

# Multivariable Spatial Prediction and Model Validation

---

Stefano Fochesatto

# What is Multivariable Spatial Prediction

- Methods we have discussed in class offer prediction of one variable at a time.
- Recall,

$$\hat{y}(s_0) = \sum_{i=1}^n \lambda_i y(s_i)$$

- Ordinary Kriging;  $\lambda_i$ 's are constrained by relative distances.
- Universal Kriging;  $\lambda_i$ 's are also constrained by trends and covariates.
- For Universal Kriging we found  $\lambda_i$ 's by minimizing MSPE,

$$MSPE = \mathbb{E} \left[ \left( Y(s_0) - \sum_{i=1}^n \lambda_i y(s_i) \right)^2 \right],$$

by the following constraint (lagrange multipliers),

$$\bar{\lambda}^T X = x_0^T.$$

- To find the  $\bar{\lambda}$  using LeastSquares it was necessary to estimate a variogram for  $Y$ .

# What is Multivariable Spatial Prediction

- What if we don't have  $x_0$  where we want to predict  $s_0$ , but we still have secondary data that has information?!
- Cokriging methods follow analogously in derivation; and allow us to directly use existing spatial correlations from secondary-data,

$$\hat{y}(s_0) = \sum_{i=1}^n \lambda_i y_1(s_i) + \sum_{j=1}^k \lambda_j y_2(s_j) + \dots$$

- In solving for  $\lambda_i$  and  $\lambda_j$  using LeastSquares it becomes necessary to estimate variograms and crossvariograms for all  $Y_i$ .

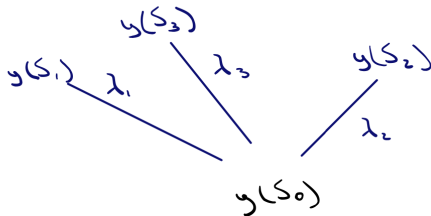
# What is Multivariable Spatial Prediction

- The goal is the spatial prediction of multiple variables simultaneously.
- Multivariable Spatial Prediction is an extension of Cokriging.
- 'It is shown that the cokriging predictor for one variable at a time is identical to the predictor of that same variable in the multivariable predictor.' Ver Hoef & Cressie (1993)
- The constraints for solving for the weights change when cokriging a variable at a time.

# Kriging Overview

- 'The truth of the matter is, when someone says 'kriging' I kind of blend it all together in my mind.' - M. Short.
- same tbh

## Kriging



Ordinary:

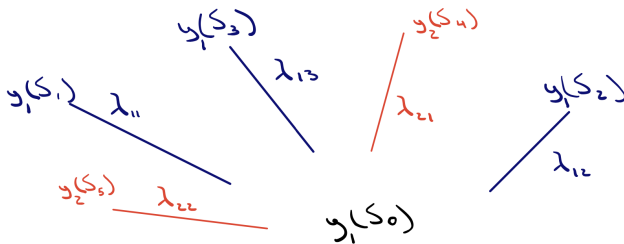
$\lambda_i$ : Constrained by relative distance

Universal:

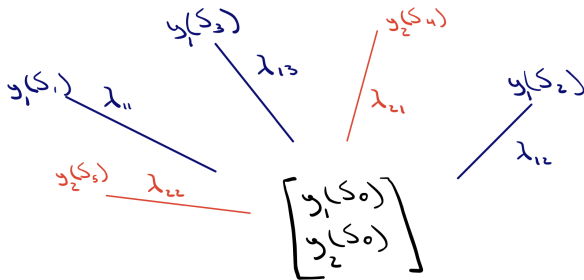
$\lambda_i$ : Constrained by possible covariates

# Kriging Overview

Lo Kriging



## Multivariable Spatial Prediction





## Application to Model Validation in Ecology



# Application to Model Validation in Ecology

- Study put together through several Icelandic environmental agencies, in conjunction with the University of Iceland and the UAF Institute of Arctic Biology.
- Machine learning models (Random Forest and TreeNet) were used to model RIO (relative index of occurrence).
- Nationwide and long term (1860-2021) Rock ptarmigan occurrence data (GBIF).
- Separate occurrence data from the Icelandic Institute of Natural History (2005-2010) for model validation.
- 11 Environmental Layers (May, June, July of 2021).

# Application to Model Validation in Ecology

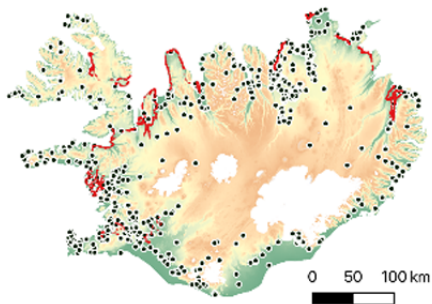
## Legend

- Training data
- Testing data

Elevation a.s.l. (m)

- 0
- 200
- 400
- 600
- 1000
- glaciers

## d. Rock ptarmigans



- Why use Multivariable Spatial Prediction?
  - Separate occurrence data does not have the associated predictors (environmental layers).
  - Current Models are validated with OOB/Cross-Validation.
  - Temper Expectations.

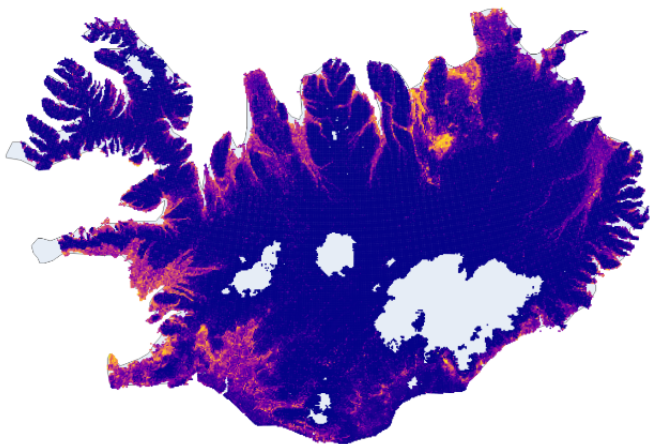
# Application to Model Validation in Ecology

- A tiny bit of code ...

```
library(gstat)
g <- gstat(NULL, id = "NDVImax250", form = NDVImax250 ~ x + y + I(x^2) + I(y^2) + x*y, data=Data.Subsample)
g <- gstat(g, id = "JJA_tavg25", form = JJA_tavg25 ~ x + y + I(x^2) + I(y^2), data=Data.Subsample)
g <- gstat(g, id = "JJA_ppt_av", form = JJA_ppt_av ~ x + y + I(x^2) + I(y^2) + I(x*y), data=Data.Subsample)
g <- gstat(g, id = "JJA_mean_w", form = JJA_mean_w ~ 1, data=Data.Subsample)
g <- gstat(g, id = "dem250", form = dem250 ~ 1, data=Data.Subsample)
g <- gstat(g, id = "soil250", form = soil250 ~ 1, data=Data.Subsample)
g <- gstat(g, id = "veg250", form = veg250 ~ 1, data=Data.Subsample)
g <- gstat(g, id = "slope250", form = slope250 ~ 1, data=Data.Subsample)
```

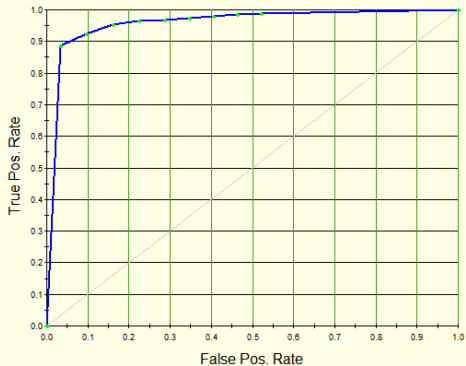
```
g <- fit.lmc(v.cross, g, fit.method = 7, correct.diagonal=1.01)
```

Random Forest RIO Map



# Results

- Accuracy of 91.98% on out of bag samples.



# Results

- Accuracy of 72.91% on outside data with kriged predictors.

