Geostatistical Analysis of Species Diversity Using Field and Topographic data (DEM)

case study: Flysch (Outer Carpathian) Forested Landslides
south Poland

Elvis Tangwa

Supervior: dr inż. Tracz

Outline

Introduction

- Study area
- Objectives
- Concepts and requirements

Material and methods

- ESDA (exploratory spatial data analysis)
- Combined variables

Results and discusssions

- Variogram modelling
- Compare and evaluate models and methods

Conclusions and outlook

Study Area

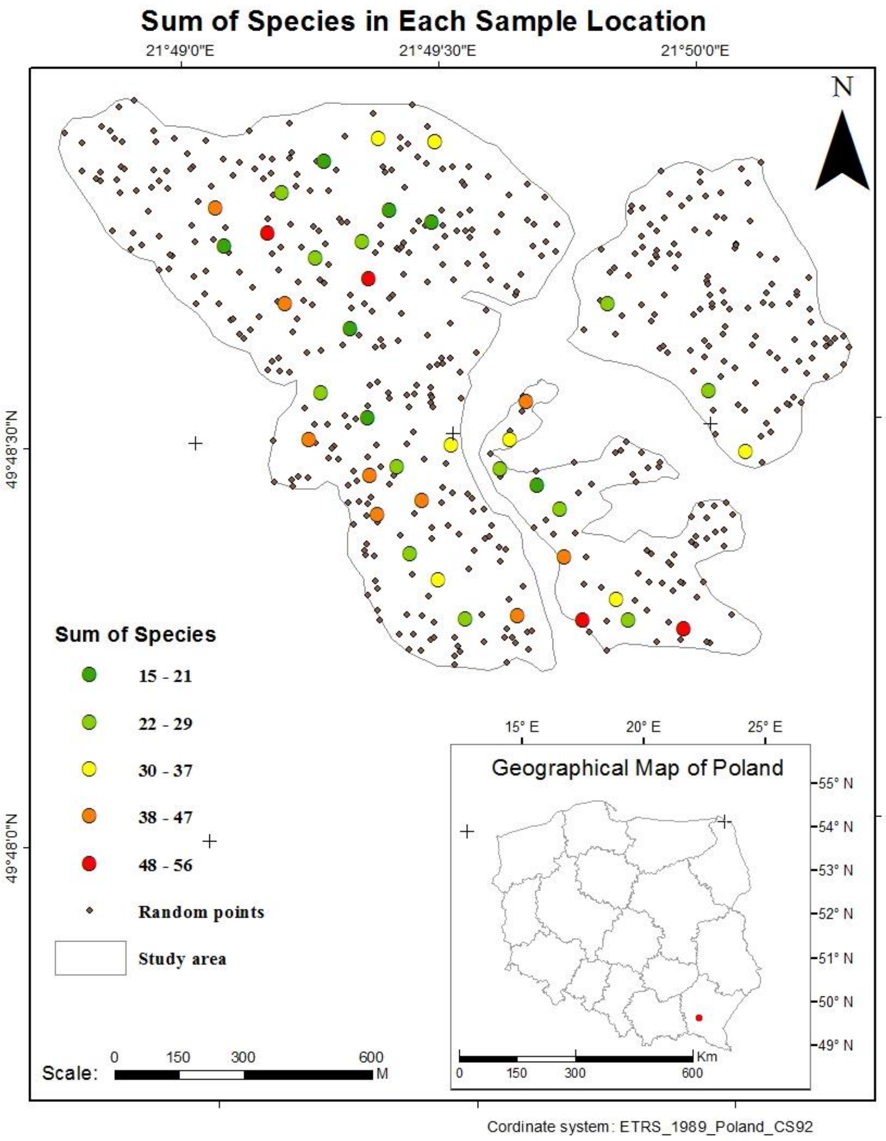


Figure 1: Location of the study area

- Landslide facilitate the evolution and succession of new species *(Alexandrowicz and Margielewski. 2010; Seiwa et al., 2013)*

- Complex terrain with strong topographic variation

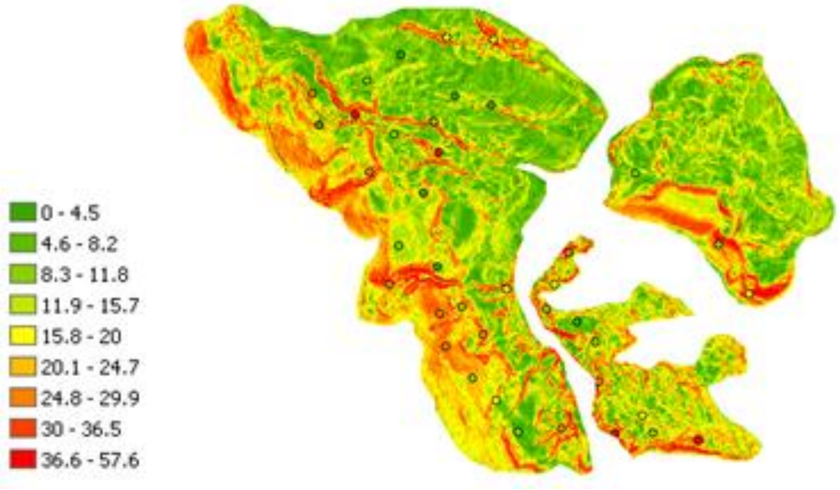


Figure 2: slope variation in study area

Aim and Objectives

Aim

- Predict species diversity using geostatistical methods

Objectives

- Create and select the best possible combined variable which explains variations in species diversity
- Test and compared the performance Ordinary kriging (OK), cokriging (CCK) and regression kriging (RK) methods
- Quantify the uncertainty of each prediction method
- Learn basic and advanced geostistical methods

Concepts and requirments

- **Stationarity** (constant mean and variance)

$$Z(s_i) = \mu + e(s_i) \qquad \text{eq.1}$$

Z = value of variable, s_i = (x, y location), μ =mean, and $e(s_i)$ = random error

- **Spatial depedence**

(autocorrelation or cross correlation)

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[Z(s_i) - Z(s_{i-h})]^2 \qquad eq. 2$$

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}\{[Z(s_i) - Z(s_{i-h})].[Y(s_i) - Y(s_{i-h})]\}$$

n =number of paired points for a given lag (h),
i - h; is a unit distant between two sample locations
$Z(s_i)$ and $Y(s_i)$ are primary and secondary variable respectively

- **Normality of distribution**

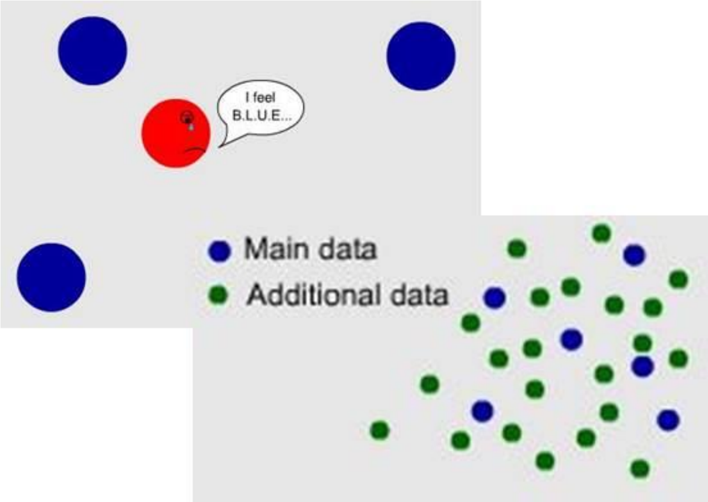


Figure 3: Spatial dependence. Source: UNIGIS ,Salzburg, Austria (April 2017)

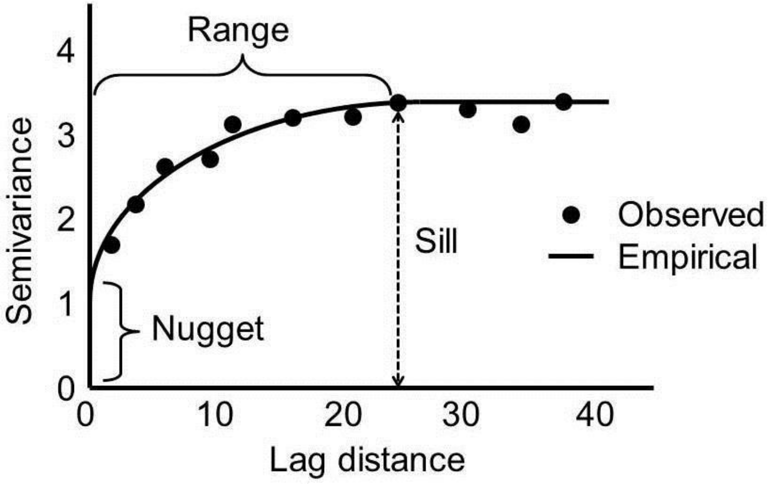


Figure 4: Illustration of a spherical model with associated parameter. Source: Biswas and Cheng Si, (2013)

Kriging and Interpolation

Ordinary Kriging (OK)

$$\hat{z}\,(s_?) = \sum_{n=1}^{N} \lambda_i \, \mathrm{Z}(s_i) \qquad \text{eq. } 4$$

Base on weighted averages determined by semivariogram model *(Webster and Olivier, 2007)*.

Ordinary Cokriging (CCK)

$$\hat{z}\,(s_?) = \sum_{i=1}^{N_1} \lambda_{1i}\, Z_1(s_{1i}) + \sum_{j=1}^{N_2} \lambda_{2i}\, Z_2(s_{1j}) \quad eq.5$$

- Primary and secondary variable should have near similar spatial structure *(Bivand et al., 2013; Krivoruchko and Wood 2014).*

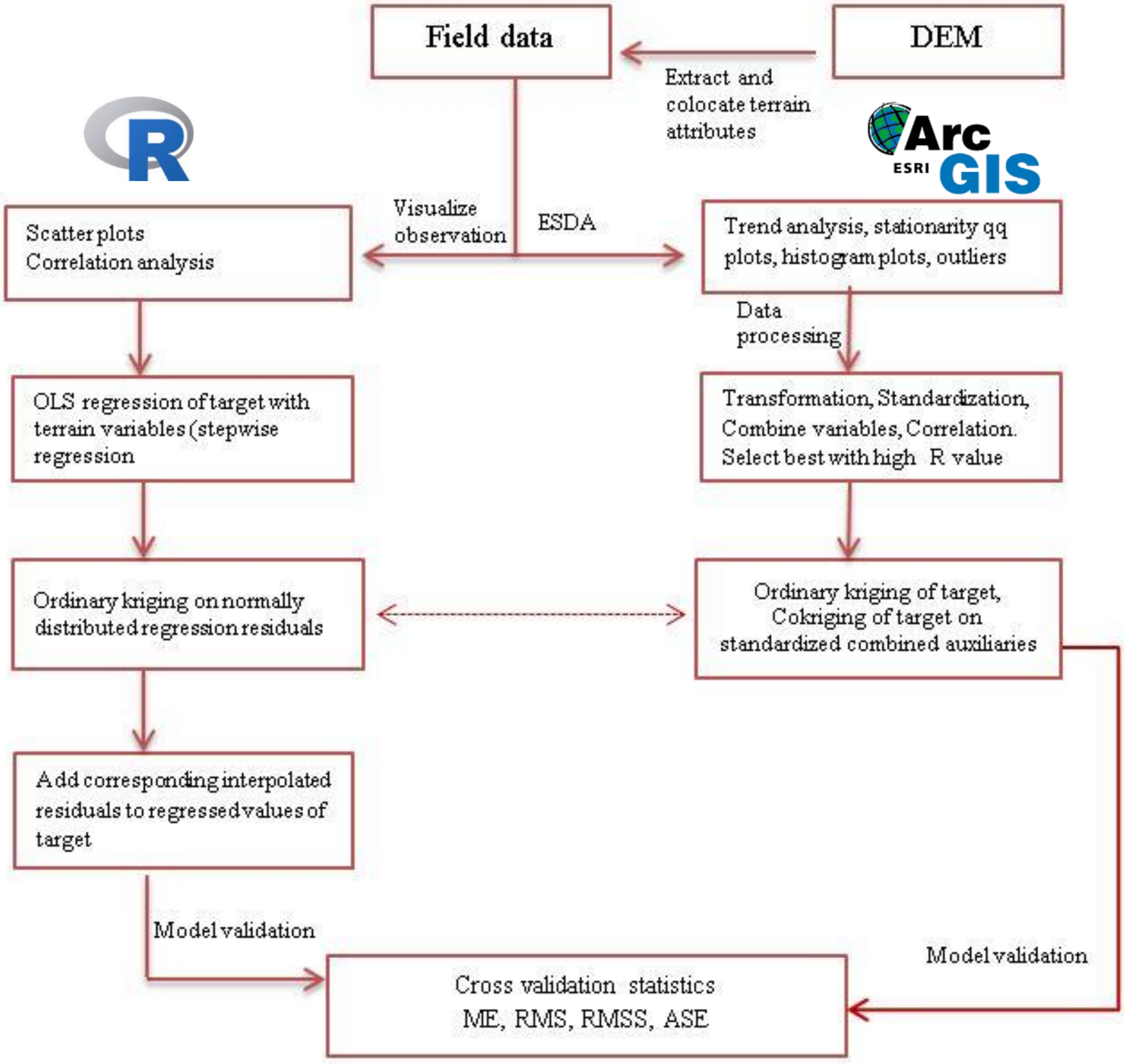
- **Regression Kriging (RK)**

$$\hat{z}\,(s_?) = \sum_{k=o}^{p} \beta_k \,.\, q_k(s_?) + \sum_{i=1}^{N} \lambda_i . e(s_i) \qquad eq.6$$

- OLS regression + OK, a linear relationship must exist and residuals must be autocorrolated *(Hengl et al., 2004a; Odeh et al., 1995).*

Where: $\hat{z}\,(s_?)$ *= value at unvisited location,* $\mathrm{Z}(s_i)$ = observed sample value, λ_i *= kriging wights, N and N2 are respctively number of primary and secondary variable in search neighborhood, p* =number of predictors, β_k*= regression coefficient, qk*-th =predictors, $e(s_i)$ *is the regression residual at location* s_i

Material and Methods



Field data

DEM

Extract and colocate terrain attributes

Visualize observation

ESDA

Scatter plots
Correlation analysis

Trend analysis, stationarity qq plots, histogram plots, outliers

Data processing

OLS regression of target with terrain variables (stepwise regression

Transformation, Standardization, Combine variables, Correlation. Select best with high R value

Ordinary kriging on normally distributed regression residuals

Ordinary kriging of target, Cokriging of target on standardized combined auxiliaries

Add corresponding interpolated residuals to regressed values of target

Model validation

Model validation

Cross validation statistics
ME, RMS, RMSS, ASE

Figure 5: Methodology and workflow

Combined variables!

How?

linearly merged standardized terrain attributes originating from the same location

$$SV_A = \frac{V_A - \mu}{S} \qquad \text{eq. 7}$$

Where SV_A = standardized value of variable V_A = unstandardized value of terrain variable at location A,

μ= mean value for a given terrain variable and S = standard deviation

R Packages :

Rgdal: reads shapefiile

Gstat: variogram modelling

SP: spatial classes, methods and functions

ArcGIS:

Geostatistical Analyst toolbox

Trend: mostly 2nd order polynomial

Results and Discussions

Table 3: Correlation between standardized combined

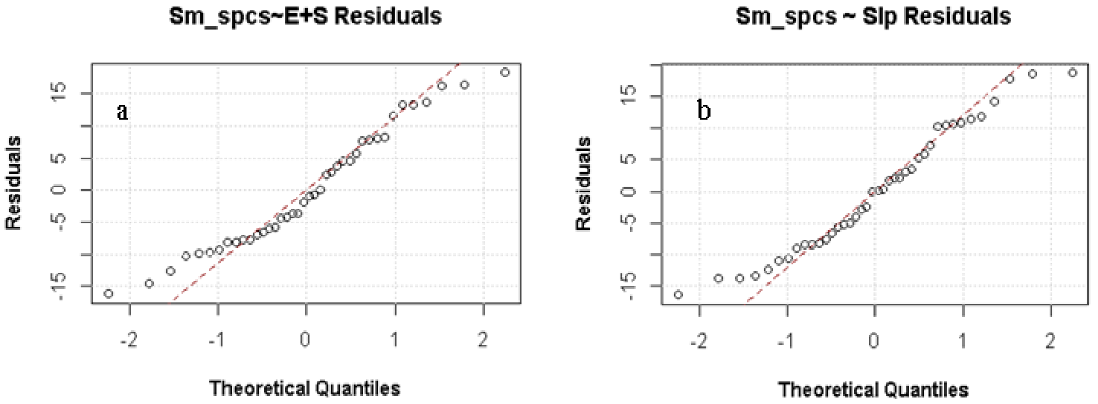
Combinations	Abbreviation	R
Slope	-	**0.53**
Slope + elevation	SE	0.25
Slope + aspect	SA	0.45
Elevation + aspect + slope	EAS	0.33

- Weak to moderate correlation with target

- Only slope and elevation were significant predictors

- Slope + elevtation explained ~36% variation in species distribution compared ~ 28% with slope alone

- Autocorrolation ~ 200 m



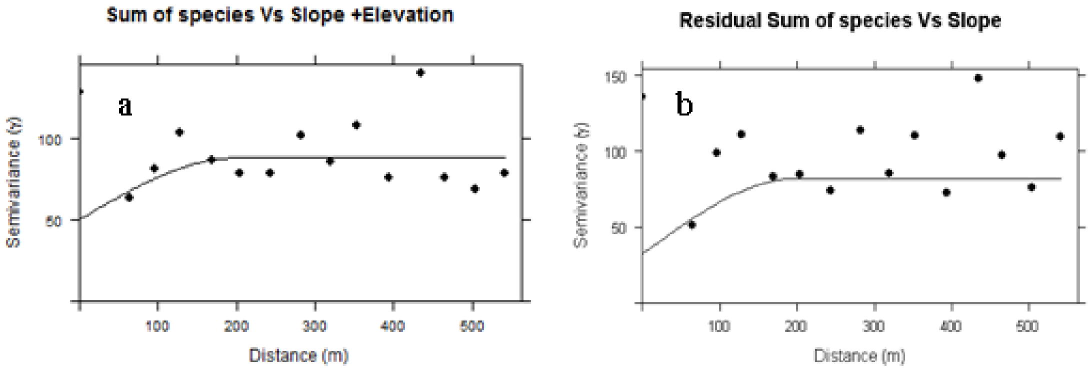
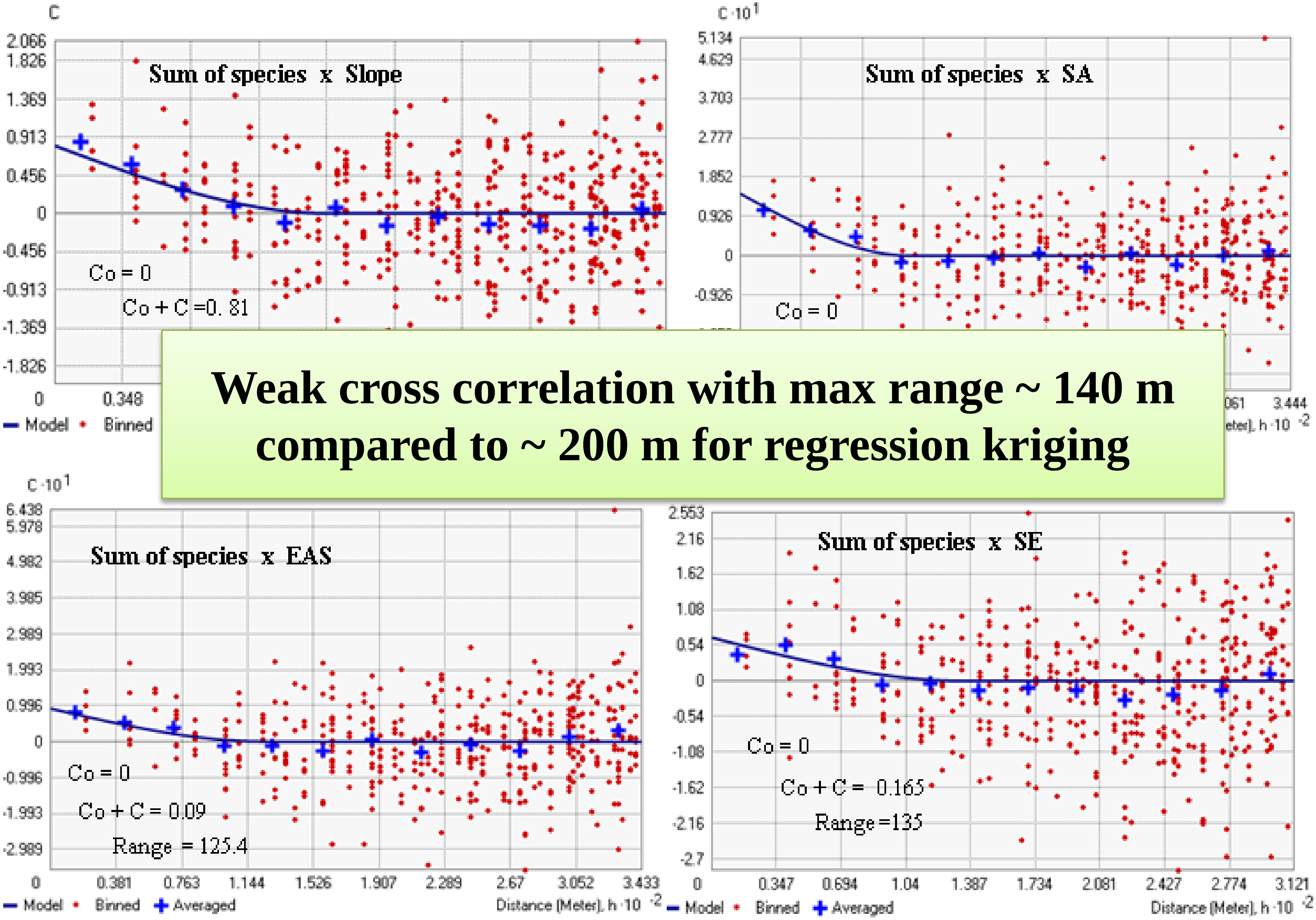


Figure 6: Variogram of regression residuals *slope + elevation x target and (b) slope x target*

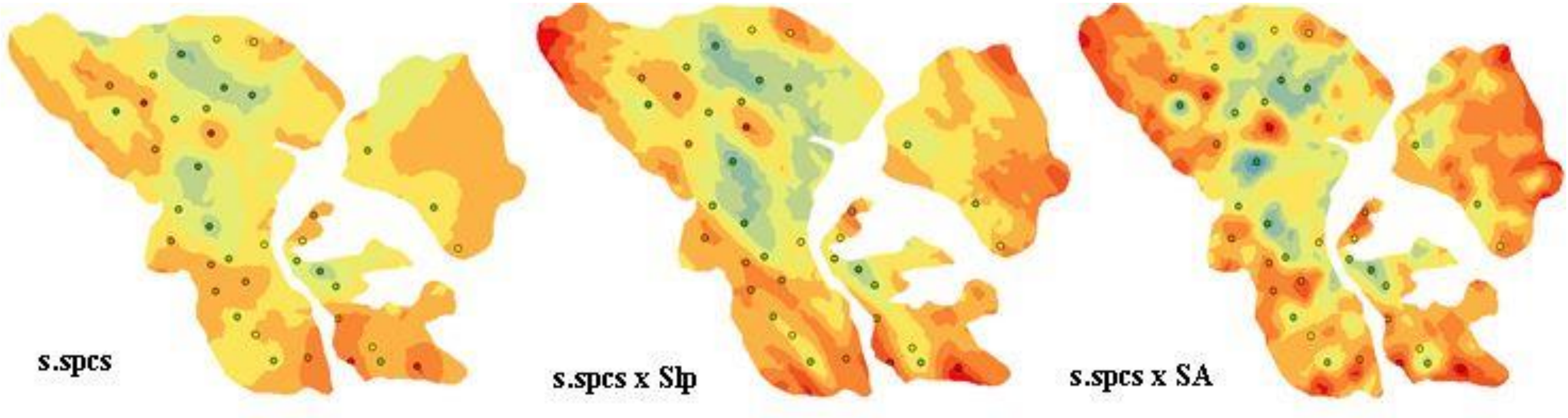
Results and Discussion Conti...

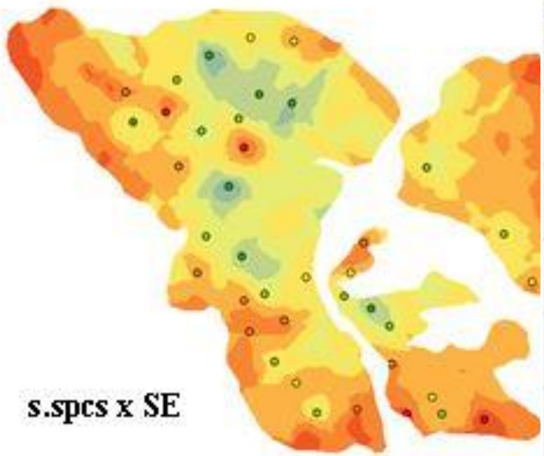


Weak cross correlation with max range ~ 140 m compared to ~ 200 m for regression kriging

Figure 8: Cross-variogram (target x auxiliaries) Co= nugget effect, Co + C= total sill

Species distribution





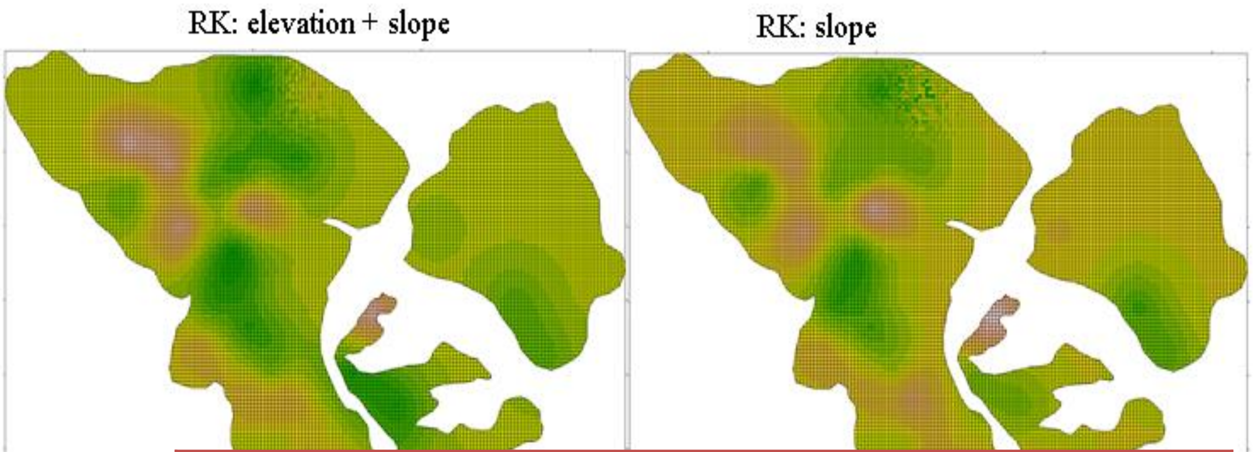


Figure 9: Predicted species dive
cokriging (CCK)

Max value under estimated and Min value over estimated

	Observed	Prd.OK sm_spcs	Prd.CKK _SA	Prd.CCK _Slp	Prd.CCK _SE	Prd.RK _SE	Prd.RK _Slp
Min	**15**	25.5	23.13	23.10	24.1	**17.2**	22.3
Max	**56**	42.7	47.49	51.64	49.2	**53.0**	47.9
Std	11.5	3.9	5.6	6.49	5.27	7.28	6.16

Table 3: Correlation between standardized combined

Results and discussion cont...

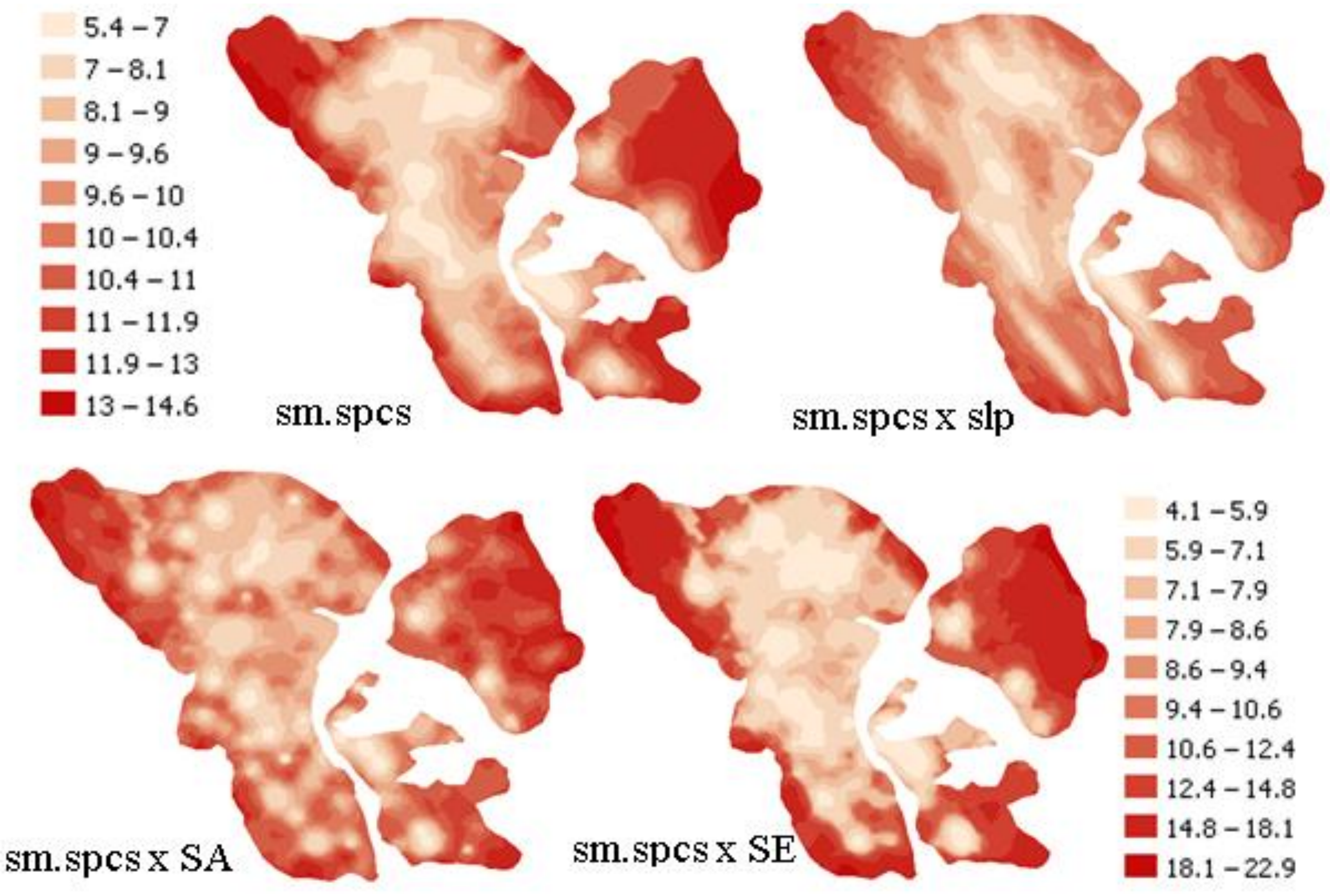


Figure 11: Predicted error based on Ok and CCK

- Almost similar variation pattern in predicted error
- Much higher error based on CCK compared to OK

Cross validation

- **ME** (mean error): unbias model should be ~ 0

$$ME = \frac{1}{n}\sum_{i=1}^{n} Z(x_i) - z(x_i) \qquad eq.8$$

- **RMSE** (root mean squared error): Model precision.

should be as small as possible:

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(Z(x_i) - z(x_i))^2 \qquad \text{eq.9}$$

- **RMSS** (root mean squared standardized) should be ~1 : Model stability.

$$RMSS = \frac{1}{n}\sum_{i=1}^{n}\frac{(Z(x_i)-z(x_i))^2}{\sigma_k^2(x_i)}) \qquad eq.10$$

- **ASE** (average standard error)

- RMSE = ASE: variability and validity

Where: $Z(x_i)$ *=predicted value,* $z(x_i)$ *=observed value and* $\sigma_k^2(x_i)$ *=kriging variance.*

Table 3: Cross validation statistics

Models	ME	RMSE	RMSS	ASE
Sum of species (OK)	-0.2056	13.70	1.001	14.36
Sum of species x Slope (CCK)	0.7906	13.67	1.084	13.25
Sum of species x SE (CCK)	0.1776	13.88	1.045	14.21
Sum of species x SA (CCK)	0.3610	13.25	1.051	13.54
Sum of species x SE (RK)	-0.0234	10.27	-	-
Sum of species x Slope (RK)	0.0115	11.11	-	-

- Fairly unbias
- RK is optimal because of low RMSE
- low precision for OK and CCK but fairly valid

Cross validation conti…

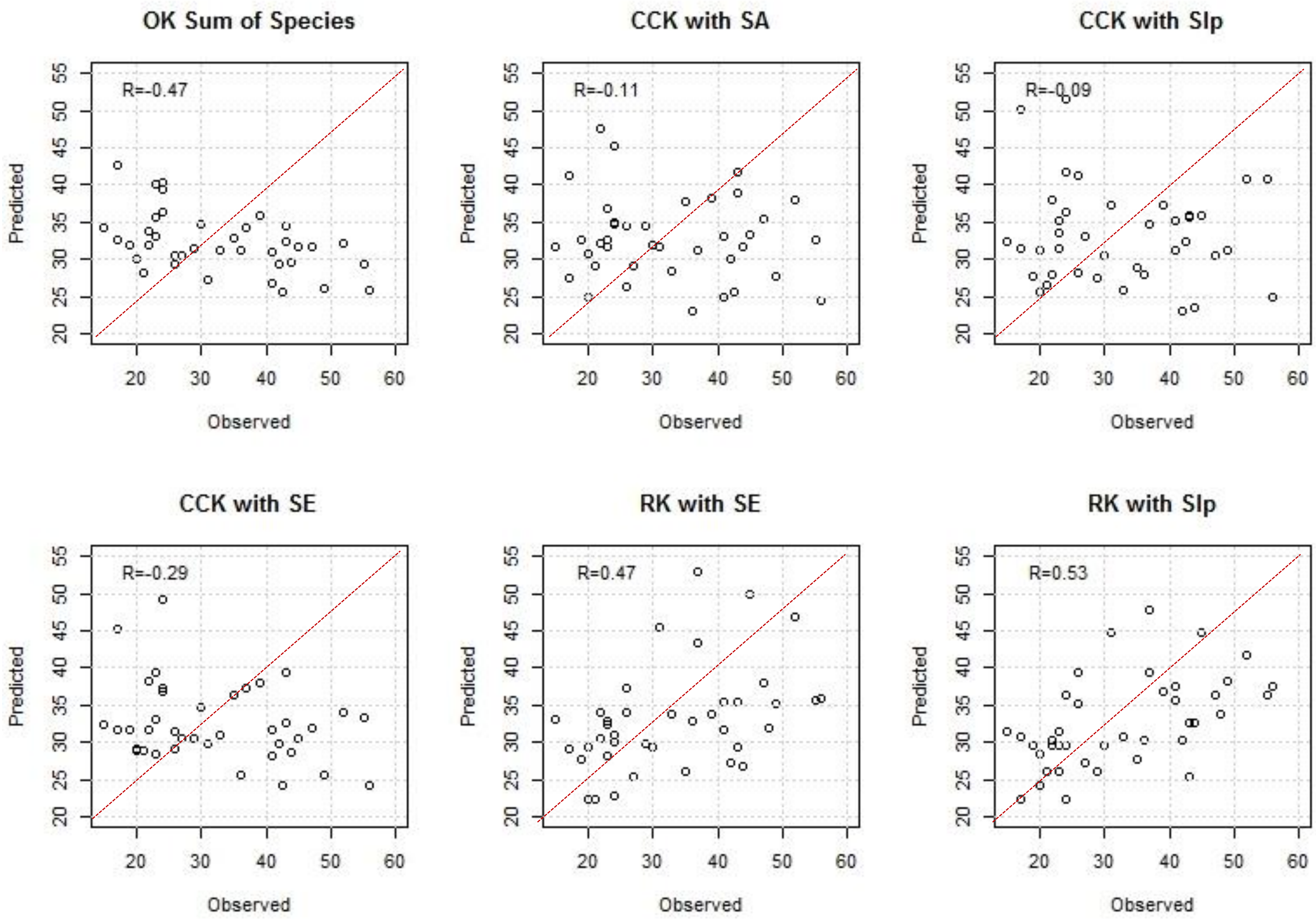


Figure 12: Comparison between observed and predicted species diversity based on OK, CCK and RK methods

Summary and conclusions

- The proposed methodology was inappropriate and somewhat misleading as it did not improve correlation with target variable

- Performance CCK was below expectation because of :
  - ✓ Difference in spatial structure between target and covariables, which made it difficult to fit appropriate coregionalized models
  - ✓ Topographic variations

- Regression kriging with slope + elevation was optimal, more flexible and robust to topographic variations than CCK

- Regression kriging should seriously be considered if two or more variables are to be used for cokriging

- A little more sampling could espcially improve results CCK results

- Based or RK with slope + elvation ,there is probably an ongoing succession of species in the south and northwest as opposed to well established species in other areas

References

- **Babak O. Deutch C. 2009**: Improved spatial modeling by merging multiple secondary data for intrinsic collocated cokriging. Journal of Petroleum Science and Engineering 69 (2009) 93–99

- **Bivand R. Pebesma E. Gómez-Rubio V. 2013** .Applied Spatial Data Analysis with R. Series, Springer New York Heidelberg Dordrecht London, 2nd ed. 2013. ISBN 978-1-4614-7618-4 (eBook)

- **Johnston K. Hoef J. Krivoruchko K. Neil l. 2001.** ArcGIS 9.3: Using ArcGIS Geostatistical Analyst. (ESRI open source document)

- **Hengl T. Heuvelink B. Stein A. 2004a**. A generic frame work for spatial prediction of soil variables based on regression kriging. Geoderma 122 (1-2): 75–93.

- **Krivoruchko K. 2011.** Spatial Statistical Data Analysis for GIS Users (chapter 6 and 8) Esri Press, 380 New York Street, Redlands, California 92373-8100

- **Webster R. Oliver A. 2007**. Geostatistics for Environmental Scientists. England, John Wiley & Sons Ltd. 2001.00393.x

Thank you for your attention!