Spatial predictions

Abhi Datta

February 9, 2018

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

Review of last lecture

- Spatial linear regression model for univariate point-referenced spatial data
- Modeling unknown surfaces with Gaussian Processes
- MLE, choosing initial values of spatial parameters using the semivariogram
- Kriging: Predictions at new locations
- Out of sample prediction
- Model comparison: AIC, BIC, RMSPE, CP, CIW
- Analysis in R

Modeling with GPs

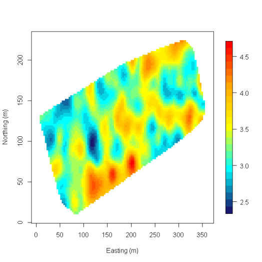
Spatial linear model

$$y(s) = x(s)'\beta + w(s) + \epsilon(s)$$

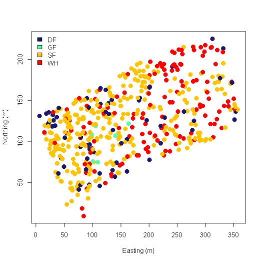
- Spatial random effects $w(s)$ modeled as $GP(0, C(\cdot \,|\, \theta))$ (usually without a nugget)
- $\epsilon(s) \stackrel{\text{iid}}{\sim} N(0, \tau^2)$ is the measurement error
- $y(s)$ becomes a GP with mean $x(s)'\beta$ and covariance function C_1 such that $C_1(s_i, s_j|\theta) = C(s_i, s_j|\theta) + \tau^2 Ind(s_i = s_j)$
- $y = (y(s_1), \ldots, y(s_n))' \sim N(X\beta, C(\theta) + \tau^2 I)$

Western Experimental Forestry (WEF) data

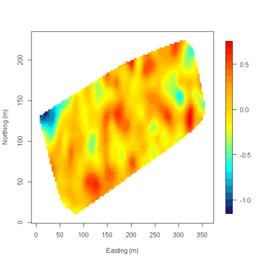
- Data consist of a census of all trees in a 10 ha. stand in Oregon
- Response of interest: log(Diameter at breast height), i.e., log(DBH)
- Covariate: Tree species (Categorical variable)



DBH

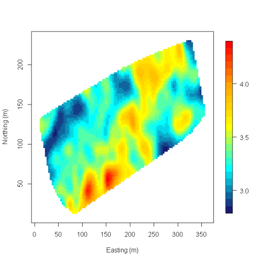


Species

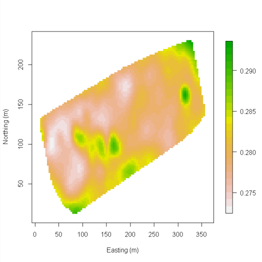


Lin. reg. Residuals

WEF data: Kriged surfaces after spatial regression



DBH Predictions



Variances

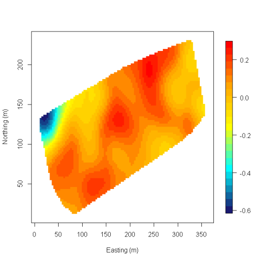
Recovering the spatial random effects

- The kriging used in last class was for predicting $y(s_0)|Y$
- The spatial random effects $w(s)$ was introduced to account for spatial variation not explained by the covariates
- Recovering the spatial surface $w(s)$ gives us an idea about what covariates may be missing
- Often of critical importance for scientists
- How do we predict $w(s_0)|Y$?

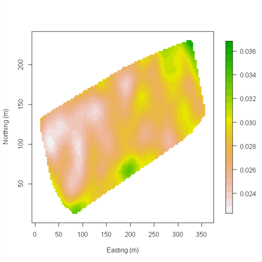
Recovering the spatial random effects

- Note that $Y = X\beta + w + \epsilon$
- $Cov(w(s_0), Y) = Cov(w(s_0), w) = (c(s_1, s_0|\theta), \ldots, c(s_n, s_0|\theta))'$
- Also, both Y and $w(s_0)$ are jointly normal
- $w(s_0)|Y \sim N(c'(C + \tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) - c'(C + \tau^2 I)^{-1}c)$
- Recall that when $s_0 \notin (s_1, s_2, \ldots, s_n)'$, then $y(s_0)|Y \sim N(x(s_0)'\beta + c'(C + \tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) + \tau^2 - c'(C + \tau^2 I)^{-1}c)$
- Setting $signal = TRUE$ and $trend.l = 0$ in *krige.conv* will predict $w(s_0)$ instead of $y(s_0)$

Predicted $w(s)$ surface for WEF data

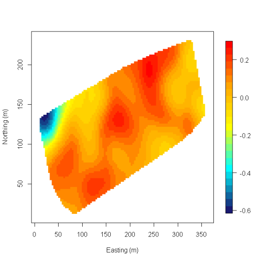


Predicted $w(s)$

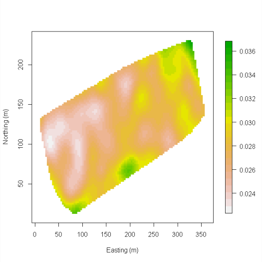


Variances of $w(s)$

Predicted $w(s)$ surface for WEF data



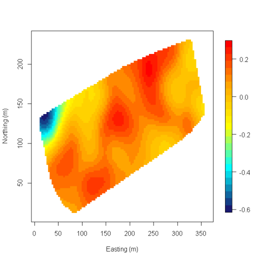
Predicted $w(s)$



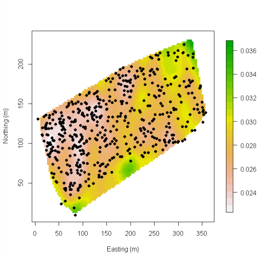
Variances of $w(s)$

Note the strong resemblance of the predicted surface with the residual surface from the non-spatial regression

Predicted $w(s)$ surface for WEF data



Predicted $w(s)$



Variances of $w(s)$ and data locations

Areas of high variance are usually located away from the data locations

The prediction variance

- Which part of the prediction variance in kriging comes from uncertainty in the covariates, i.e., from the linear regression?
- Which part is due to the random effects $w(s)$ and which is simply noise variance ?

The prediction variance

- Which part of the prediction variance in kriging comes from uncertainty in the covariates, i.e., from the linear regression?
- Which part is due to the random effects $w(s)$ and which is simply noise variance ?
- In linear regression (no spatial term), prediction at a new location is $\hat{y}_{new} = x'_{new}\hat{\beta} + \epsilon_{new}$
- So, $\widehat{var(\hat{y}_{new})} = \widehat{var(x'_{new}\hat{\beta})} + \widehat{var(\epsilon_{new})} = \hat{\tau}^2(x'_{new}(X'X)^{-1}x_{new} + 1)$
- The contributions from the covariates and noise are easily separable
- Can we do it for the spatial linear model?

The prediction variance

- $y(s_0)|Y \sim N(x(s_0)'\beta + c'(C(\theta) + \tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) + \tau^2 - c'(C + \tau^2 I)^{-1}c)$

The prediction variance

- $y(s_0)|Y \sim N(x(s_0)'\beta + c'(C(\theta) + \tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) + \tau^2 - c'(C + \tau^2 I)^{-1}c)$
- Rewrite this as:
  $y(s_0)|Y = x'\beta + E(w(s_0)|Y) + \eta(s_0) + \epsilon(s_0)$
  where $\eta(s_0) \sim N(0, c(s_0, s_0|\theta) - c'(C + \tau^2 I)^{-1}c)$ and $\epsilon(s_0) \sim N(0, \tau^2)$

The prediction variance

- $y(s_0)|Y \sim N(x(s_0)'\beta + c'(C(\theta)+\tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) + \tau^2 - c'(C+\tau^2 I)^{-1}c)$
- Rewrite this as:
  $y(s_0)|Y = x'\beta + E(w(s_0)|Y) + \eta(s_0) + \epsilon(s_0)$
  where $\eta(s_0) \sim N(0, c(s_0, s_0|\theta) - c'(C+\tau^2 I)^{-1}c)$ and
  $\epsilon(s_0) \sim N(0, \tau^2)$
- $\widehat{var}(y(s_0)|) = \widehat{var}(x(s_0)'\hat{\beta}) + \widehat{var}(E(w(s_0)|Y)) + c(s_0, s_0) - c'(C+\tau^2 I)^{-1}c + \tau^2$

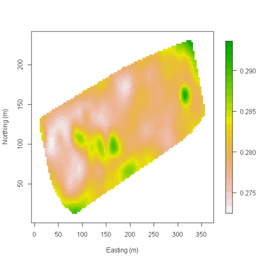
The prediction variance

- $y(s_0)|Y \sim N(x(s_0)'\beta + c'(C(\theta) + \tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) + \tau^2 - c'(C + \tau^2 I)^{-1}c)$
- Rewrite this as:
  $y(s_0)|Y = x'\beta + E(w(s_0)|Y) + \eta(s_0) + \epsilon(s_0)$
  where $\eta(s_0) \sim N(0, c(s_0, s_0|\theta) - c'(C + \tau^2 I)^{-1}c)$ and
  $\epsilon(s_0) \sim N(0, \tau^2)$
- $\widehat{var}(y(s_0)|) = \widehat{var}(x(s_0)'\hat{\beta}) + \widehat{var}(E(w(s_0)|Y)) + c(s_0, s_0) - c'(C + \tau^2 I)^{-1}c + \tau^2 + 2\widehat{cov}(x(s_0)'\hat{\beta}, E(w(s_0)|Y))$

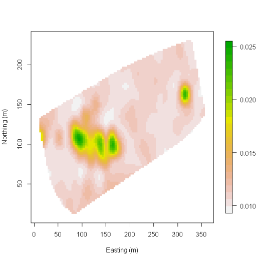
The prediction variance

- $y(s_0)|Y \sim N(x(s_0)'\beta + c'(C(\theta)+\tau^2 I)^{-1}(Y - X\beta), c(s_0, s_0|\theta) + \tau^2 - c'(C+\tau^2 I)^{-1}c)$
- Rewrite this as:
  $y(s_0)|Y = x'\beta + E(w(s_0)|Y) + \eta(s_0) + \epsilon(s_0)$
  where $\eta(s_0) \sim N(0, c(s_0, s_0|\theta) - c'(C+\tau^2 I)^{-1}c)$ and
  $\epsilon(s_0) \sim N(0, \tau^2)$
- $\widehat{var}(y(s_0)|) = \widehat{var}(x(s_0)'\hat{\beta}) + \widehat{var}(E(w(s_0)|Y)) + c(s_0, s_0) - c'(C+\tau^2 I)^{-1}c + \tau^2$
  $+2\widehat{cov}(x(s_0)'\hat{\beta}, E(w(s_0)|Y))$
- The covariance term is hard to interprete and the blue, green and red components are used to understand variation due to covariates, spatial effects and noise

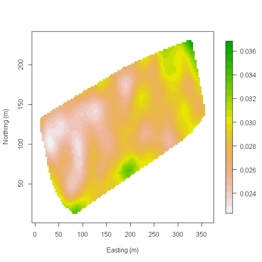
Breaking down prediction variances for WEF data



total

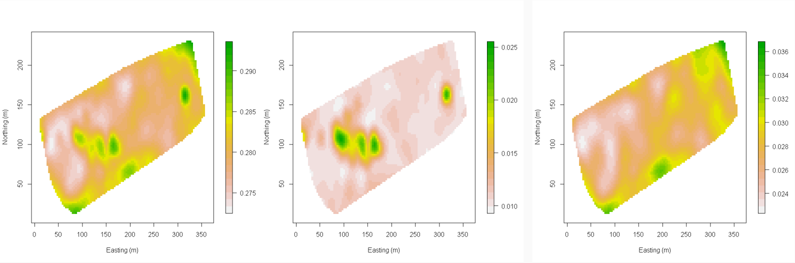


covariates



$w(s)$

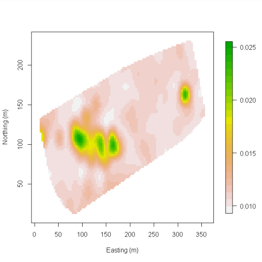
Breaking down prediction variances for WEF data



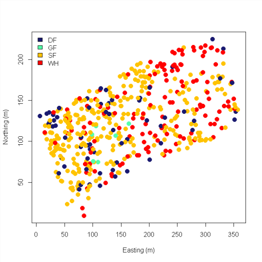
total covariates $w(s)$

- Noise part is simply τ^2 for every location
- What's going on in the patches for the variance from the covariates?

Breaking down prediction variances for WEF data

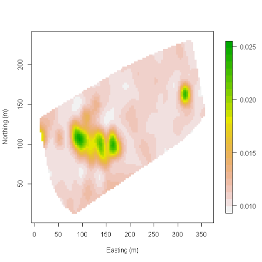


covariates

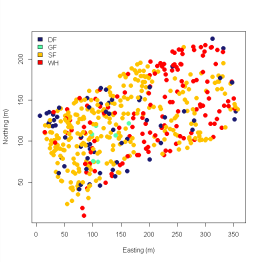


Species

Breaking down prediction variances for WEF data



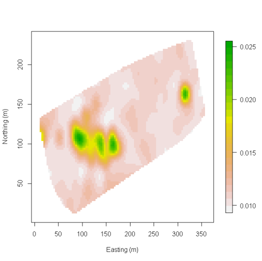
covariates



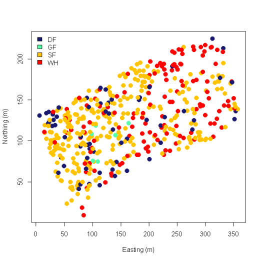
Species

- The in-sample data contains only a handful (6 out of 500) of datapoints for species GF
- Small subsample size $\Rightarrow$ high variance for the regression coefficient corresponding to GF

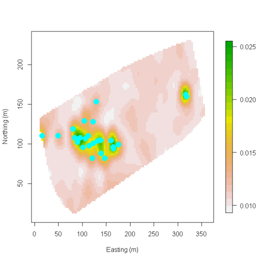
Breaking down prediction variances for WEF data



covariates



Species



locations of GF

- The in-sample data contains only a handful (6 out of 500) of datapoints for species GF
- Small subsample size ⇒ high variance for the regression coefficient corresponding to GF
- Areas of high variance coincide with spots having GF species