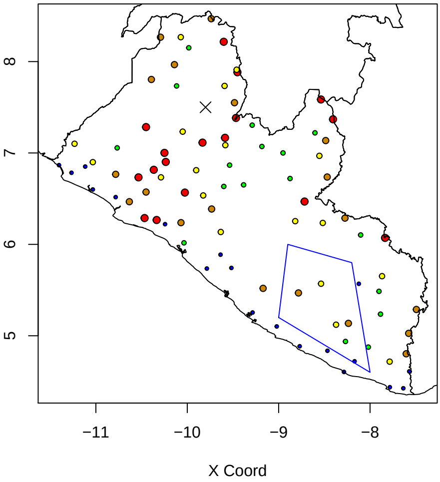
Statistical model

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

- Test: the average temperature range over a year is 15 degrees
- Estimate: what is the average temperate range over a year?
- Predict: what was the temperature range over the whole of 1996?

Information to answer these questions comes from both the data and the statistical model

Geostatistical prediction: onchocerciasis in Liberia



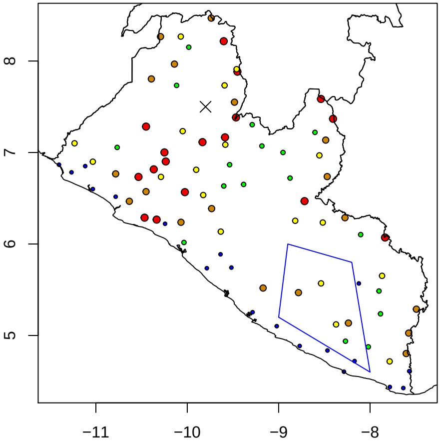
Predictive targets

1. prevalence at the marked location X
2. average prevalence over the region delineated in blue
3. does prevalence at the marked location X exceed 0.2 (20%)?
4. anything you like

Prediction without covariates

Signal $S(x)$ Data $(Y_i, x_i) : i = 1, 2, ...$

Model $S(x) \sim$ Gaussian process, $Y_i | S \sim \mathrm{N}(S(x_i), \tau^2)$



Predictive targets

1. $T = S(x)$

$$\hat{S}(x) = \sum w(x - x_i) Y_i$$

2. $T = \int_A S(x) dx$

$$\hat{T} = \int_A \hat{S}(x) dx$$

3. $T = I(S(x) > 0.2)$

$$\hat{T} = I(\hat{S}(x) > 0.2)?$$

No!

Minimum mean square error prediction

Model

- $[S^*]$ = probability distribution of underlying spatial process
- $[Y|S^*]$ = probability distribution of data conditional on underlying spatial process
- Bayes' theorem: $[S^*|Y] = [S^*][Y|S^*]/[Y]$

Mean square error

- $\hat{T} = t(Y)$ is a point predictor
- $\text{MSE}(\hat{T}) = \mathrm{E}[(\hat{T} - T)^2]$ is the mean square error

Theorem

1. $MSE(\hat{T})$ takes its minimum value when $\hat{T} = \mathrm{E}(T|Y)$.
2. $\mathrm{Var}(T|Y)$ estimates the achieved mean square error

Simple and ordinary Kriging

$$Y \sim \mathrm{MVN}(\mu 1, \sigma^2 V) \qquad V = R + (\tau^2/\sigma^2) \qquad R_{ij} = \rho(\|x_i - x_j\|)$$

Target for prediction: $T = S(x) \qquad [Y|S^*] \sim \mathrm{N}(S(x), \tau^2)$

Write $r = (r_1, ..., r_n)$ where $r_i = \rho(\|x - x_i\|)$

Standard results on multivariate Normal then give $[T|Y]$ as univariate Normal with mean and variance

$$\hat{T} = \mu + r' V^{-1}(Y - \mu 1)$$

$$\mathrm{Var}(T|Y) = \sigma^2(1 - r' V^{-1} r)$$

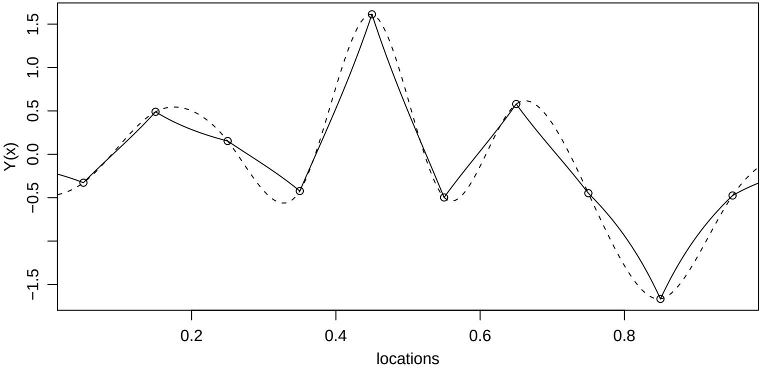
Simple Kriging: $\hat{\mu} = \bar{Y}$

Ordinary Kriging: $\hat{\mu} = (1' V^{-1} 1)^{-1} 1' V^{-1} Y$

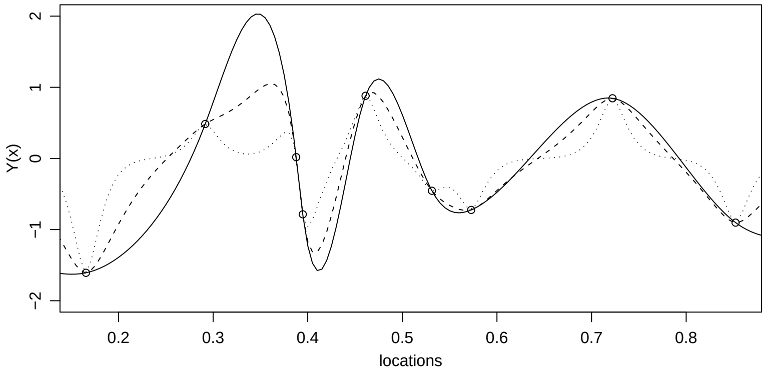
Note In both cases, $\hat{T}$ is a linear combination of the outcome data Y

Simple Kriging: three examples

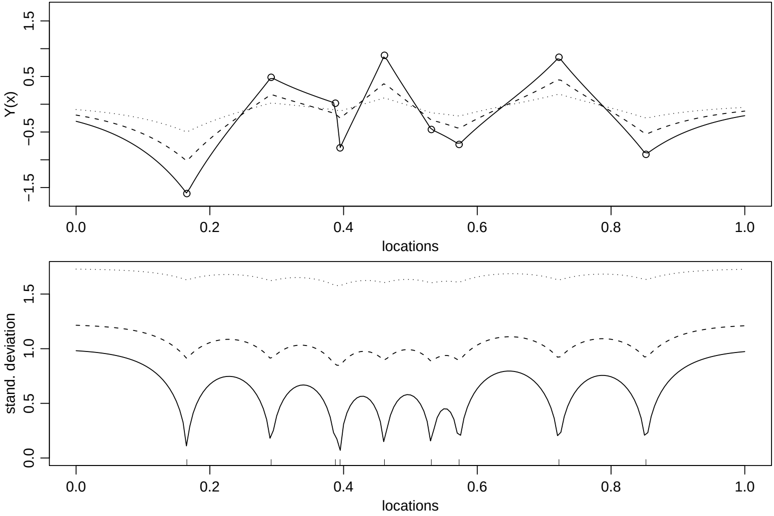
1. Varying κ (smoothness of $S(x)$)



2. Varying ϕ (range of spatial correlation



3. Varying τ^2/σ^2 (noise-to-signal ratio)



Prediction: a universal algorithm

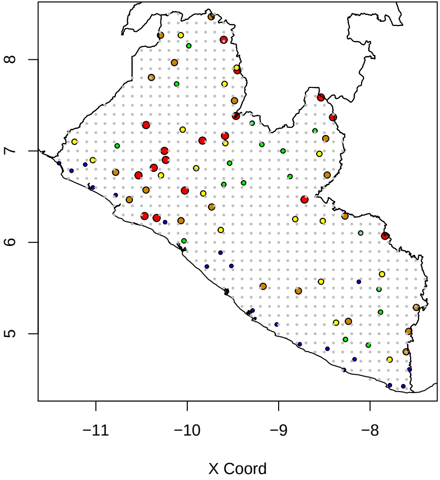
The answer to any prediction problem is a probability distribution

Peter McCullagh, FRS

- $T =$ the predictive target
- $Y =$ data that can tell us something about T.

The predictive distribution of T is the conditional probability distribution of T given Y

Geostatistical prediction of any target T



Linear Gaussian model

$$S^* = \{S(x_1^*), ..., S(x_M^*)\}$$

on prediction grid of locations to cover area of interest

- $[Y]$ = multivariate Normal
- $[S^*|Y]$ = multivariate Normal
- simulate samples from $[S^*|Y]$
- corresponding $T^* = \mathcal{T}(S^*)$ are samples from predictive distribution of T

Prediction with covariates, $d(x)$

Signal $T(x) = d(x)'\beta + S(x)$ Data $(Y_i, x_i, d(x_i)) : i = 1, 2, ...$

Model $S(x) \sim$ Gaussian process, $Y_i | S \sim \mathrm{N}(T(x_i), \tau^2)$

- Point prediction

$$\hat{T}(x) = d(x)'\hat{\beta} + \hat{S}(x)$$

- Plug-in prediction

  Sample from $[T(x)|Y]$ with parameters fixed at their maximum likelihood estimates

- Parameter uncertainty

  Sample from $[T(x)|Y]$ with parameters sampled from the multivariate Normal distribution of their maximum likelihood estimates

Transformations

- Assumptions for Gaussian model may hold more closely after point-wise transformation
- Two widely used examples:

  1. Logarithm

  – often useful when outcome is non-negative, real-valued
  – converts multiplicative relationships to additive ones

  2. Empirical logit

  – often useful when outcome is a proportion

$$el(p) = \log\{p/(1-p)\}$$

  – or if outcome is numerator y and denominator n

$$el(y) = \log\{(y+0.5)/(n-y+0.5)\}$$

  – but may be better to use binomial model (next lecture)

Bayesian inference

Model specification

$$[Y, \theta] = [\theta][Y|\theta]$$

- $[Y|\theta]$ probability distribution of Y given parameter value θ
- $[\theta]$ prior probability distribution for θ
  (before you collect any data)

Parameter estimation

- Bayes' Theorem gives posterior distribution for θ
  (adding information from data)

$$[\theta|Y] = [Y|\theta][\theta]/[Y]$$

where $[Y] = \int [Y|\theta][\theta] d\theta$

Bayesian inference for geostatistical models

Model specification

$$[Y, S, \theta] = [\theta][S|\theta][Y|S, \theta]$$

- $[S]$ is an unobserved spatial stochastic process, representng the spatial phenomenon of scientific interest

Parameter estimation

- integration gives likelihood function

$$[Y, \theta] = \int [Y, S, \theta] dS = [\theta][Y|\theta]$$

- as before, Bayes' Theorem gives posterior distribution

$$[\theta|Y] = [Y|\theta][\theta]/[Y]$$

where $[Y] = \int [Y|\theta][\theta] d\theta$

Bayesian inference for geostatistical models (2)

Prediction

S denotes the spatial process of interest at data-locations

S^* denotes the same process at data and prediction locations

- expand model specification to

$$[Y, S^*, \theta] = [\theta][S|\theta][Y|S,\theta][S^*|S,\theta]$$

- plug-in predictive distribution is

$$[S^*|Y, \hat{\theta}]$$

- Bayesian predictive distribution is

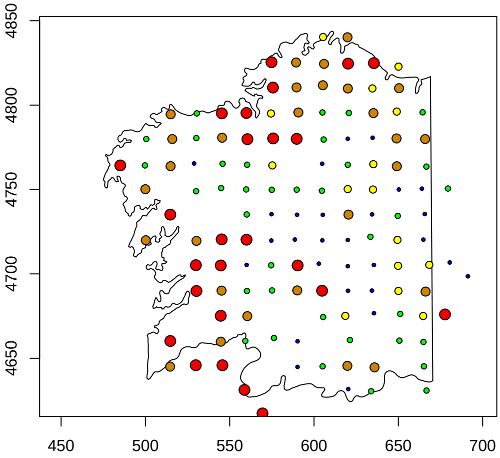
$$[S^*|Y] = \int [S^*|Y,\theta][\theta|Y]d\theta$$

- for any target $T = t(S^*)$, required predictive distribution $[T|Y]$ follows by direct calculation

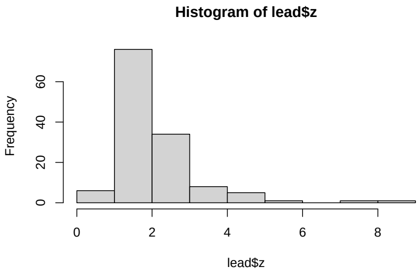
Notes

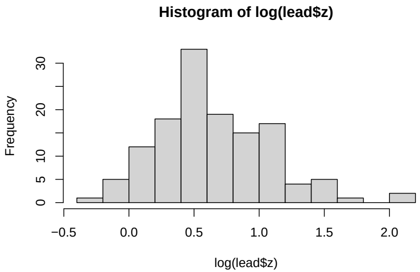
- likelihood function is central to both classical and Bayesian inference
- Bayesian prediction is a weighted average of plug-in predictions, with different plug-in values of θ weighted according to their conditional probabilities given the observed data.
- Bayesian prediction is usually more conservative than plug-in prediction
- Non-Bayesian alternative is to sample parameter values from the multivariate Normal distribution of their maximum likelihood estimates

Lead concentrations in Galicia, year 2000

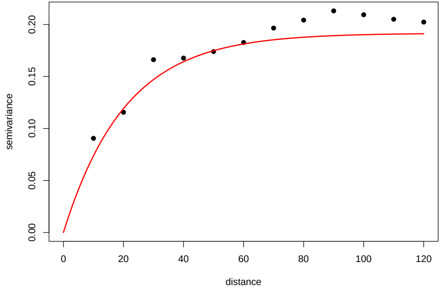


Lead pollution in Galicia: data and variograms



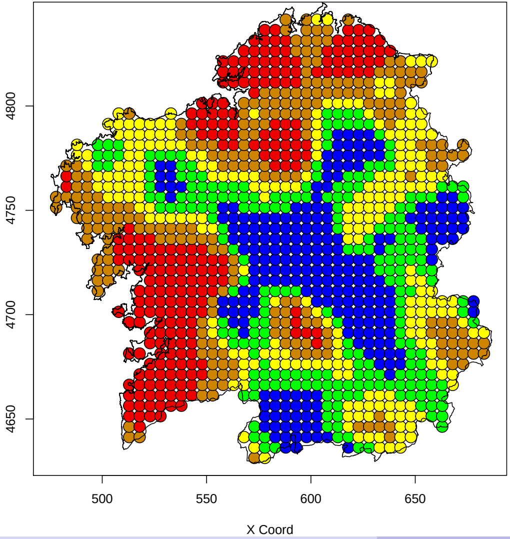


- Fit model to log-transformed lead concentrations
- $V(u) = \tau^2 + \sigma^2\{1 - \exp(-u/\phi)\}$

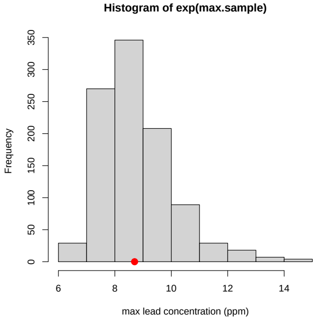


Lead pollution in Galicia: predictions

Point prediction

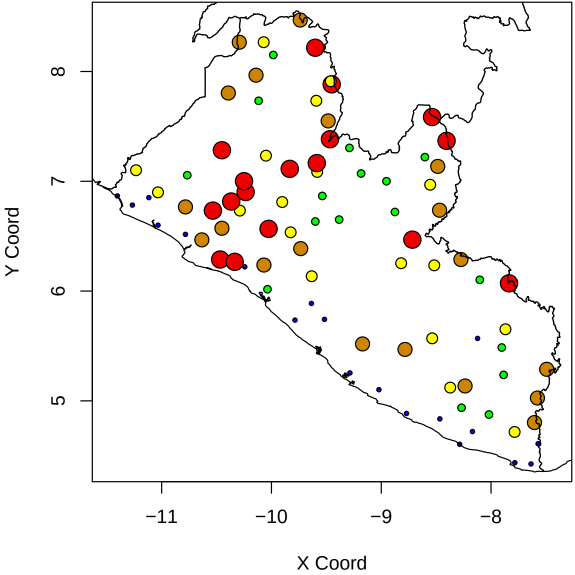


Maximum lead concentration



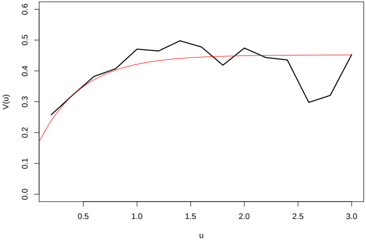
Note: maximum observed value of lead pollution indicated by red dot

Onchocerciasis in Liberia

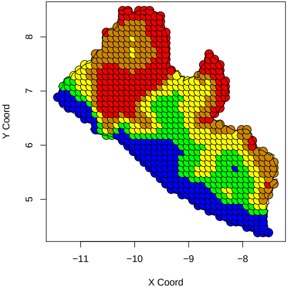


Onchocerciasis in Liberia: predictions

- Fit linear model to logit prevalence
- Longitude and latitude as covariates
- $V(u) = \tau^2 + \sigma^2\{1 - \exp(-u/\phi)\}$



Probability that prevalence exceeds 0.2



0.000-0.205
0.205-0.435
0.435-0.591
0.591-0.698
0.698-0.983

Geostatistical design: where to sample?

Classical ideas from survey sampling design apply

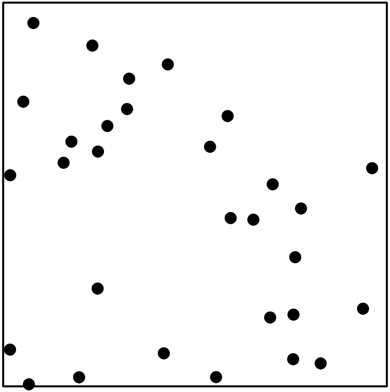
- randomize to avoid subjective bias
- stratify to controls for large-scale spatial variation...and for operational convenience

But spatial correlation $\Rightarrow$ completely random sampling is inefficent

- constrain randomisation to achieve a more even spatial coverage
- supplement with a few close pairs of locations if possible, to estimate nugget variance

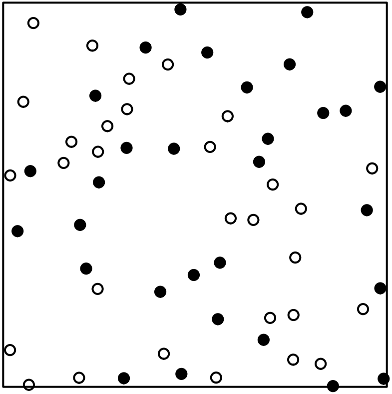
Spatially regulated sampling designs

Sample at random subject to a minimum distance constraint



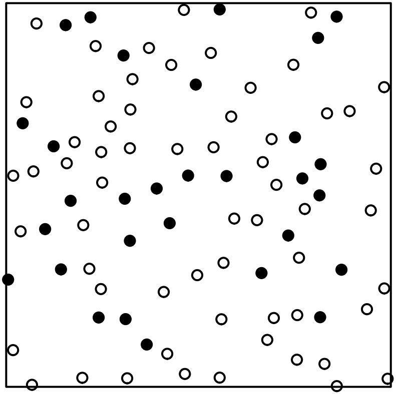
Spatially regulated sampling designs

Sample at random subject to a minimum distance constraint



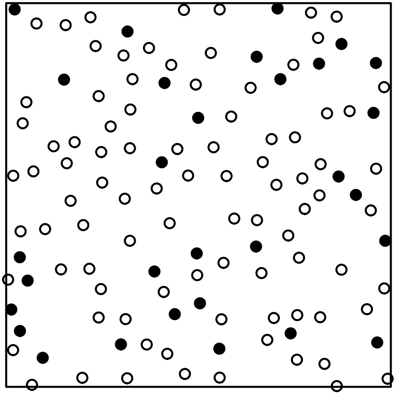
Spatially regulated sampling designs

Sample at random subject to a minimum distance constraint

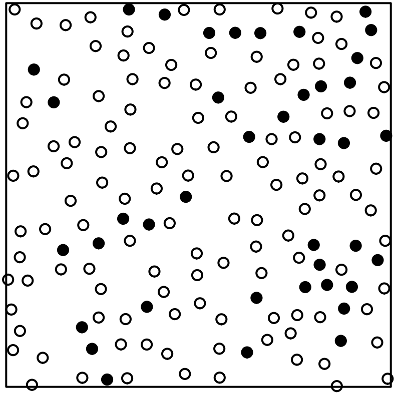


Spatially regulated sampling designs

Sample at random subject to a minimum distance constraint

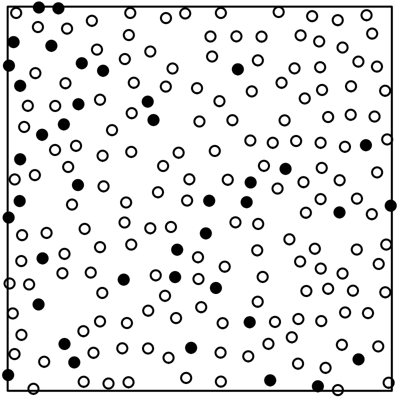


Sample at random subject to a minimum distance constraint

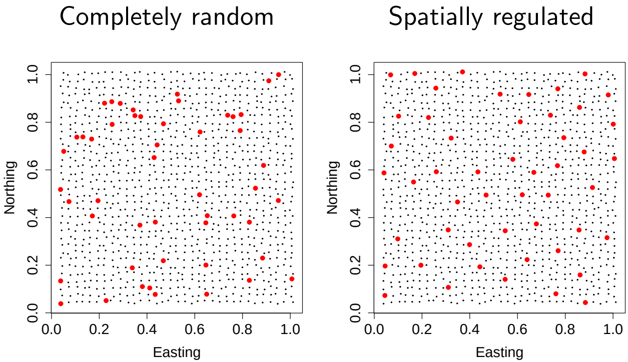


Spatially regulated sampling designs

Sample at random subject to a minimum distance constraint



Spatially regulated sampling from a pre-specified set of locations



- Adding a few close pairs is still a good idea
- But geographical constraints may work against this