DSC 383: Advanced Predictive Models for Complex Data

Section: Spatial Statistics > Subsection: Point-Referenced Spatial Data and Gaussian Processes

INSTRUCTOR:
Catherine (Kate) Calder
Department of Statistics & Data Sciences
University of Texas at Austin

LECTURE: Point-Referenced/Geostatistical Data

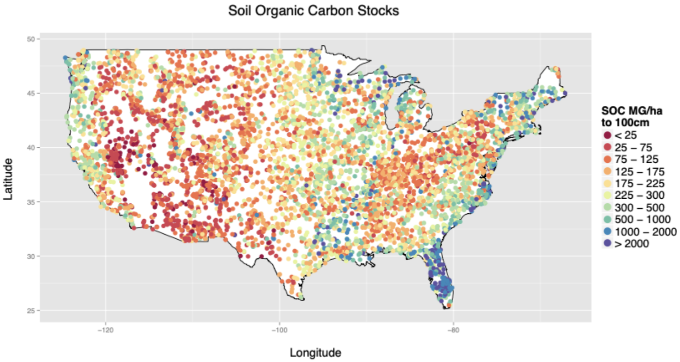
SETTING

- Point-referenced/geostatistical spatial data are observations associated with a fixed set of locations in "space"
  - "space" is often geographic space, but does not have to be...
  - the locations of the observations are considered fixed and the observations/attributes/values associated with the locations are treated arising from a random process

- Goals of statistical analyses of point-referenced/geostatistical data:
  1. inference on the unknown parameters of the random process that generated the observed data
  2. prediction of the process at unobserved locations (with estimates of uncertainty)

SOME EXAMPLES

- Environmental monitoring: soil organic carbon measurements collected as part of the Rapid Carbon Assessment (RaCA) Project



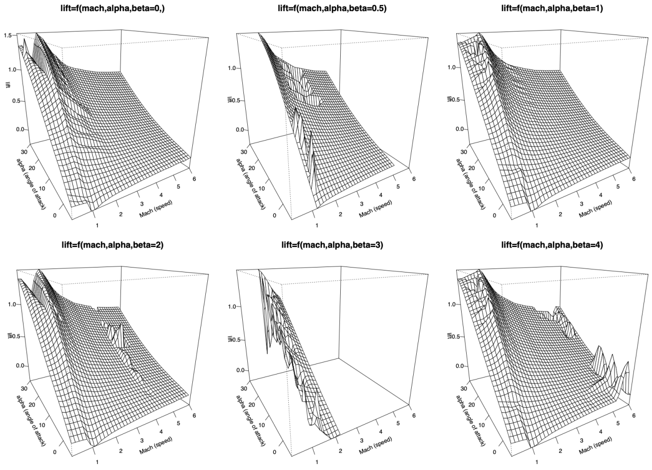
(from Risser, Calder, Berrocal, and Berrett, 2019)

- Time-series data with irregularly space observations

    1. a pediatrician's records of a child's weight over time

    2. the amount of money an individual withdraws from an ATM over time

- Computer experiment: rocket design simulation



from (Gramercy and Lee, 2008)

LECTURE: Gaussian Processes

DEFINITION

- **Definition:** A process, $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subseteq \mathbb{R}^d\}$, is a Gaussian process, if $(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))$ is multivariate normal for every set $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$

$\rightarrow$ A Gaussian process is a continuously-indexed stochastic process

GP PREDICTION/KRIGING

- Assume that $\mathcal{D} = [0, 1]$ and that we observe $Y(s_1), \ldots, Y(s_n)$

- Goal: Estimate $Y(s^*)$ for any $s^* \in \mathcal{D}$

- Challenge:

- Possible solutions:

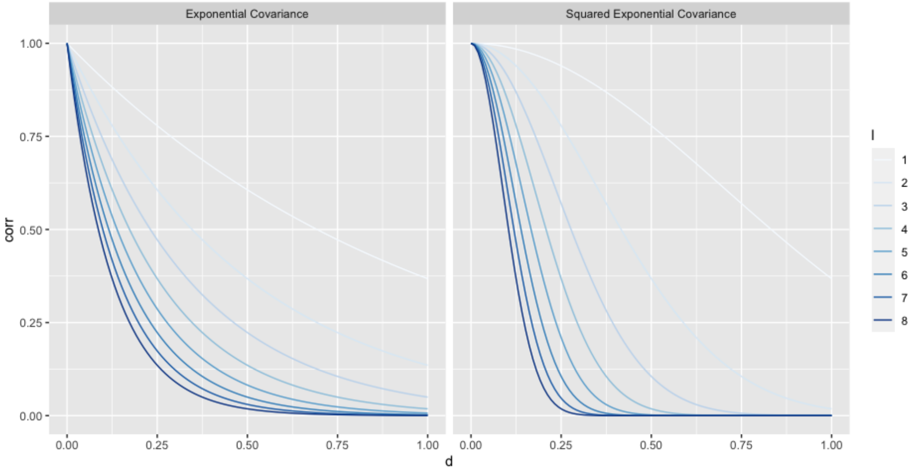
  1. Assume stationarity

  2. Assume a simple parameterization of $\boldsymbol{\Sigma}$

PARAMETRIC COVARIANCE FUNCTIONS

- Exponential:

- Squared Exponential:



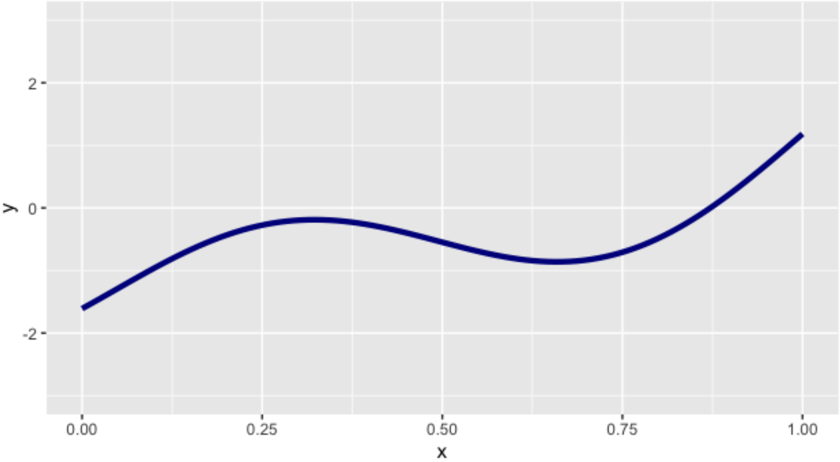
Exponential Covariance
Squared Exponential Covariance
corr
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
d
l
1
2
3
4
5
6
7
8

PREDICTION

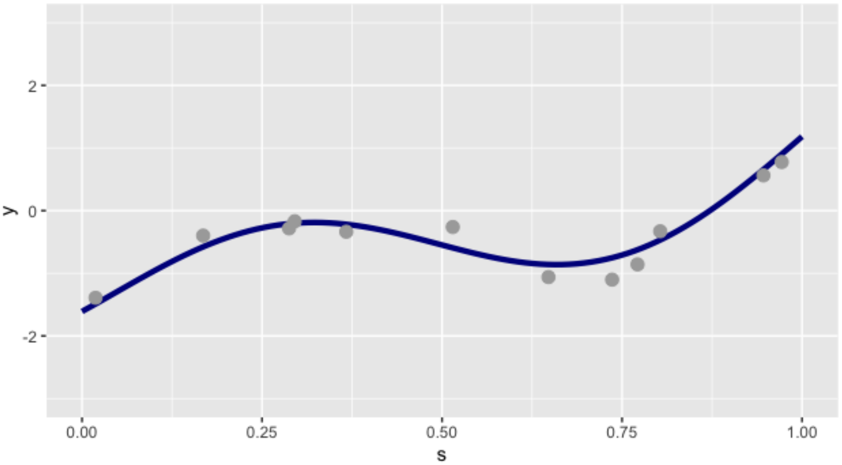
- Example: Let $Y(0.25) = 1$ and $Y(0.8) = -0.5$. Assume $\mu(s) = 0$ for all $s \in \mathcal{D}$ and $\mathbf{\Sigma}_{i,j} = exp(-|s_i - s_j|)$ for all $s_i, s_j \in \mathcal{D}$.

  Estimate $Y(0.5) \quad \rightarrow$

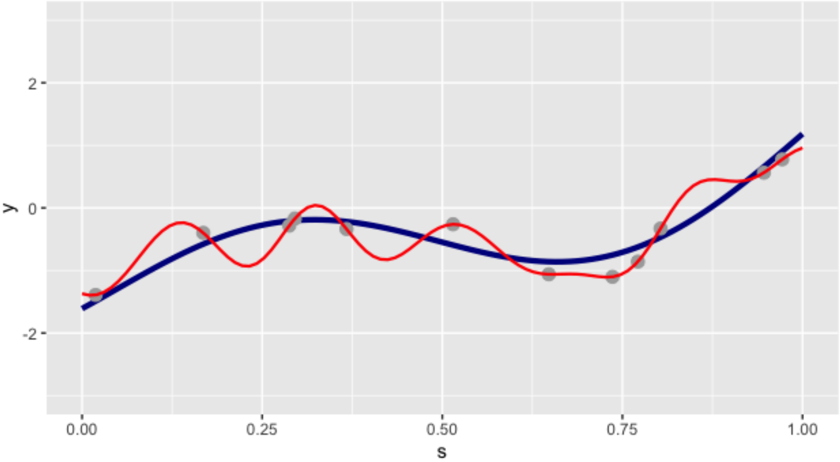
- Example: Truth



- Example cont.: Data = Truth + Noise

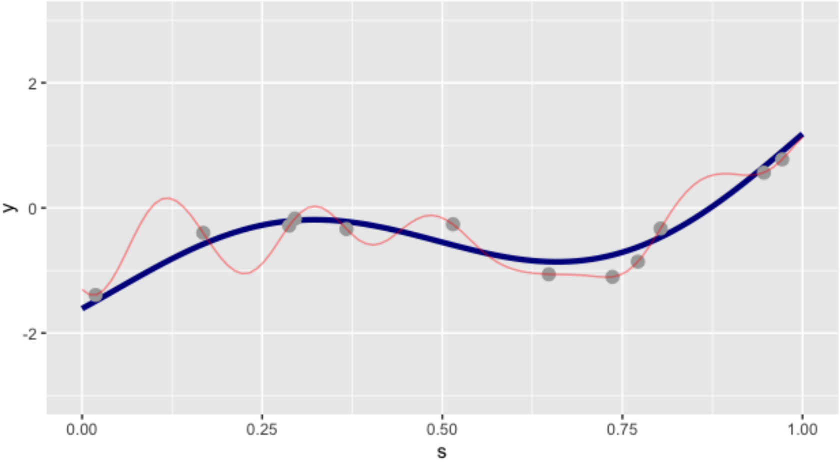


- Example cont.: Predictive mean



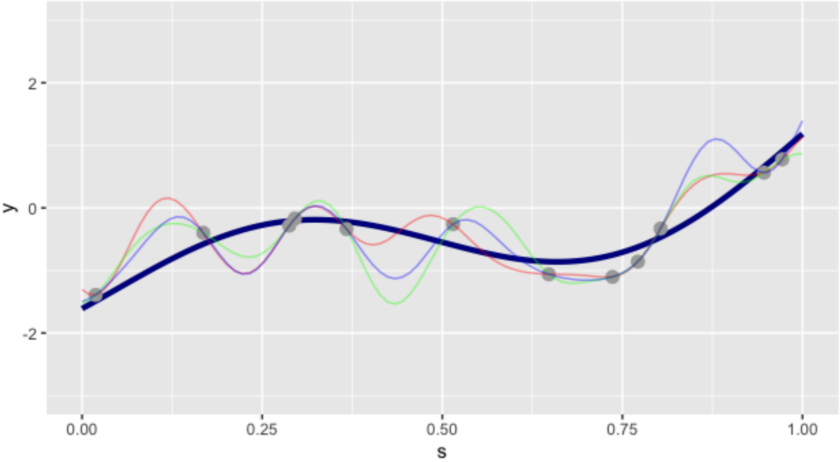
$$\sigma^2 = 1, \ell = 10$$

- Example cont.: Sample from the predictive distribution



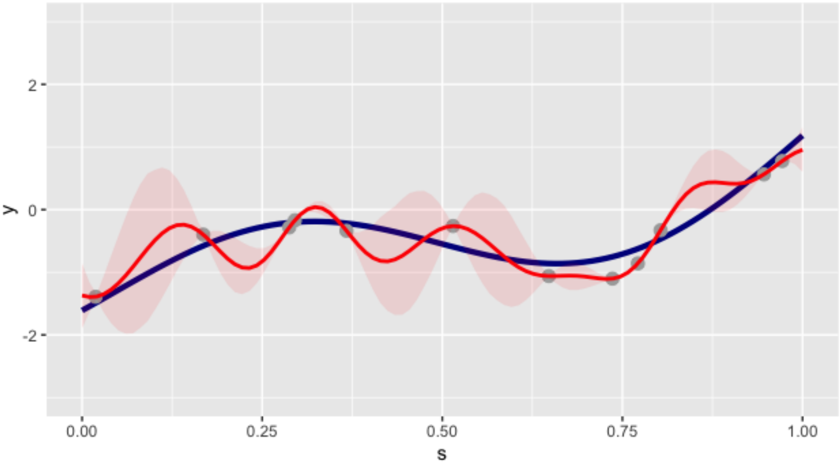
$$\sigma^2 = 1, \ell = 10$$

- Example cont.: **Multiple** samples from the predictive distribution



$$\sigma^2 = 1, \ell = 10$$

- Example cont.: Mean and pointwise 95% intervals based on 1000 samples from the predictive distribution



$$\sigma^2 = 1, \ell = 10$$

NUGGET

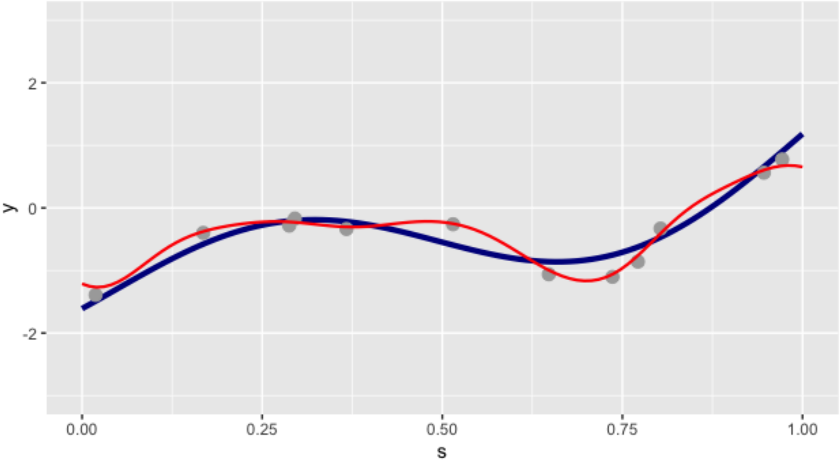
- **Question:** Should the predictions go through the observed data?

  Nugget effect:

  Measurement error:

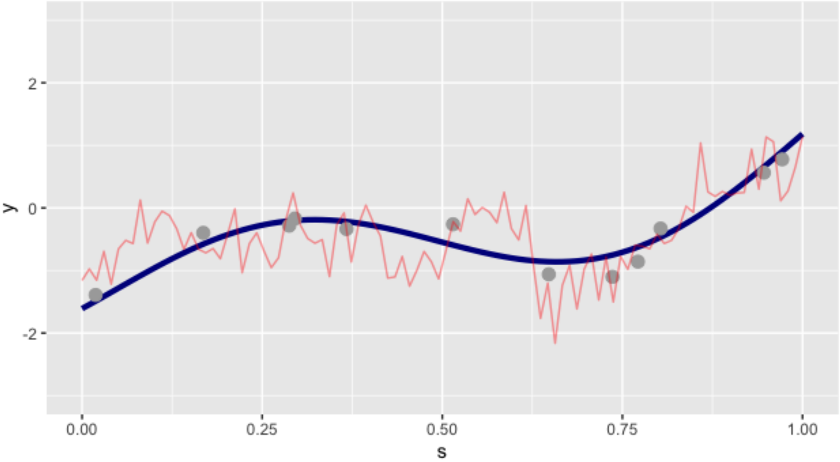
- Covariance functions with nuggets:

- Example cont.: Predictive mean



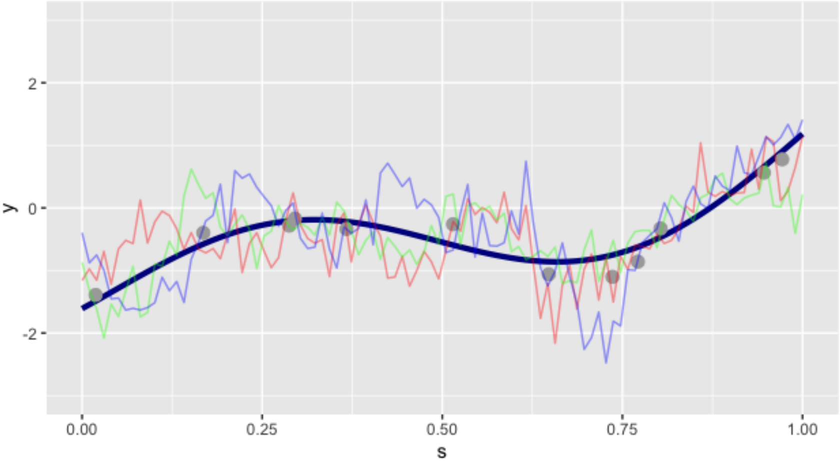
$$\sigma^2 = 1, \ell = 10, \sigma_e^2 = 0.1$$

- Example cont.: Sample from the predictive distribution



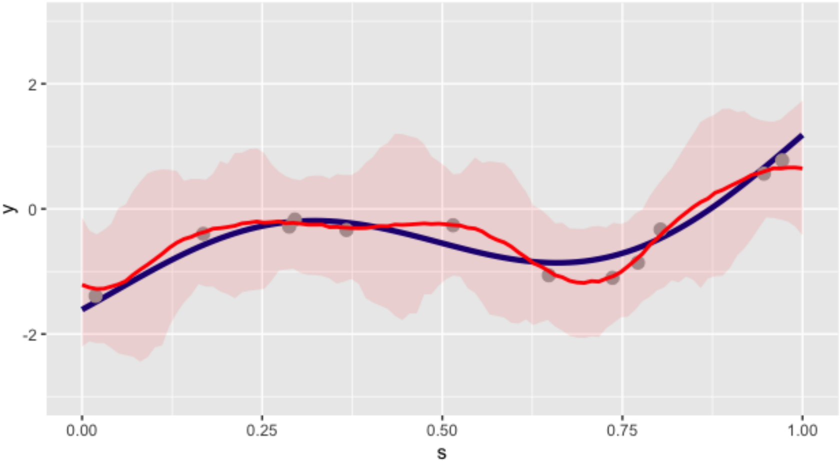
$$\sigma^2 = 1, \ell = 10, sigma_e^2 = 0.1$$

- Example cont.: **Multiple** samples from the predictive distribution



$$\sigma^2 = 1, \ell = 10, \sigma_e^2 = 0.1$$

- Example cont.: Mean and pointwise 95% intervals based on 1000 samples from the predictive distribution



$$\sigma^2 = 1, \ell = 10, \sigma_e^2 = 0.1$$