# Geostatistical analysis in ecological Studies

Taibanganba Watham
Scientist/Engineer -SD and
Forestry and Ecology Department
taibang@iirs.gov.in

# Geostatistics

Geostatistics is a subset of statistics specialized in analysis and interpretation of geographically referenced data.

- How does a variable vary in space?
- What controls its variation in space?
- How many samples are needed to represent its spatial variability?
- What is a value of a variable at some new location?
- What is the uncertainty of the estimate?

But the main use of geostatistics is to predict values of a sampled variable over the whole area of interest, which is referred to as spatial prediction or spatial interpolation

# Environmental variables

Environmental variables are the quantitative or descriptive measures of different environmental features.

- Relevant and detailed geoinformation is a prerequisite for proper NRM
- Expensive
- Limited number of sampling locations
- We are often forced to build complete maps by using a sample of <1%

| Environmental features | Common variables of interest to decision making |
|---|---|
| Freshwater resources and water quality | DO, BOD, concentration of herbicides; trends in concentrations of pollutants; temperature change; |
| Land degradation: erosion, landslides, surface runoff | soil loss; erosion risk; quantities of runoff; dissolution rates of various chemicals; landslide susceptibility; |
| Soil fertility and productivity | organic matter, nitrogen, phosphorus and potassium in soil; biomass production; yields; leaf area index; |
| Distribution of animal species (wildlife) | occurrence of species; biomass; animal species density; biodiversity indices; habitat conditions; |
| Distribution of natural vegetation | land cover type; vegetation communities; occurrence of species; biomass; density measures; species richness; habitat conditions; |
| Climatic conditions and changes | mean, min and max temperature; monthly rainfall; wind speed and direction; total incoming radiation; |
| Air quality in urban areas | NOx, SO2 concentrations; emission of greenhouse gasses; ozone concentrations; Air Quality Index; |
| Global and local sea conditions | chlorophyll concentrations; biomass; sea surface temperature; emissions to sea; |

# Spatial prediction models

Everything is related to everything else. But near things are more related than distant things.

From the statistical perspective, an environmental variable can be viewed as an information signal consisting of three components.

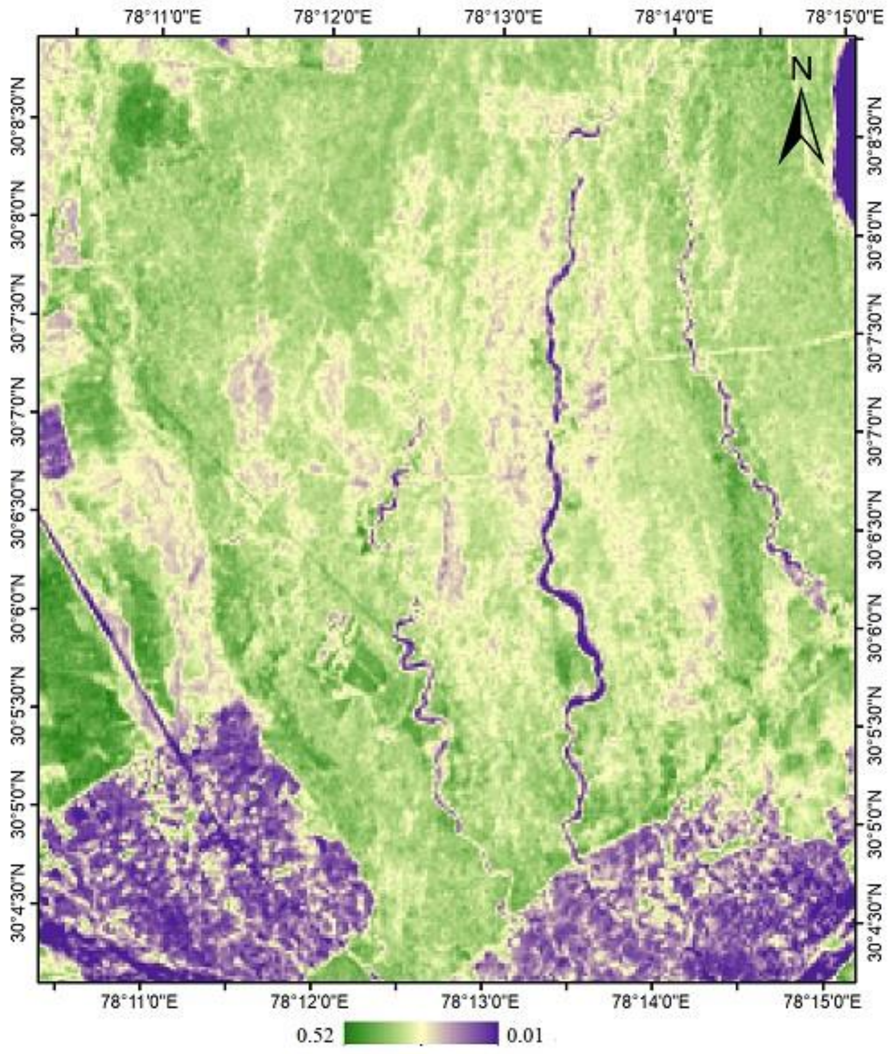
$$Z(s) = Z^*(s) + \varepsilon'(s) + \varepsilon''$$

Where,

$Z^*(s)$ is the deterministic component,

$\varepsilon'(s)$ is the spatially correlated random component

$\varepsilon''$ is the pure noise, usually measurement error.



# Spatial prediction models

Successful geostatistical mapping: To understand the sources of variability in the data.

The variability in data is a sum of two components:

(a) the natural spatial variation

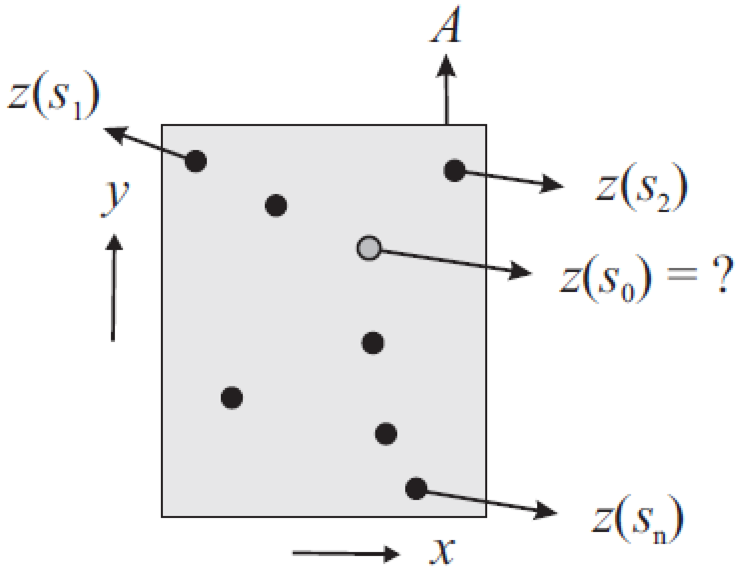
(b) the inherent noise

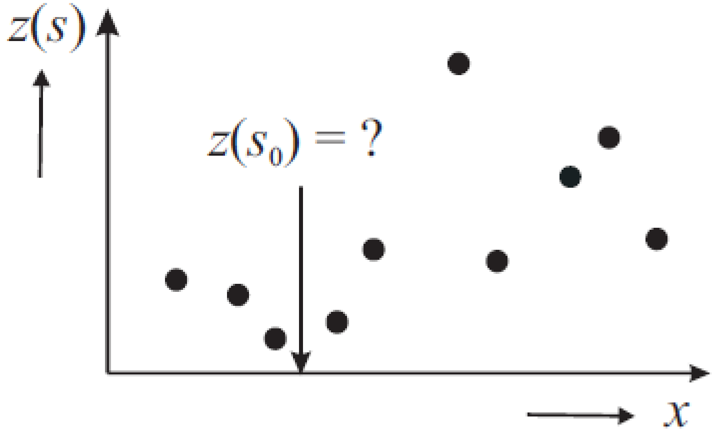
Interested in is the natural spatial variation, which is mainly due to the physical processes that can be explained by a mathematical model (up to a certain level).

To consider all aspects of natural variation.

# Spatial prediction models

Assuming that the samples are representative, unbiased and consistent, values of the target variable at some new location $s_0$ can be derived using a spatial prediction model.




$$\hat{z}(s_0) = E\{Z|z(s_i), qk(s_0), \gamma(h), s \in A$$

where $z(s_i)$ is the input point dataset, $qk(s_0)$is the list of deterministic predictors and $\gamma(h)$is the covariance model defining the spatial autocorrelation structure.

# Spatial prediction models

Classification based on the amount of statistical analysis included

| MECHANICAL/ EMPIRICAL MODELS | STATISTICAL (PROBABILITY) MODELS |
|---|---|

- Inverse distance Weighting (IDW) interpolation
- Thiessen polygons
- Splines

- Environmental correlation
- kriging (plain geostatistics)
- Mixed models (Regression-kriging)

Support size: Point/Block

Proximity effect: Local/Global

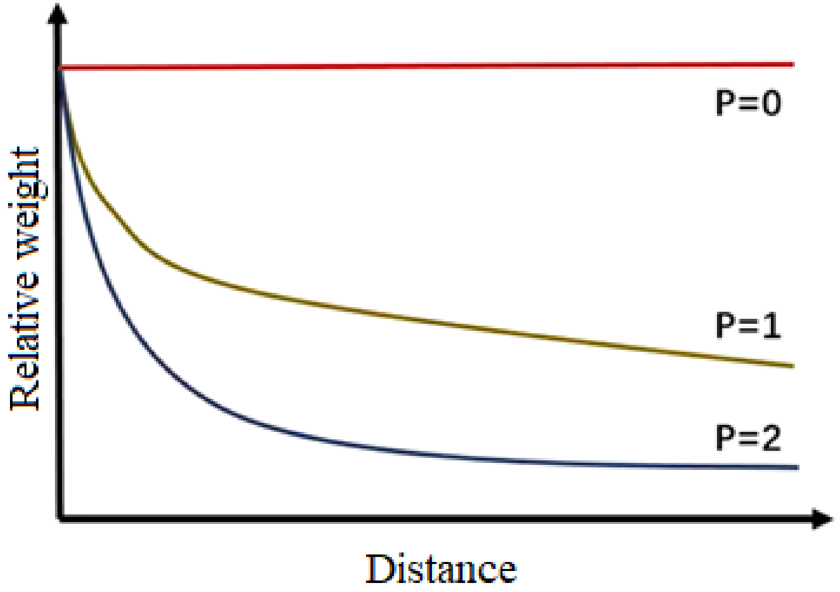
Convexity effect: Convex/Non-convex

Smoothing effect: Exact/Approximate

# Inverse distance Weighting interpolation (IDW)

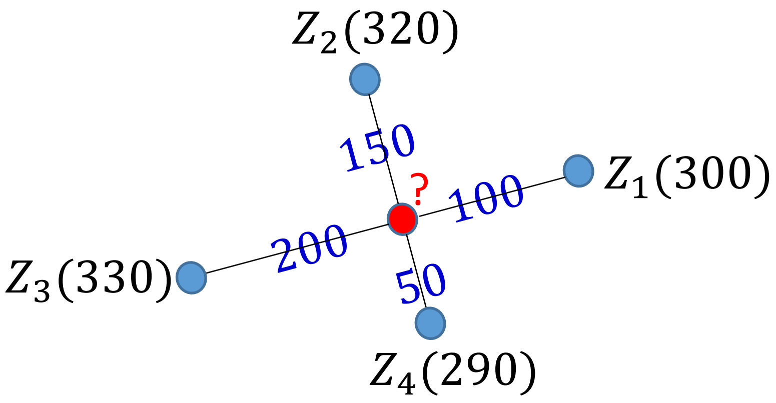
- One of the oldest spatial prediction technique
- This method is based on the assumption that value at any point is influenced by its neighboring points and influence (weights) is inversely proportional to distance.
- Weight is decided by a weighting function,

$$W = \frac{1}{d^p}$$



# Inverse distance interpolation (IDW)

$$Z_p = \frac{\sum_{i=1}^{n}\left(\frac{Z_i}{{d_i}^p}\right)}{\sum_{i=1}^{n}\left(\frac{1}{{d_i}^p}\right)}$$




$Z_p$= Predicted value
$Z_i$ = Value at $i$ measured point
$d_i$ = distance of $i$ measured location to predicted point
$p$ = power function
$n$ = number of points to be used

$$Z_p = \frac{(300/100)^2 + (320/150)^2 + (330/200)^2 + (290/50)^2}{(1/100)^2 + (1/150)^2 + (1/200)^2 + (1/50)^2}$$

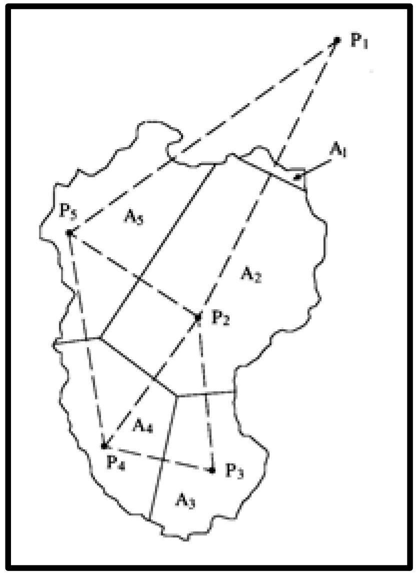
$$Z_p = \frac{0.03 + 0.014222222 + 0.00825 + 0.116}{0.0001 + 0.00004 + 0.000025 + 0.0004}$$

$$Z_p = 295.85$$

- IDW is exact interpolation method
- For large dataset it will take more time
- For the weight we need to define some 'p' i.e. power value

# Thiessen polygons

- Prediction at unsampled location is provided by nearest single data point.
- Delaunay Triangulation rule says that 'there should be no point inside the circumcircle of any triangle'
- Draw the perpendicular bisector for the sides of the triangles formed
- Extend (or shorten) bisectors to connect forming polygons surrounding each sample location
- Now you can notice that each Thiessen polygon contains only one station
- Each polygon will be given the value of the sample location inside it.




Thiessen polygons

# Environmental correlation

If some exhaustively-sampled auxiliary variables or covariates are available in the area of interest and if they are significantly correlated with our target variable (spatial crosscorrelation), and assuming that the point-values are not spatially auto-correlated, predictions can be obtained by focusing only on the deterministic part of variation.

$$\mathrm{W} = f\{q(s)\} + \epsilon$$

where $q$ are the auxiliary predictors that can be used to explain the deterministic part of spatial variation.

$$\mathrm{V} = f\begin{Bmatrix} S(x,y,t)C(x,y,t)O(x,y,t) \\ R(x,y,t)P(x,y,t)A(x,y,t) \end{Bmatrix}$$

# Kriging

Problem with the deterministic interpolation is that there is no priori method of knowing that whether the best weights have been chosen.

- How many points should be taken to compute the local average.
- Whether best weights are estimated.
- What is the error / uncertainty associated with interpolated values.

Kriging estimates are best linear unbiased estimates (BLUE) of the ReV at a set of locations, provided that the surface is stationary and the correct form of the theoretical variogram has been determined

# Kriging

Stationarity, isotropy, intrinsic hypothesis and unbiasness are the basic assumptions of Geostatistical techniques (Sluiter, R., 2009)

Stationarity: same probability distribution function is everywhere in study region i.e statistical properties do not change with distance

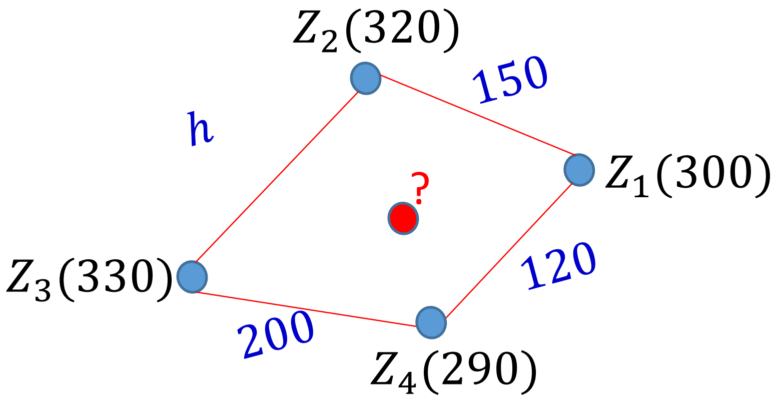
Intrinsic hypothesis: difference between two values taken at two different locations come from a distribution which depends only on the distance (and possibly relative direction) of the two locations

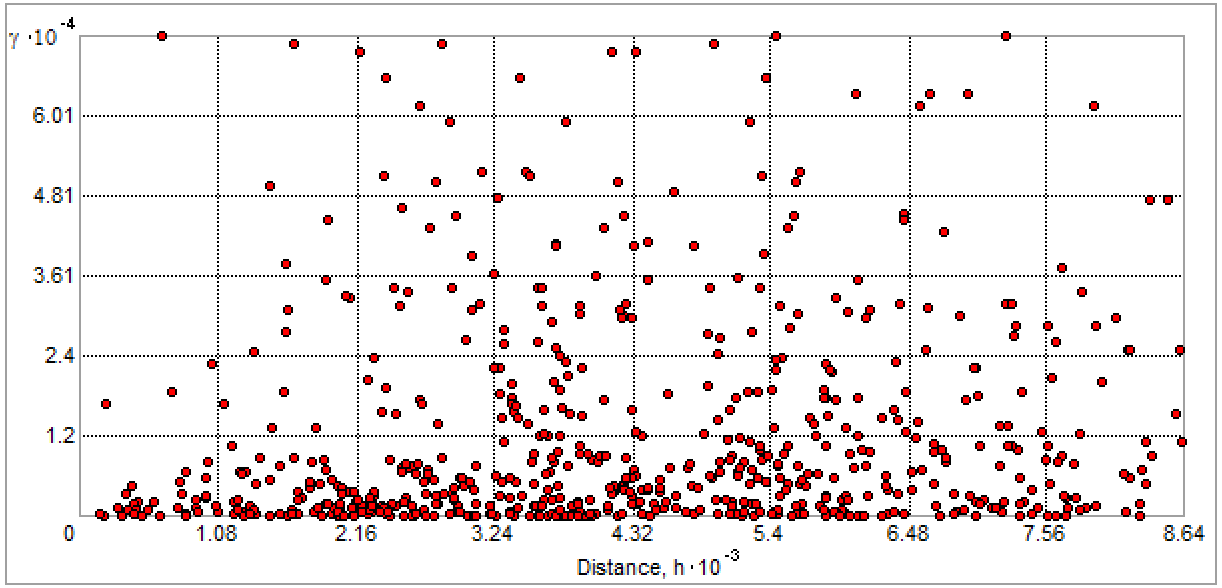
Isotropy: change in target variable is directionally independent

Unbiasness: all sample values are having equal importance for interpolation

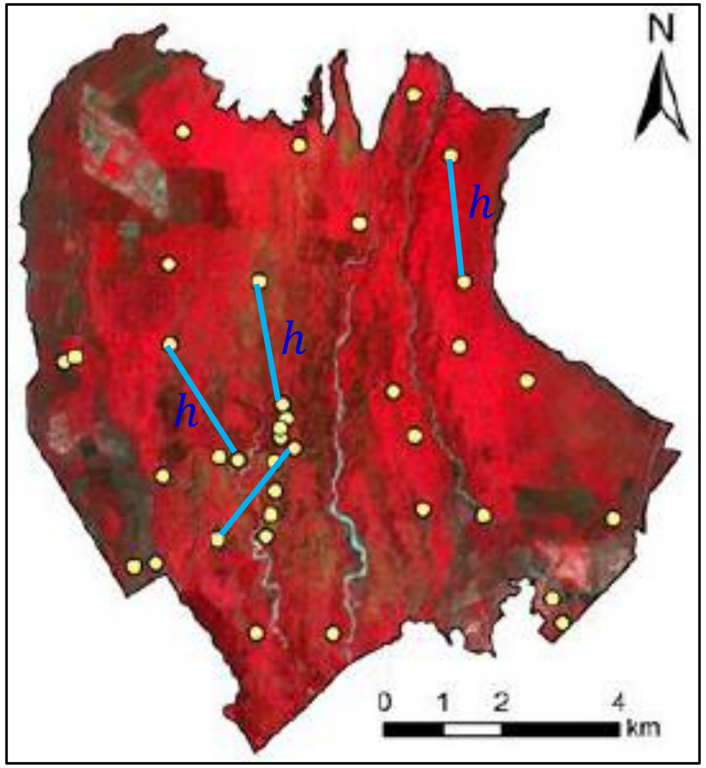
# Spatial auto-correlation

$$\gamma\left(\vec{h}\right) = \frac{1}{2n(\vec{h})}\sum_{i=1}^{n(\vec{h})}\left(z_{i+\vec{h}} - z_i\right)^2$$

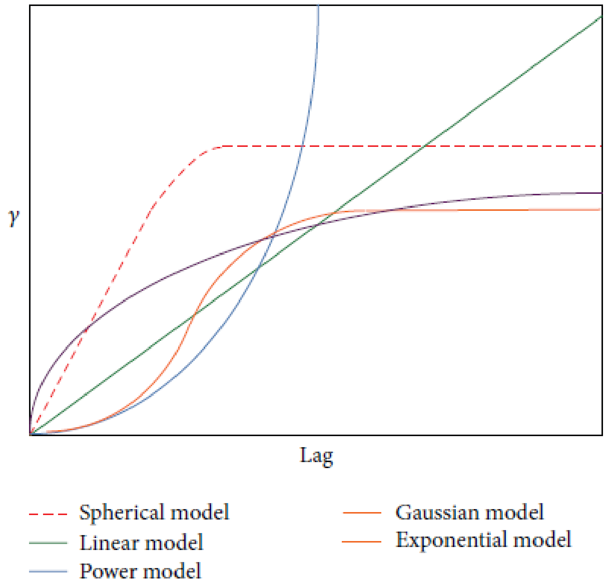


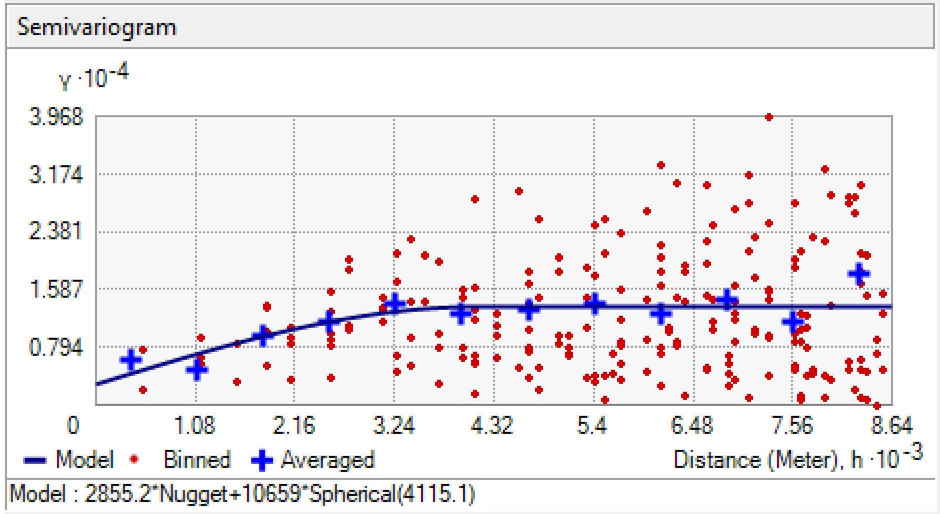



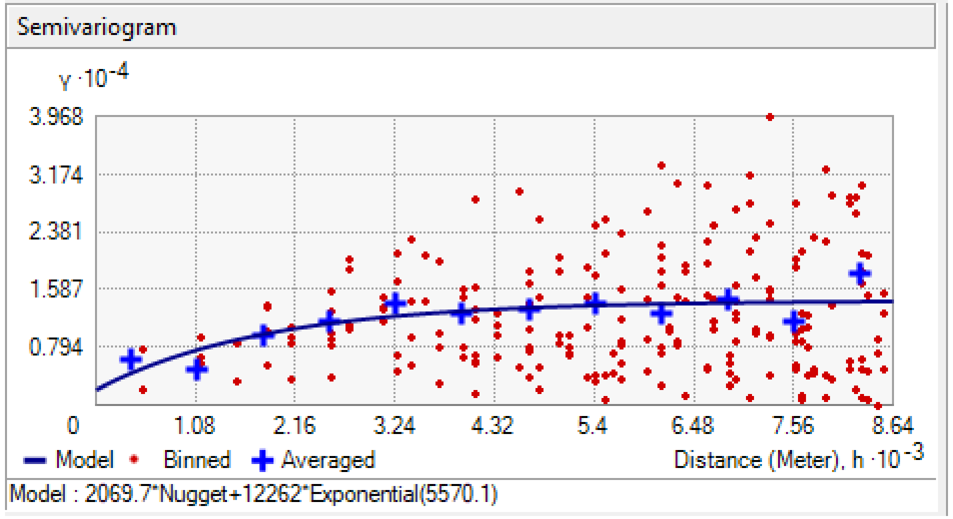

Variogram cloud

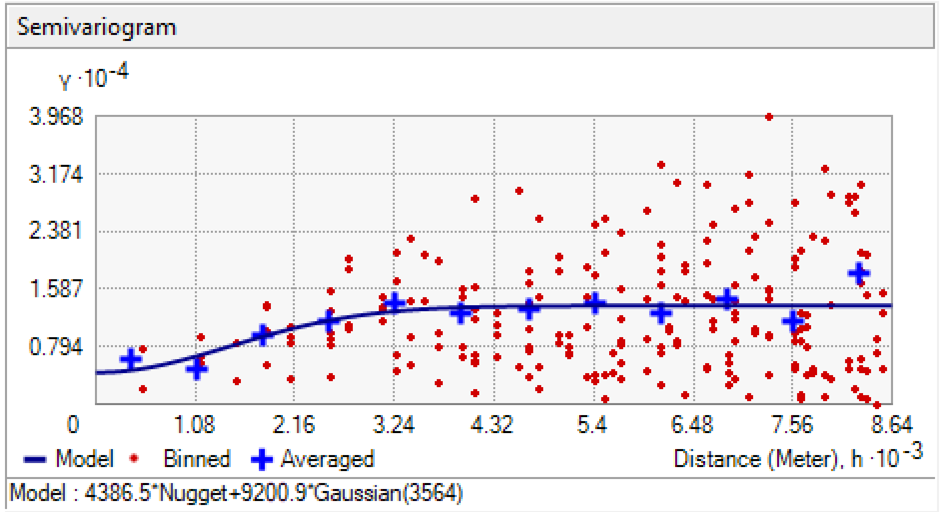


# Variogram

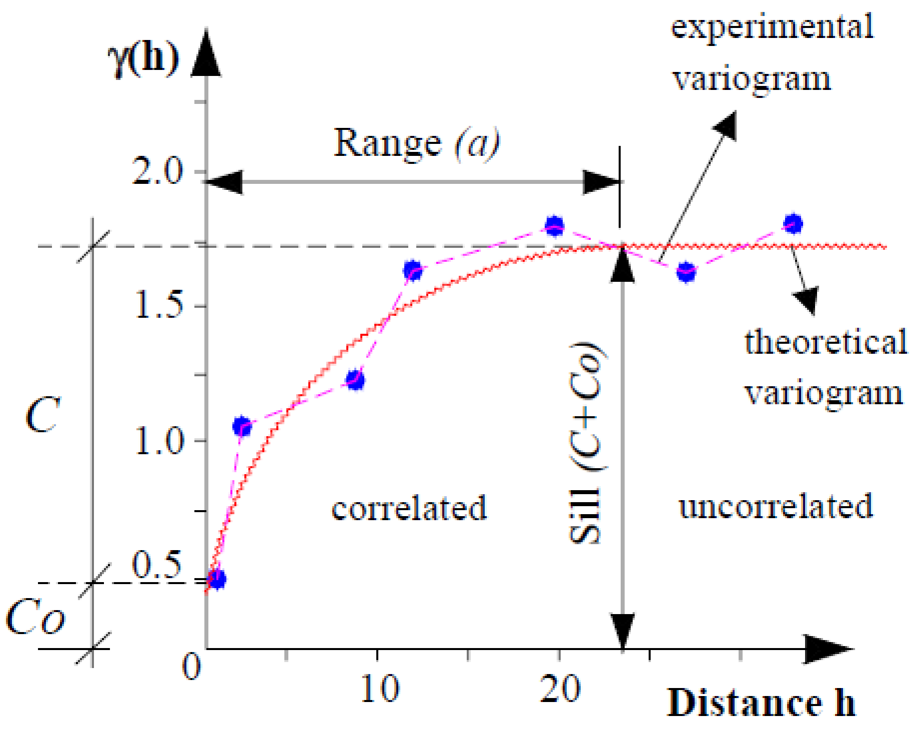

γ
Lag
Spherical model
Linear model
Power model
Gaussian model
Exponential model


Semivariogram
γ ·10^-4
3.968
3.174
2.381
1.587
0.794
0
1.08
2.16
3.24
4.32
5.4
6.48
7.56
8.64
Model
Binned
Averaged
Distance (Meter), h ·10^-3
Model : 2855.2*Nugget+10659*Spherical(4115.1)


Semivariogram
γ ·10^-4
3.968
3.174
2.381
1.587
0.794
0
1.08
2.16
3.24
4.32
5.4
6.48
7.56
8.64
Model
Binned
Averaged
Distance (Meter), h ·10^-3
Model : 2069.7*Nugget+12262*Exponential(5570.1)


Semivariogram
γ ·10^-4
3.968
3.174
2.381
1.587
0.794
0
1.08
2.16
3.24
4.32
5.4
6.48
7.56
8.64
Model
Binned
Averaged
Distance (Meter), h ·10^-3
Model : 4386.5*Nugget+9200.9*Gaussian(3564)

# Variogram



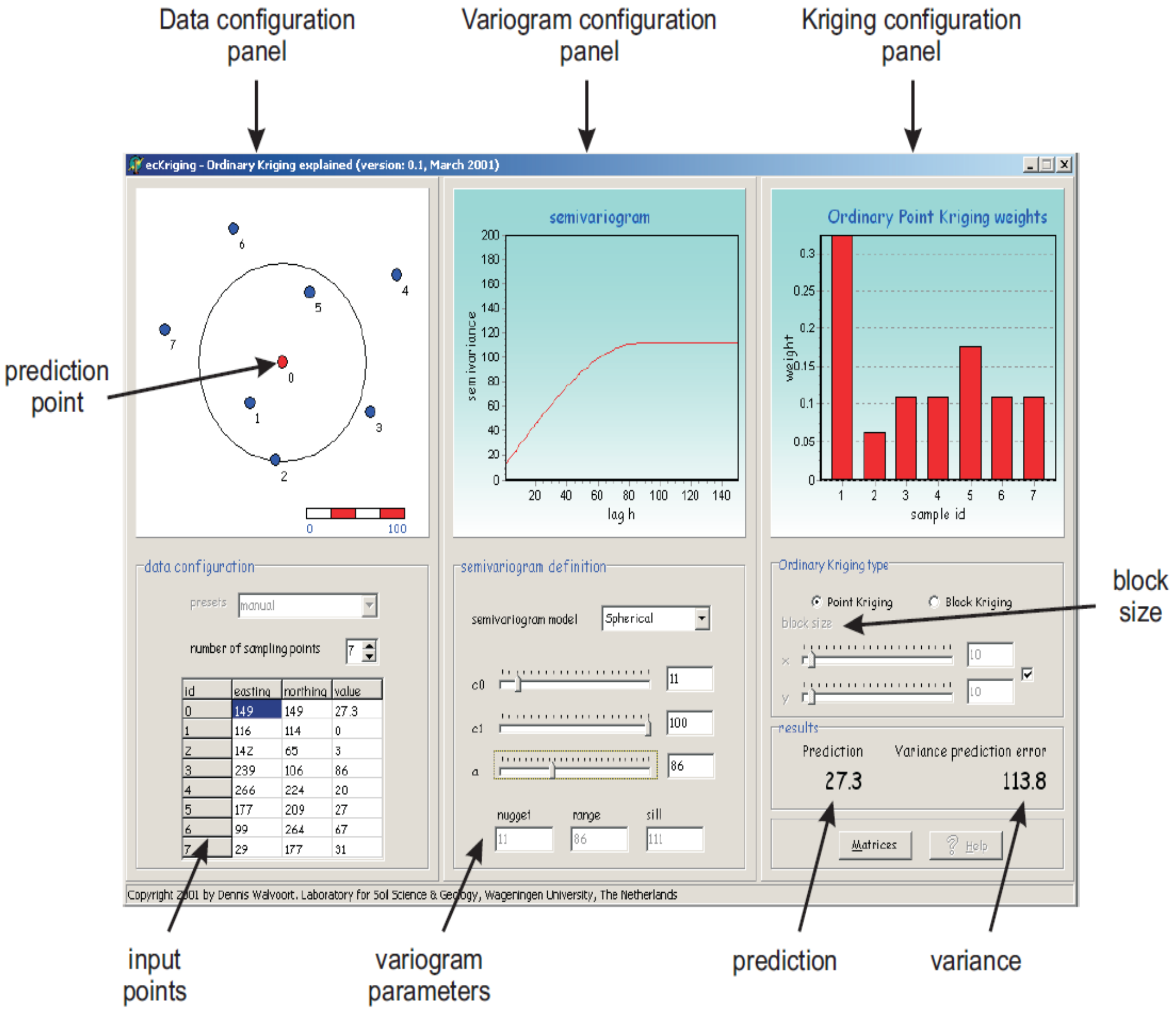

**Range (*a*):** indicates the distance between locations beyond which observations appear independent, i.e. variance no longer increases.

**Nugget (C0):** describes the unexplained variance of the variable when modelled for very short distances between samples, i.e. shorter than the sampling distance.

**Sill (C + C0):** is the value of the variation chart, which corresponds to its range (a). From this point forward, one assumes that there is no more spatial dependence between the samples, as variance of the differences between the pairs of samples becomes constant with distance.

# Variogram & Kriging

Kriging uses the variogram model to compute the weights of neighboring points based on the distribution of those values kriging is letting the localized pattern produced by the sample points define the weights (in a systematic way).




Source: EZ-Kriging(Dennis J.J. Walvoort)

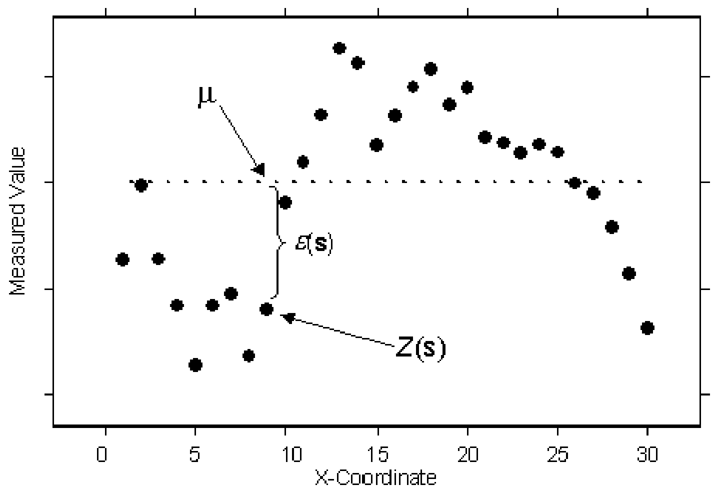
# Kriging types

## 1. Ordinary Kriging

## 2. Simple Kriging

$$Z(\mathbf{s}) = \mu + \varepsilon(\mathbf{s})$$

where μ is a unknown constant

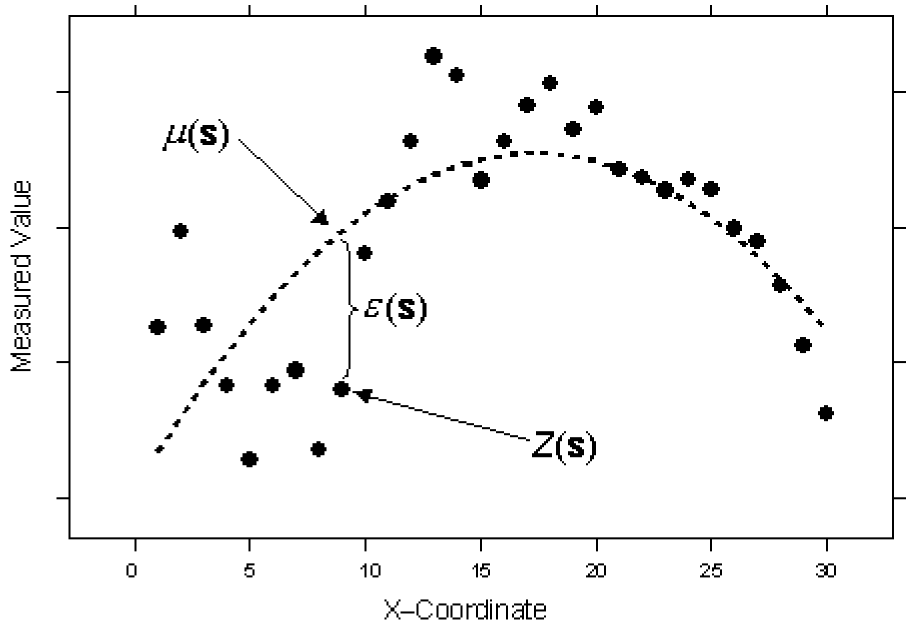



where μ is a known constant

# 3. Universal Kriging

- If there is trend present in the data; i.e. condition of stationarity is not satisfied
- Then, we choose we use another technique for prediction named 'Universal Kriging'
- Local mean is also not constant
- Trend surface is fitted for mean
- Any order polynomial trend can be fitted
- Rest process is like OK

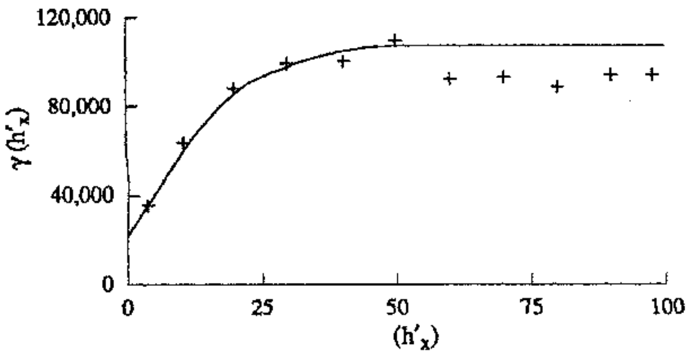
$$Z(s) = \mu + \varepsilon(s)$$

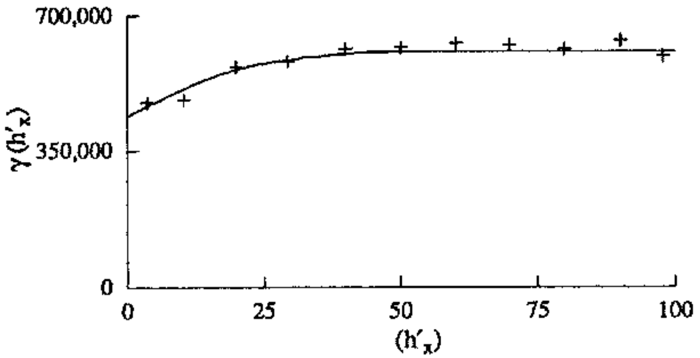


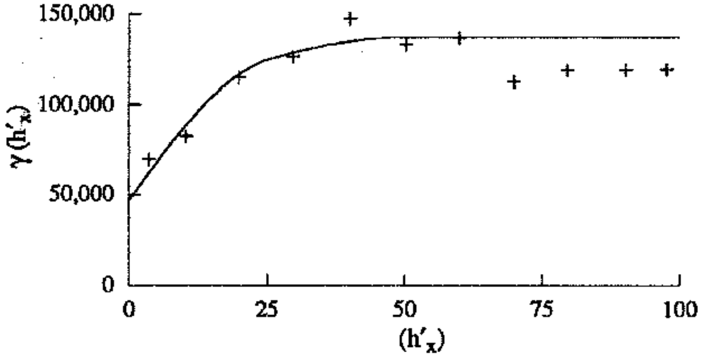

where $\mu(\mathbf{s})$ is some deterministic function

# 4. Co-Kriging

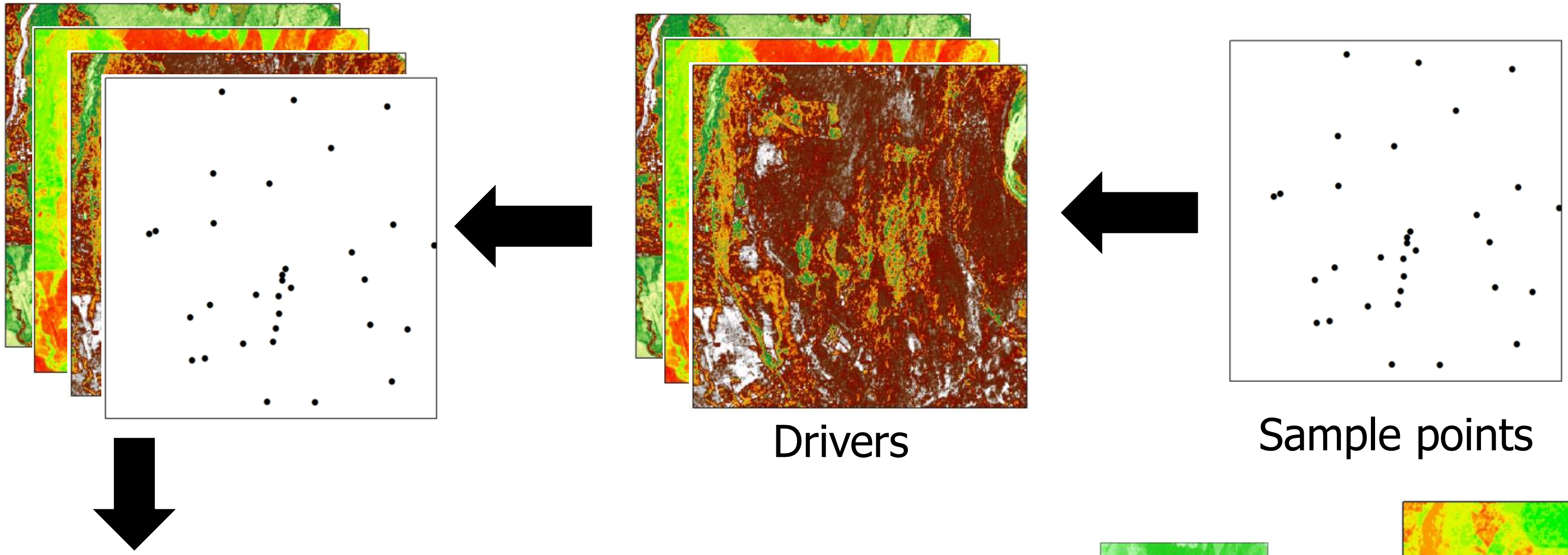
- Cokriging is an extension of ordinary kriging that uses a high spatial resolution secondary variable network to improve the estimation of the primary variable.
- Like ordinary kriging, a variogram model is developed for the primary variable then second variogram model is developed for the secondary variable.
- A third variogram model is generated from the cross correlation of the primary and secondary variables.
- The secondary data is "transformed" to the scale of the primary variable in order to derive a cross variogram. The secondary data are then treated identically to the primary data during spatial estimation.



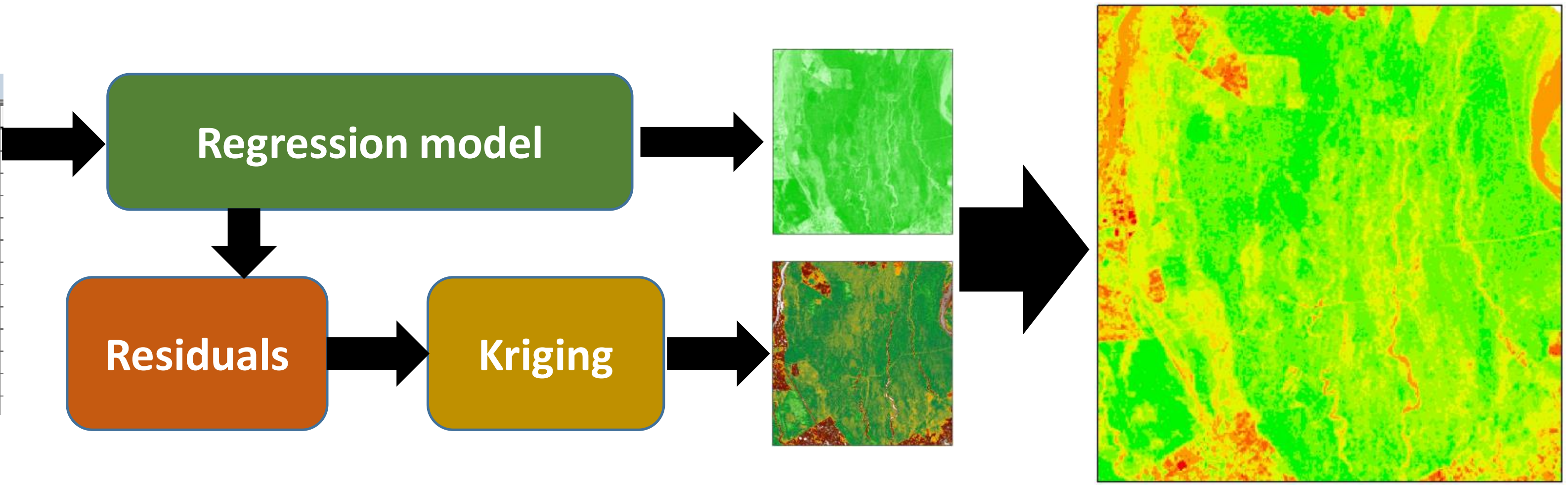



# Mixed model :Regression Kriging

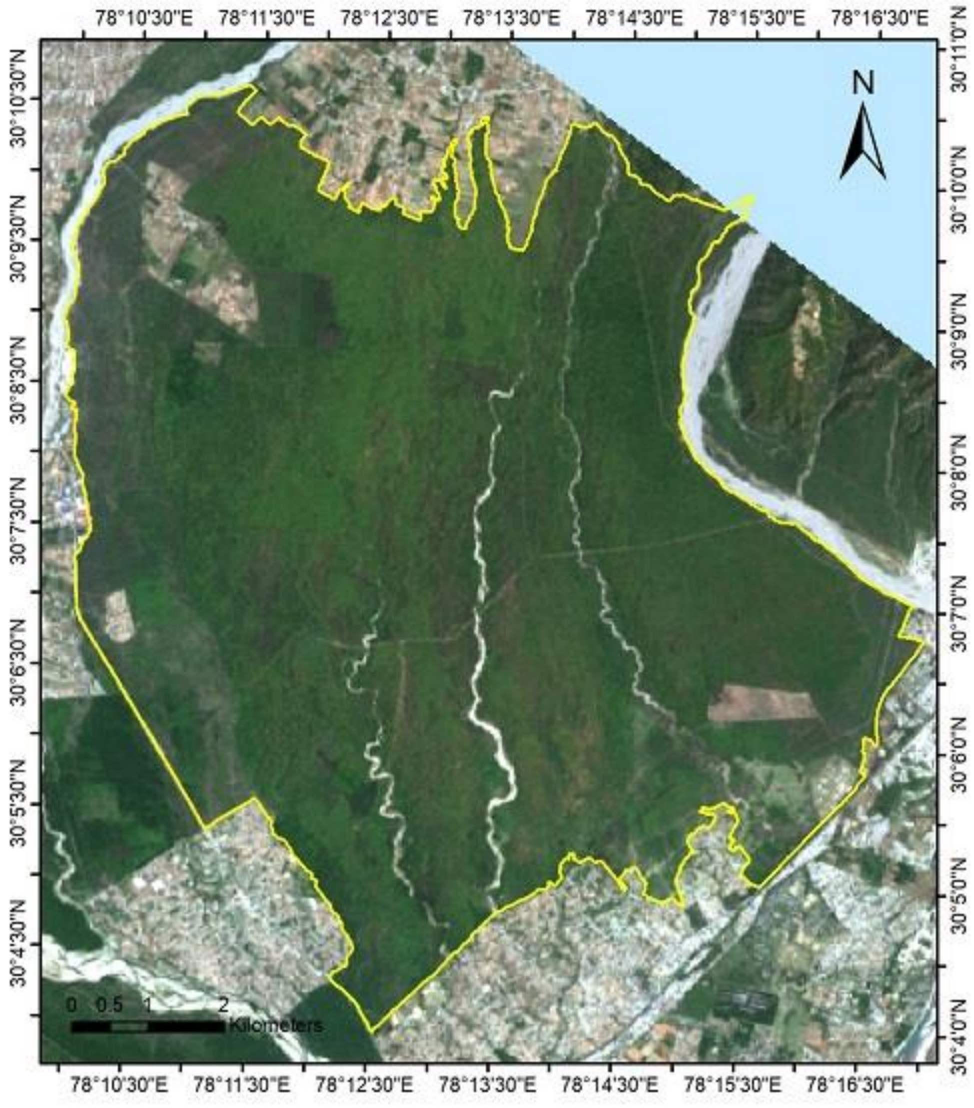


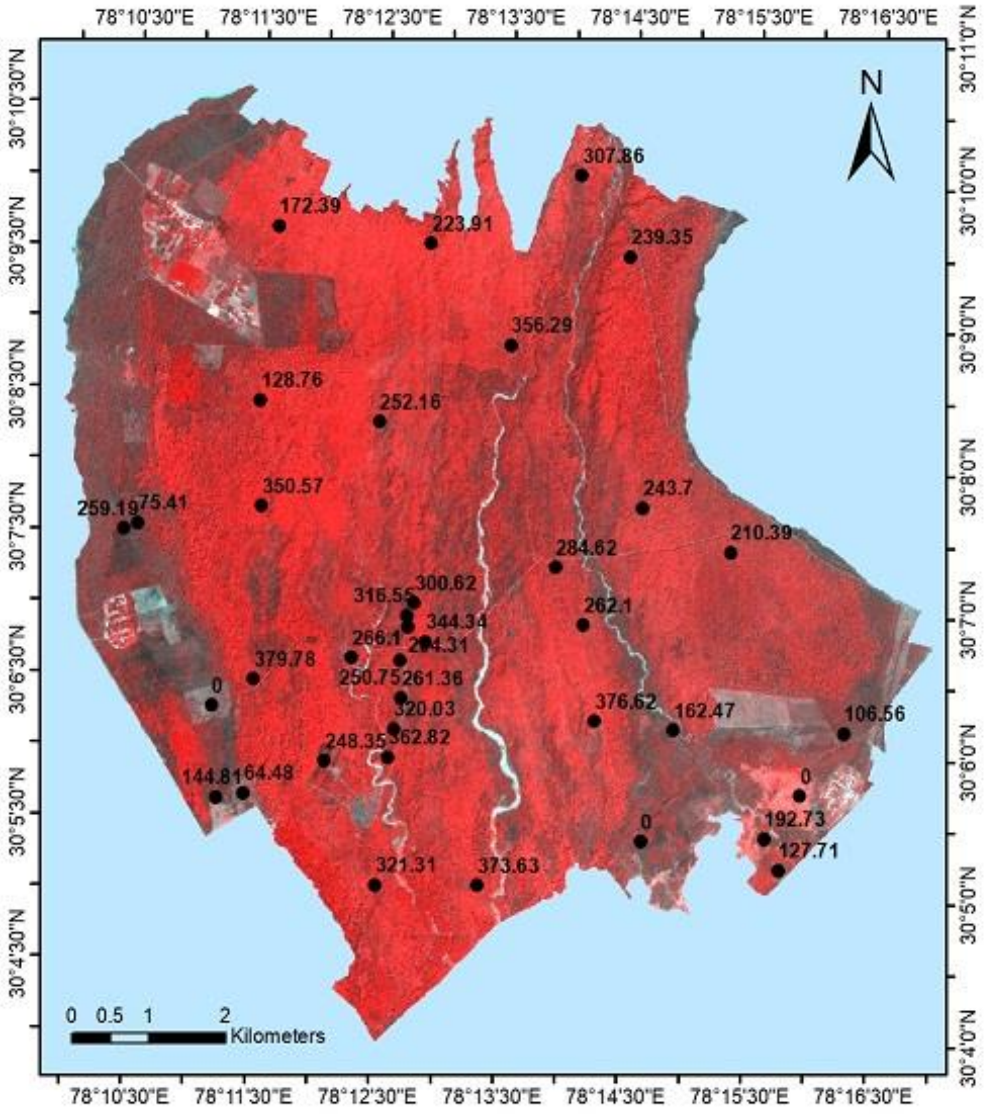

Point_biomass

| FID | PLOT_LOC_ | Biomass | PAC1 | EVI | NDVI |
|---|---|---|---|---|---|
| 0 | 1 | 533.511 | 0.291483 | 0.317946 | 0.390994 |
| 1 | 2 | 556.076 | 0.236732 | 0.436377 | 0.508417 |
| 2 | 3 | 680.905 | 0.237352 | 0.411407 | 0.48982 |
| 3 | 4 | 771.951 | 0.215542 | 0.465128 | 0.540807 |
| 4 | 5 | 683.628 | 0.189536 | 0.48581 | 0.55788 |
| 5 | 6 | 626.187 | 0.19282 | 0.456755 | 0.53131 |
| 6 | 7 | 673.516 | 0.183785 | 0.481988 | 0.555586 |
| 7 | 8 | 639.614 | 0.18108 | 0.47185 | 0.552231 |
| 8 | 11 | 732.628 | 0.192124 | 0.502488 | 0.567674 |
| 9 | 13 | 605.573 | 0.166864 | 0.582712 | 0.63597 |
| 10 | 14 | 557.657 | 0.171918 | 0.589029 | 0.630294 |
| 11 | 15 | 801.324 | 0.204966 | 0.506648 | 0.563465 |
| 12 | 17 | 566.178 | 0.24923 | 0.34999 | 0.429209 |

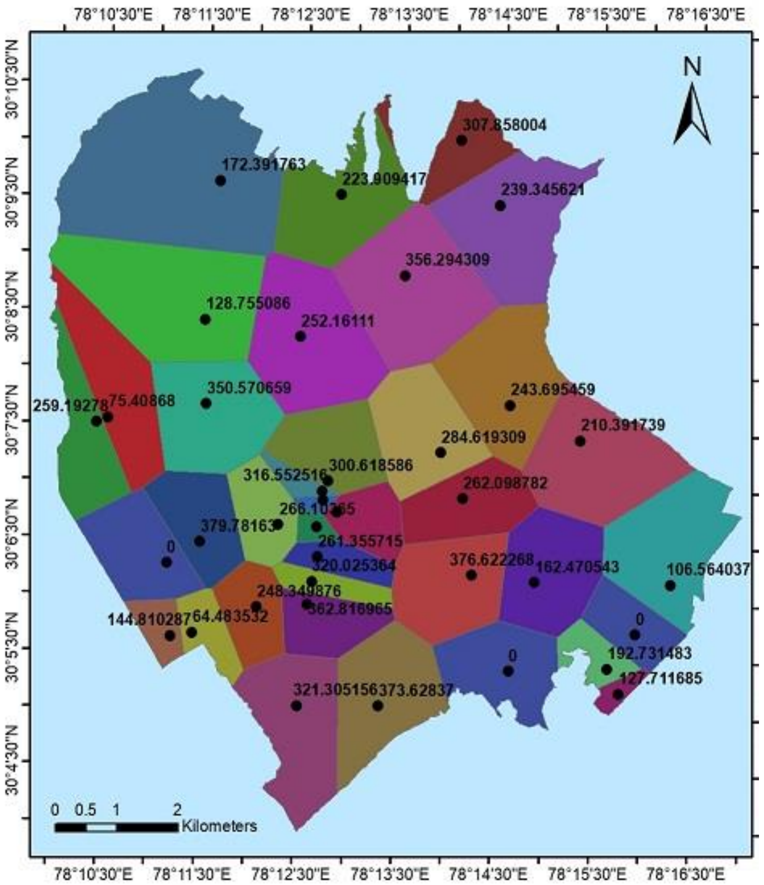


# Case Study: Total vegetation Forest carbon

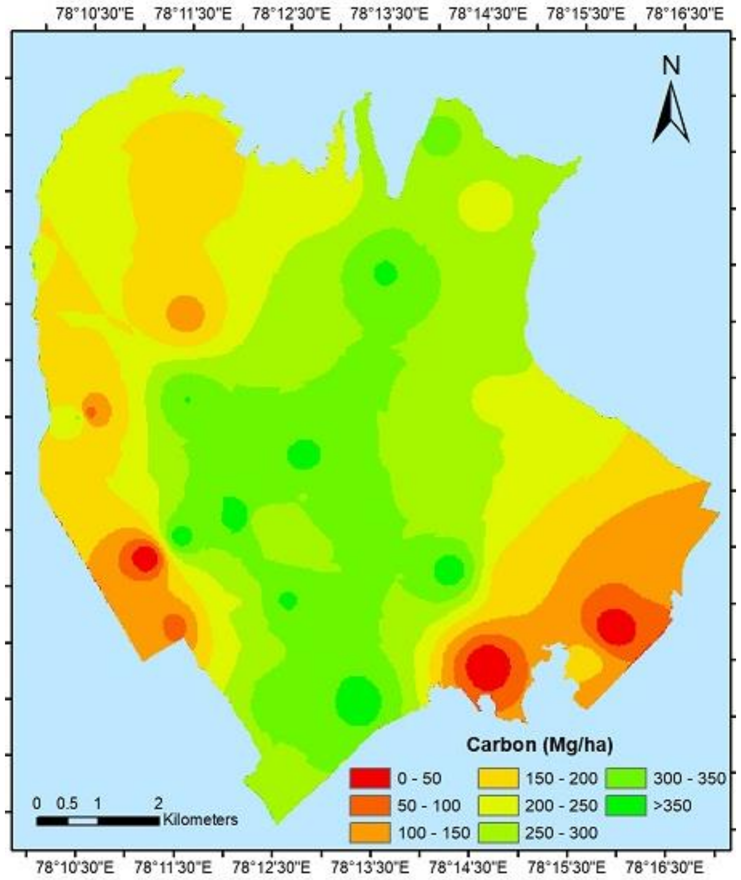


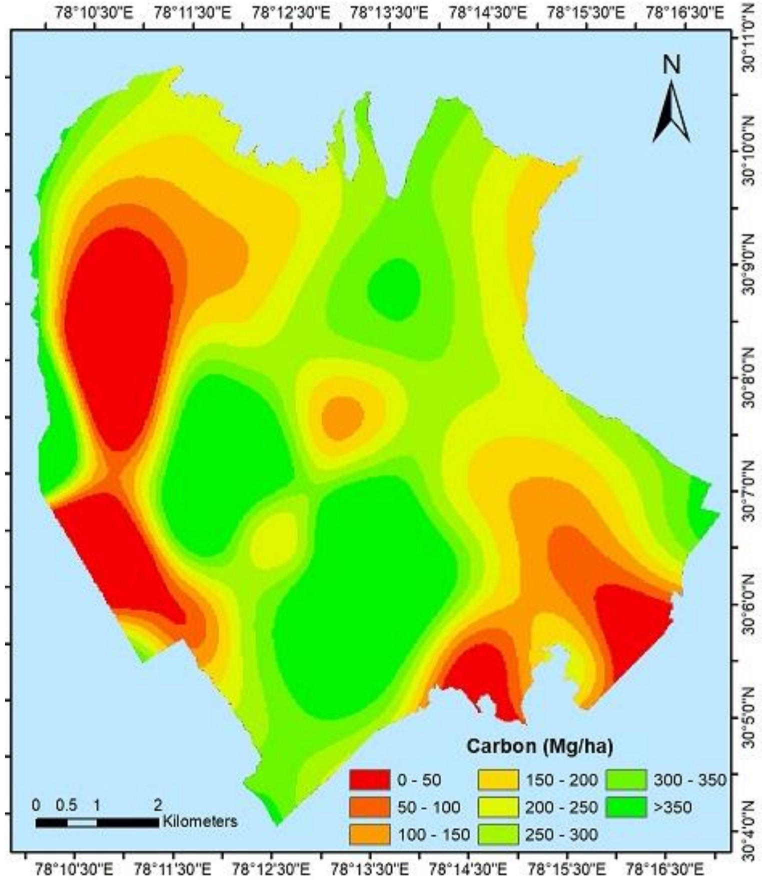
## Thiessen polygons



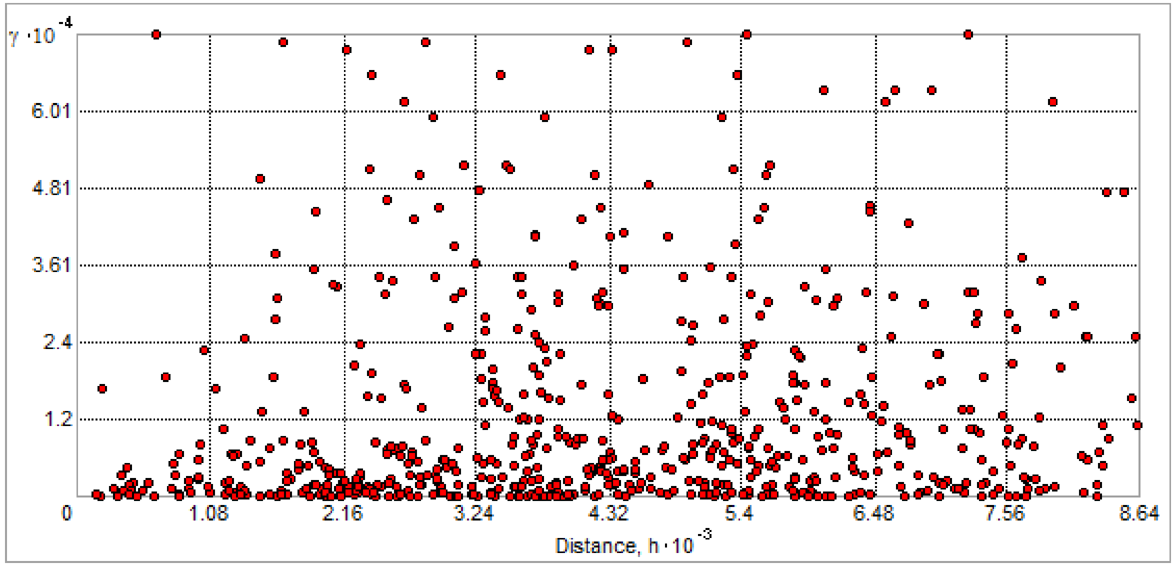

## IDW

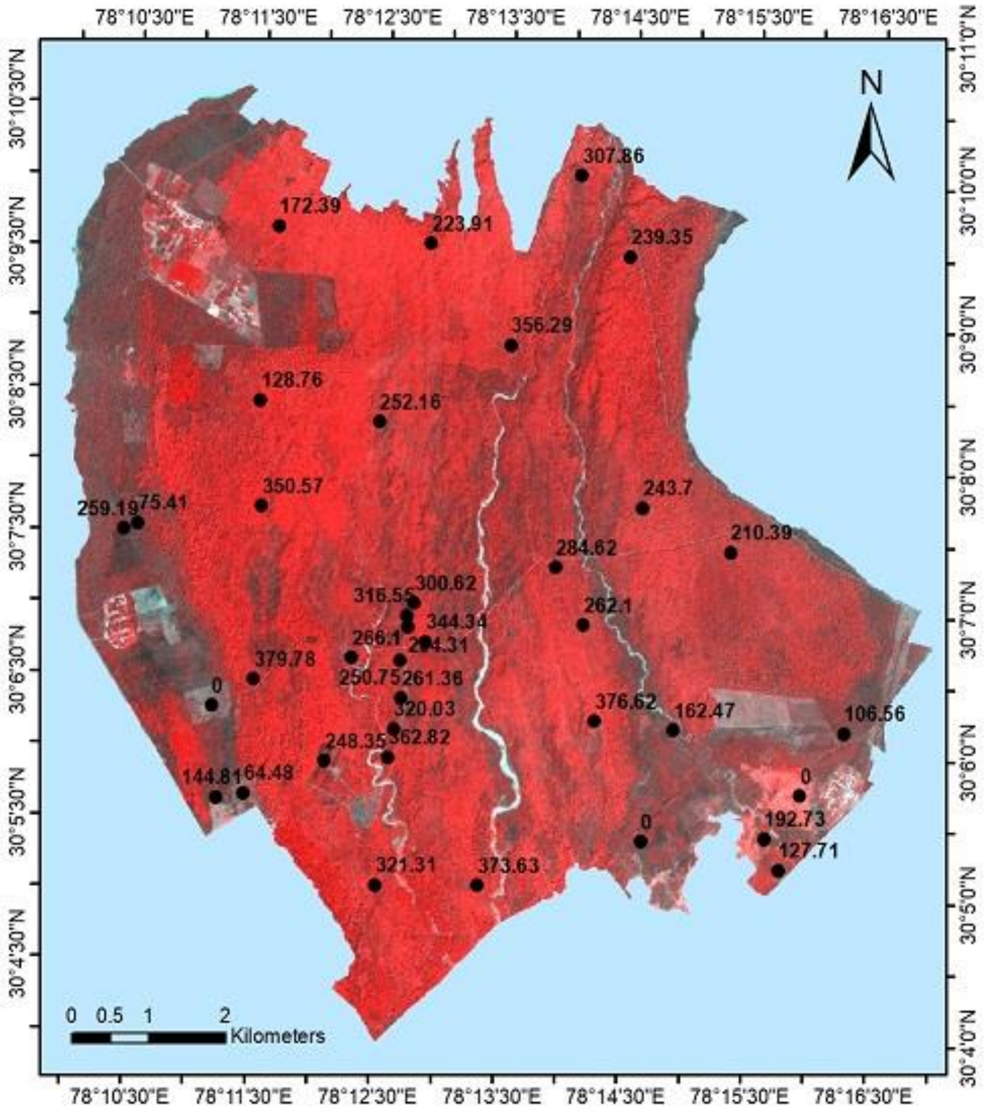



## Spline

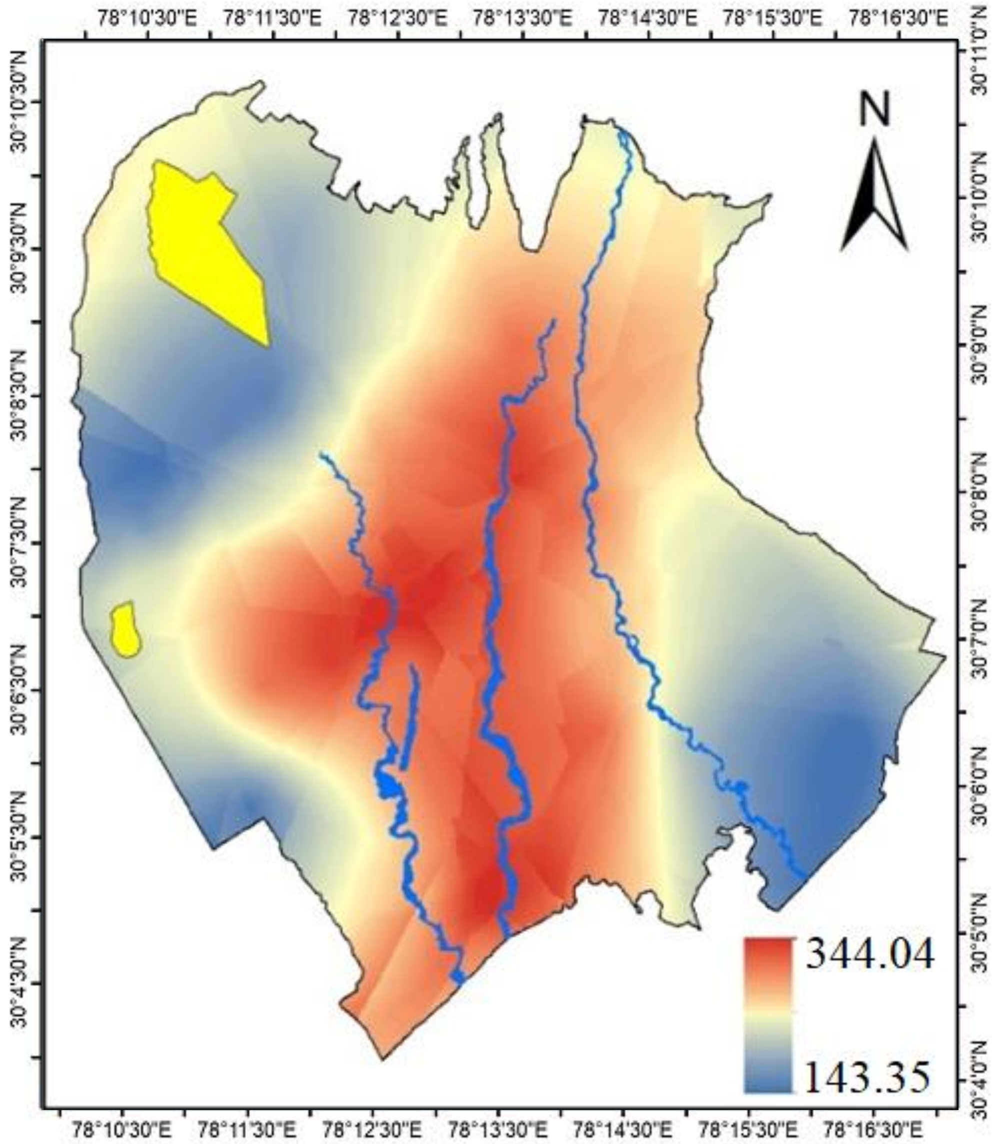


# Variogram Cloud

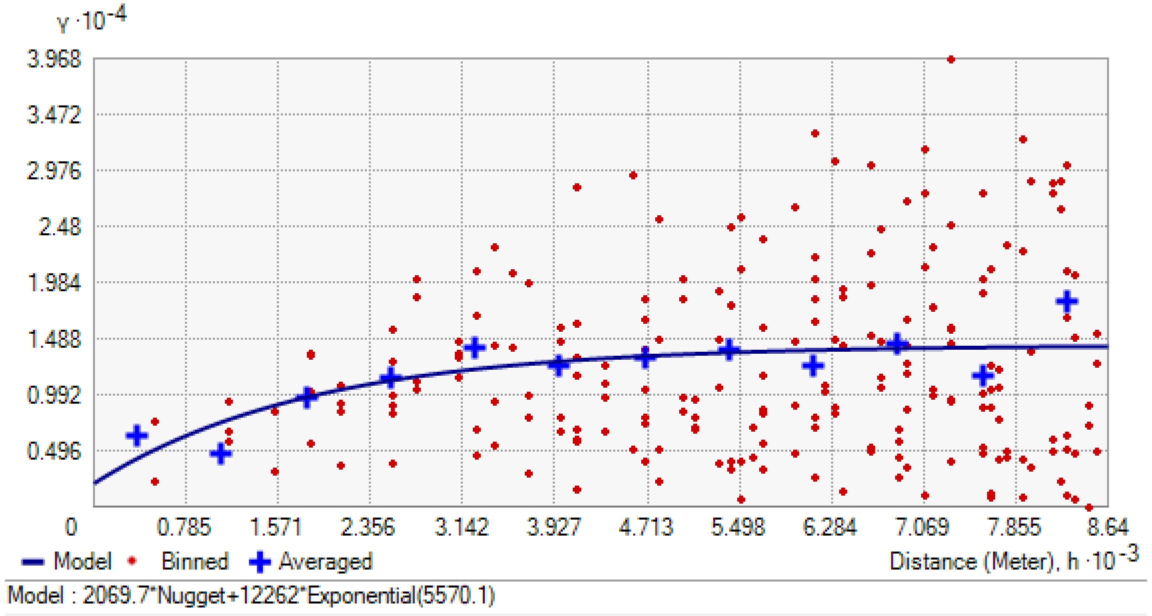


# Ordinary kriging

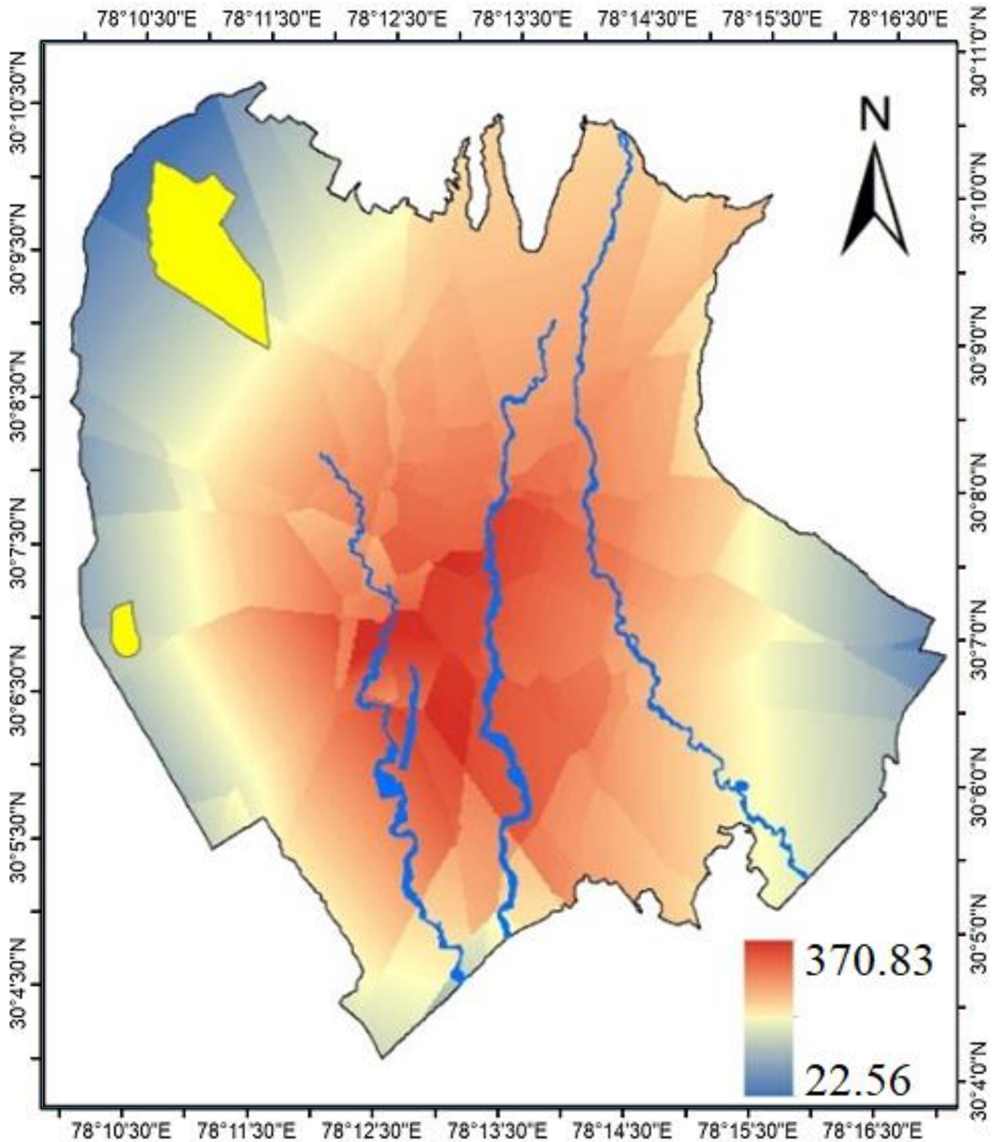



| | Nugget | Partial sill | RMSS |
|---|---|---|---|
| Ordinary kriging: | | | |
| Spherical | 0.365 | 0.350 | 1.02 |
| Exponential | 0.206 | 1.226 | 0.99 |
| Gaussian | 0.252 | 0.472 | 1.11 |
| Circular | 0.255 | 0.463 | 1.10 |

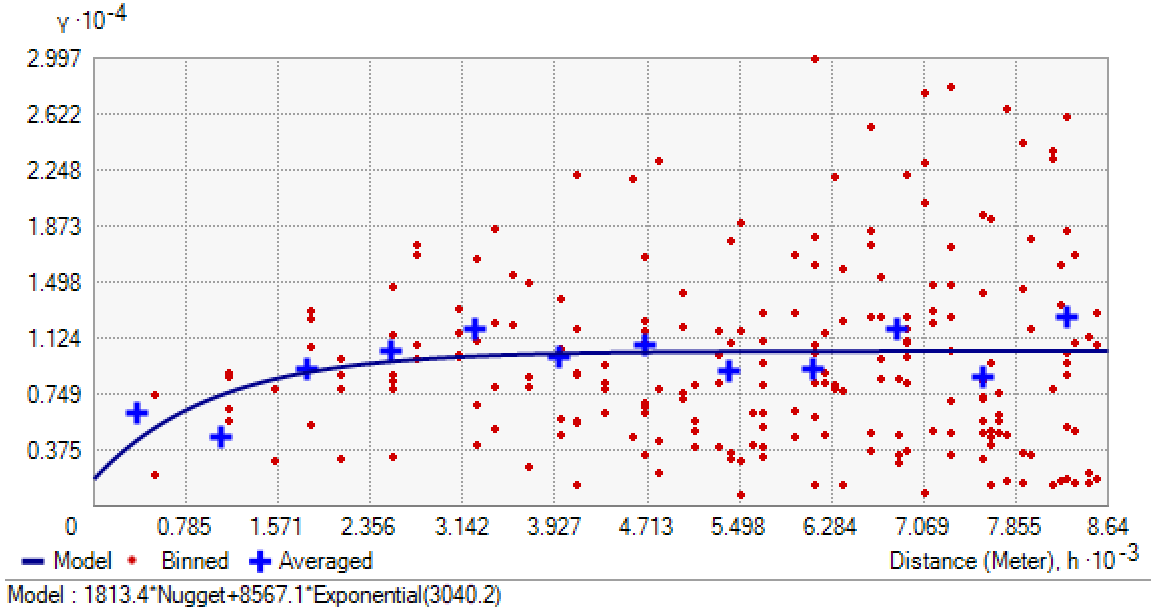



RMSE of 121.78 Mg $ha^{-1}$

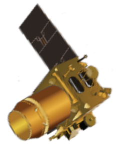
# Universal kriging




| | Nugget | Partial sill | RMSS |
|---|---|---|---|
| Universal kriging: | | | |
| Spherical | 0.229 | 0.120 | 1.32 |
| Exponential | 0.181 | 0.856 | 1.31 |
| Gaussian | 0.259 | 0.091 | 1.33 |
| Circular | 0.252 | 0.097 | 1.33 |

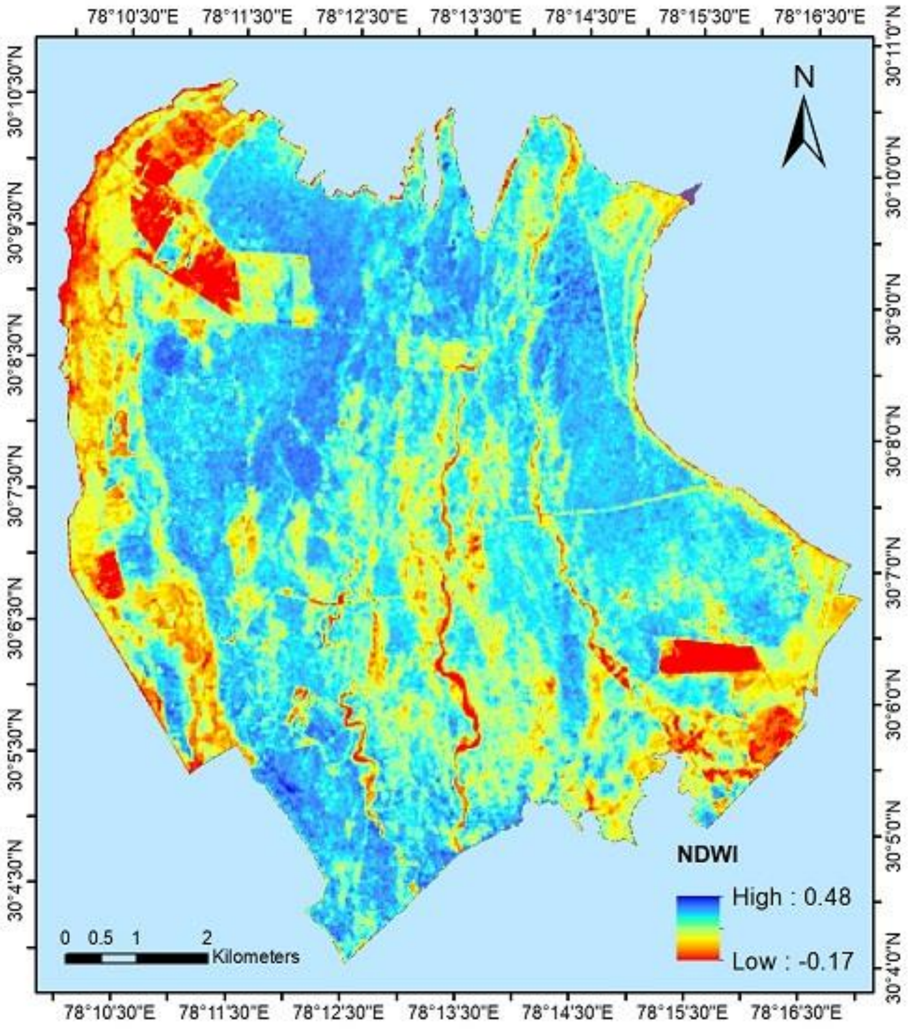



RMSE of 139.48 Mg ha$^{-1}$

# Environmental Correlation

| Indices | Formula |
|---|---|
| MidIR | $\frac{SWIR1}{SWIR2}$ |
| MSI | $\frac{SWIR1}{NIR}$ |
| NDWI1 | $\frac{NIR - SWIR1}{NIR + SWIR1}$ |
| NDWI2 | $\frac{NIR - SWIR2}{NIR + SWIR2}$ |
| NDVI | $\frac{NIR - Red}{NIR + Red}$ |
| WDRVI | $\frac{\propto * NIR - Red}{\propto * NIR + Red} + \frac{1-\propto}{1+\propto}$ |
| EVI | $2.5 * \frac{NIR - Red}{NIR + 6 * Red - 7.5 * Blue + 1}$ |
| VARI | $\frac{Green - Red}{Green + Red - Blue}$ |
| RDVI | $\frac{NIR - Red}{SQRT(NIR + Red)}$ |
| OSAVI | $\frac{NIR - Red}{NIR + Red + 0.16}$ |



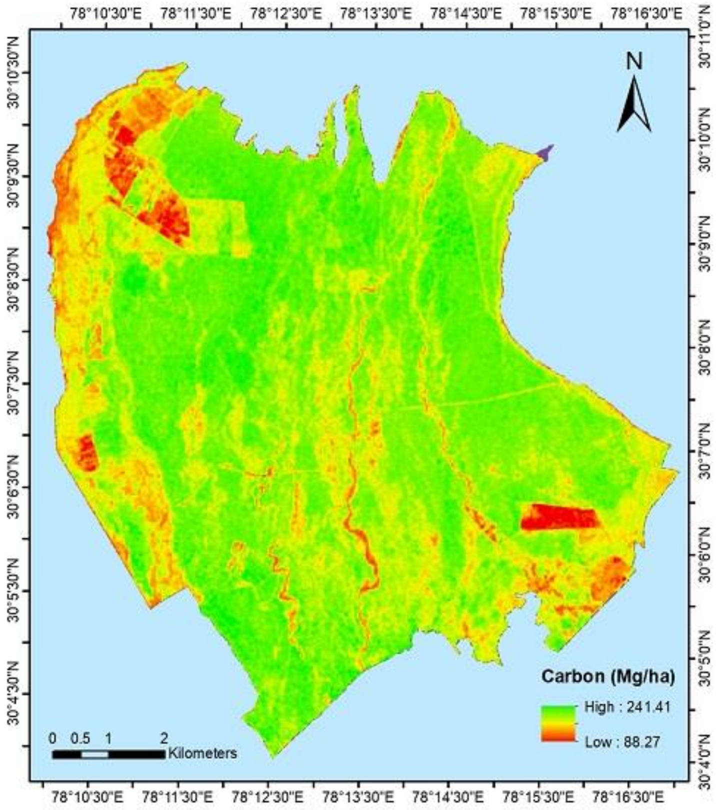

**NDWI**

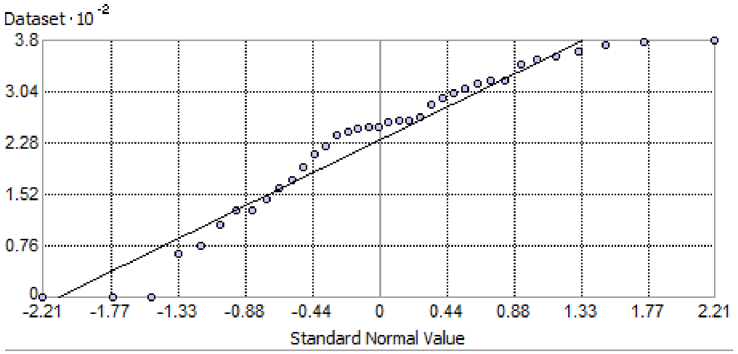
| Variable | R |
|---|---|
| SWIR-2 | -0.45 |
| NDWI2 | 0.44 |
| MIDIR | 0.43 |
| MSI | -0.43 |
| LSWI-1 | 0.39 |
| SWIR-1 | -0.38 |
| PC | -0.35 |
| NIR | 0.34 |

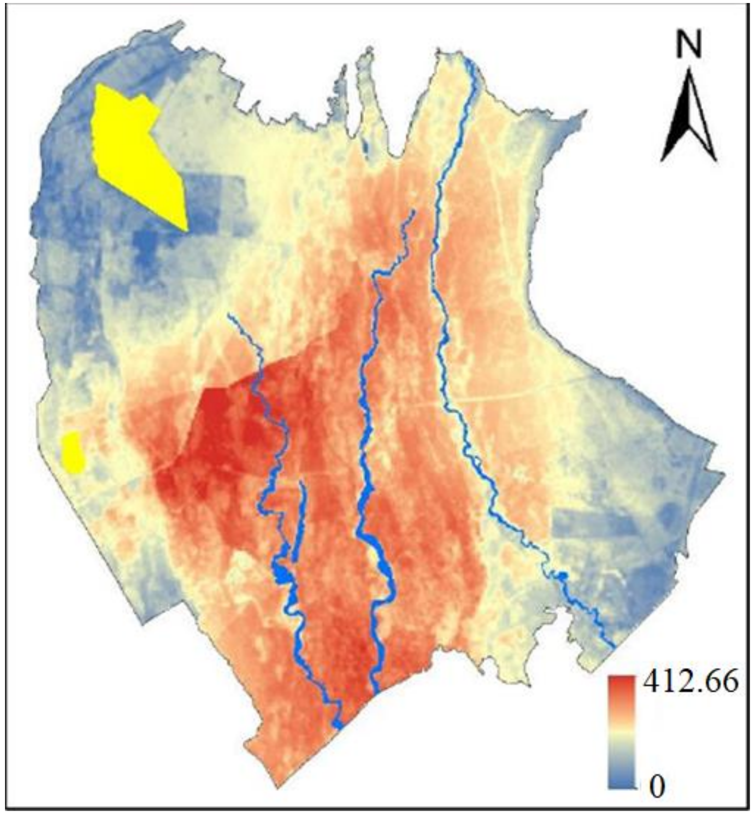
| Variable | R |
|---|---|
| OSAVI | 0.25 |
| RDVI | 0.24 |
| NDWI1 | 0.19 |
| EVI | 0.16 |
| WDRVI | 0.16 |
| GREEN | 0.13 |
| VARI | -0.09 |
| RED | 0.06 |

# Environmental Correlation

# Regression kriging

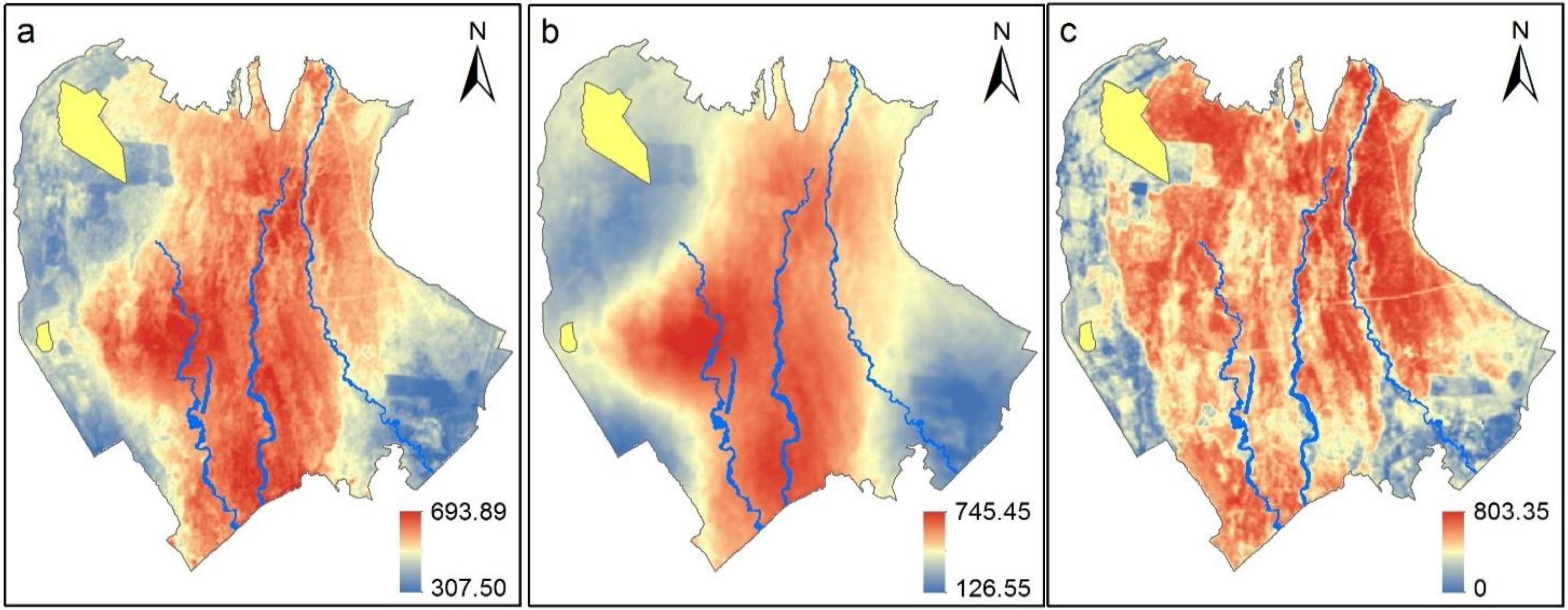



| variable | Linear fit function | RMSE (Mg $ha^{-1}$) |
|---|---|---|
| NDWI2 | 272.90 + 499.99 × LSWI-2 | 122.26 |
| SWIR-2 | 782.59 - 3008.24 ×SWIR-2 | 128.40 |
| MIDIR | -113.50 + 323.60 ×MIDIR | 128.46 |

# Co-kriging

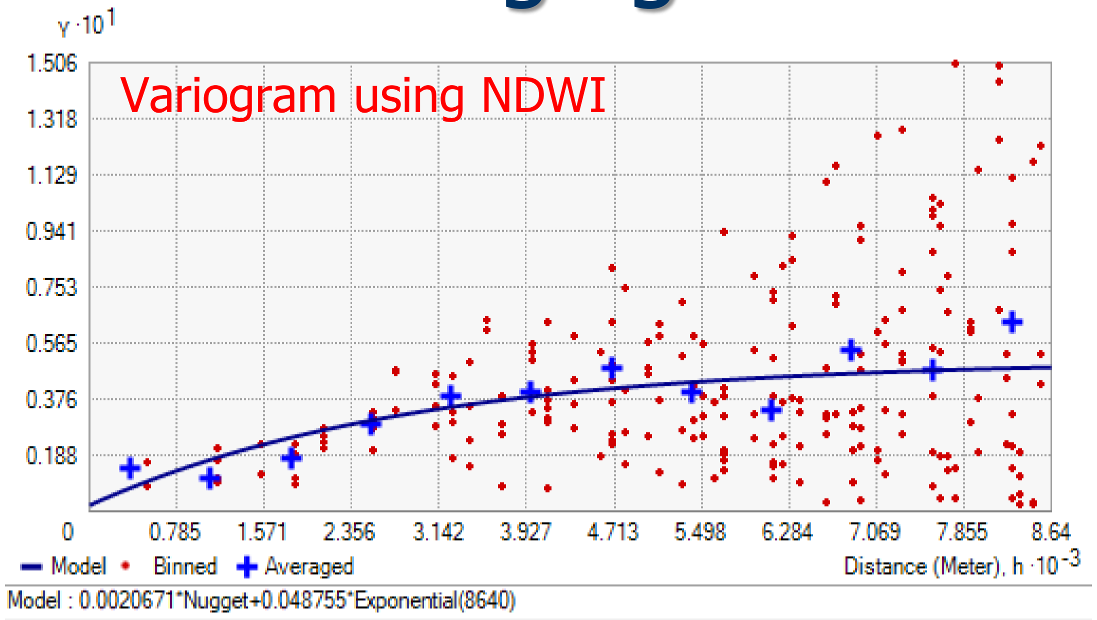



NIR

PCA

NDWI

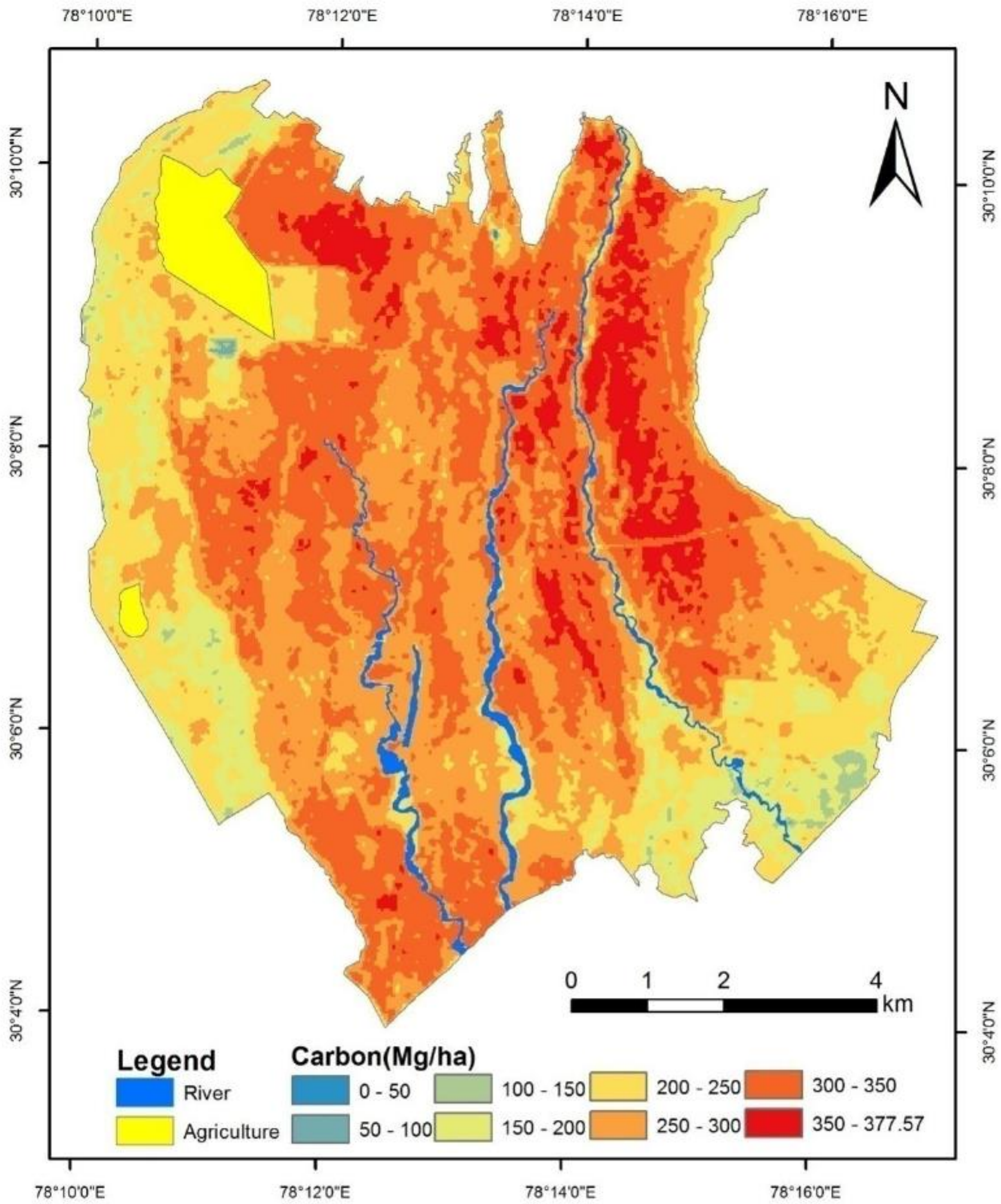
# Co-kriging



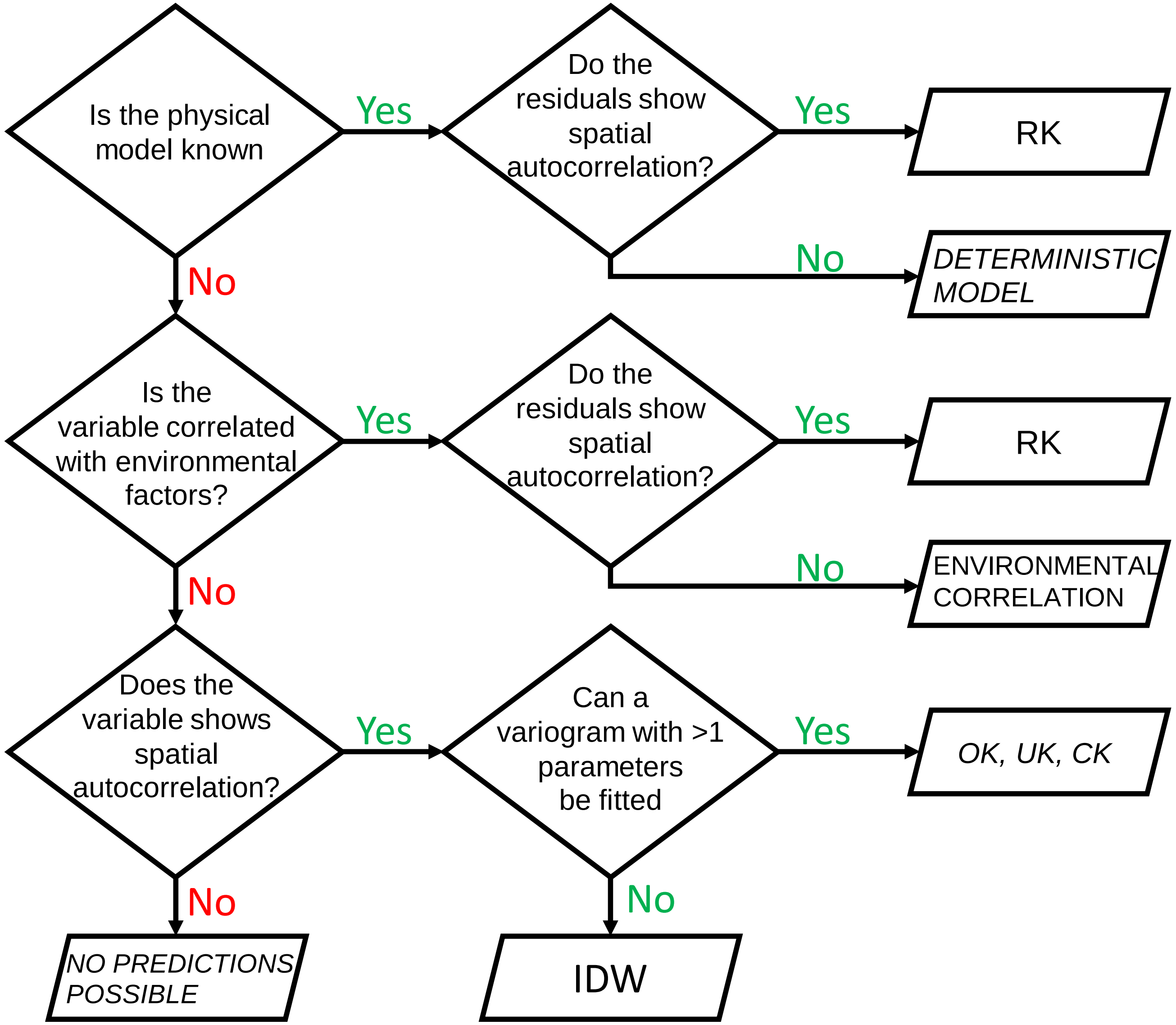

The total carbon stored by Barkot forest excluding agriculture and water body was estimated to be 2240797.37 Mg C, with an average of 276.13 Mg C $ha^{-1}$.

RMSE of 58.77 Mg $ha^{-1}$

| Location | Carbon (Mg C $ha^{-1}$) |
|---|---|
| Thano Forest range | 280.783 |
| Lachhiwala Forest Range | 236.962 |
| Rajpur forest range | 224.745 |



# Decision tree for selecting a suitable spatial prediction model




| Application | Reference |
|---|---|
| Air pollution mapping | Shukla et al 2020 |
| Offshore wind power prediction | Hu et al 2020 |
| Ground water potential | Rostami et al 2020 |
| Animal territory-thiessen polygons | Lotte et al 2013 |
| Spatial patterns of wildlife conflict | Siljander et al 2020 |
| Spatio-temporal prediction of LST | Bhattacharjee et al 2020 |
| Soil organic carbon stocks | Owusu et al 2020 |
| Distribution patterns of urban birds | Walker et al 2008 |
| Lake surface water quality | Chatterjee & lataye 2019 |
| Animals travel: path reconstruction | Fleming et al 2015 |
| Mean daily temperature | Sekulić et al 2020 |

SAGA, ArcGIS, IDRISI, ILWIS, GRASS, GSLIB,

R+gstat

# Thank You

Email- Taibang@iirs.gov.in
Tel- 0135-252-4176