GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

WEEK 6: GEOSTATISTICAL MODELLING

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

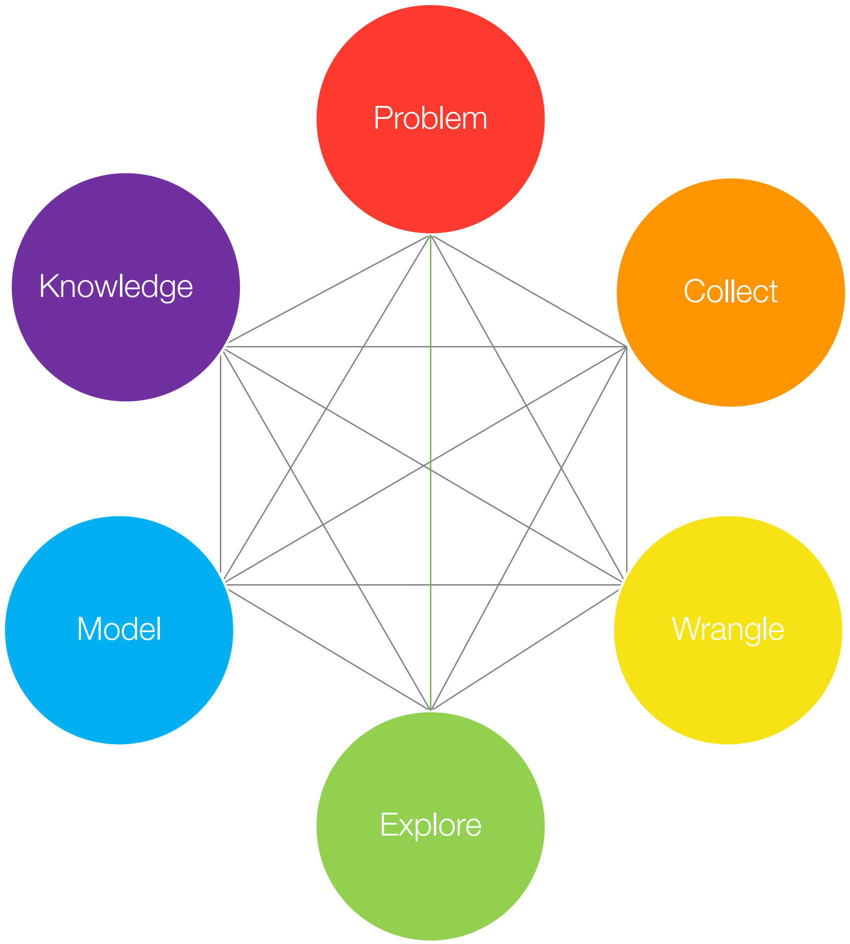
Contents

1. What is Geostatistics?

2. Inverse Distance Weighting (IDW)

3. Kriging modelling
   - Variogram analysis (empirical and theoretical variogram)
   - Graphing the semivariogram
   - Estimation of various parameters (sill, nugget and range)
   - Types of Theoretical Variogram to be Mindful off…

4. Workflow for performing a Kriging



QUICK RECAP

1. Niche Models are used on outcomes that are typically a point-process (i.e., events that occur at random e.g., wildfires, crime, road accidents etc.,).

2. Niche Models are used to predict were point-process are likely to occur, or where environments are ‘suitable’ for such outcomes (presence-only, presence-absence data etc.)

3. In week 3, we spoke of the various ways to represent geographic space through a device called a **Spatial Weight Matrix**
   - Contiguity-based (i.e., Queens, Rook & Bishop)
   - Distance-based (i.e., inverse, negative exponential)

Let’s rewind a bit to last week,
And to the Week 3 and 5

What is Geostatistics?

Definition of Geostatistics:

Geostatistics is class of spatial analytical methods used for **predictive inference.** They are a set of tools for predicting outcomes that's **continuous**, and the prediction is made at **unsampled locations** which are based on **sampled fixed points (or neighbouring point observations)**.

The key focus and main ingredients for geostatistics are:

1) Fixed point locations (i.e., coordinates [longitude & latitude]) with measured outcome (i.e., **dependent variable**) and other covariate data (i.e., **independent variable(s)**)

2) Raster grid template for making the geospatial predictions

The procedure for predicting phenomena at a location that have not been sampled based on nearby sampled points is called "Spatial Interpolation"

In terms of geostatistical methodologies, there are two main branches of models:

	Deterministic	Stochastic
Definition(s)	These type of models have parameter values that are typically arbitrarily defined	The parameter values for the these set of models have to be estimated
Model Types	**Inverse Distance Weighting (IDW)**	**Kriging**

- In terms of similarity between IDW and Kriging – it weights the surrounding measured values to derive a prediction for an unmeasured location.
- Differences are in the assumptions:
  - IDW assumes that spatial autocorrelation between neighboring points is proportional to the distance (and that it can be defined by distance reverse function).
  - Kriging assumes that distance (mainly) or directionality between sampling points reflects the spatial autocorrelation, and functions can be fitted to describe the correlation between points (and explain the variation on the surface)

Notes 1: Both rely on the similarity of nearby sample points to create or predict the surface, the deterministic relies purely on mathematical functions for interpolation, while statistical models are a combination of both statistics and mathematical methods to create the predicted surface and also produce levels of uncertainty about the predictions.

Notes 2: Kriging is a much better model, it is an approach that is more sophisticated than the IDW. Because of its deterministic nature, the prediction from IDWs tend to be less accurate (but with Kriging, a regression (either it be a null, or a simple or multivariable) model is calibrated into it.


Mining


Soil science

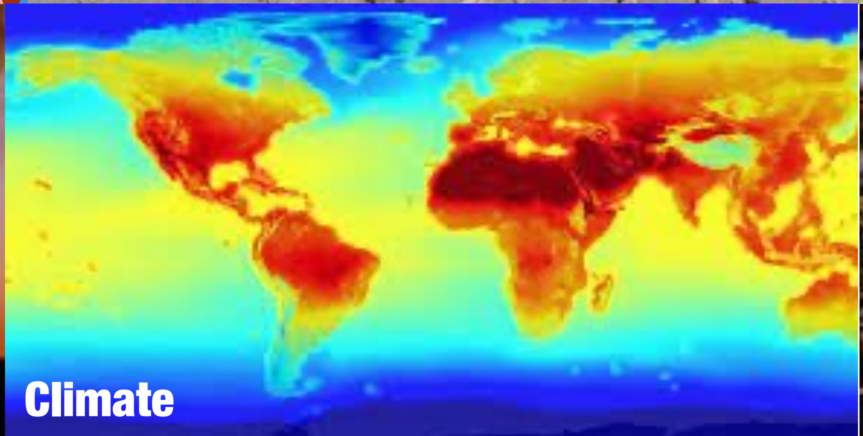

Environmental Criminology


Environmental & Spatial Epidemiology


Palaeontogy and Archology


Landscape ecology


Pollution Science


Climate


Humanitarian crisis

Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) [1]

Inverse Distance Weighting (IDW) is a deterministic approach which assumes that each input point has a local influence that diminishes with distance. It is a method of spatial interpolation that predicts a spatial location that is unsampled by using nearby sampled locations to apply **distance-based weights** and then averaging those sampled values.

The key characteristics of IDWs:

- It heavily relies on some distance decay function which applies weights of greater value to **sampled points** closer to the "**point of interest**" (we want to predict) than sampled points that further away.

- A specified number of points, or all points within a specified radius can be used to determine the output value of each location. Use of this method assumes the variable being mapped decreases in influence with distance from its sampled location.

Inverse Distance Weighting (IDW) [2]

Inverse Distance Weighting (IDW) is a deterministic approach which assumes that each input point has a local influence that diminishes with distance. It is a method of spatial interpolation that predicts a spatial location that is unsampled by using nearby sampled locations to apply **distance-based weights** and then averaging those sampled values.

Mathematical formulation for IDW method:

$$x^* = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n}{w_1 + w_2 + w_3 + \cdots + w_n} \equiv x^* = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \text{ where } w_i = \frac{1}{d(x^*, x_i)} \quad [1]$$

Expanded form **Condensed form**

x^* is the unknown spatial location we want to predict. The x_i is the known sampled locations which we will use to predict x^*. Lastly, w_i is a weight derived from computing inverse distance between unknown location x^* we want to predict and known sample location x_i.This distance is donated as $d(x^*,\ x_i)$

$\equiv$ This symbol mean they are identical

Distance and Spatial Weights [1]

Remember the estimation of a spatial weight $\boldsymbol{w_{ij}}$ which is based on distance $\boldsymbol{d_{ij}}$ between some location i and j. Here, we use point locations, or the centroids of such given area, to compute the distances whereby the coordinates for i is (x_i, y_i) and j is (x_j, y_j) are used. **The goal here is to integrate distance decay in IDW models.**

Euclidean Distance:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad [1]$$

Inverse Distance:

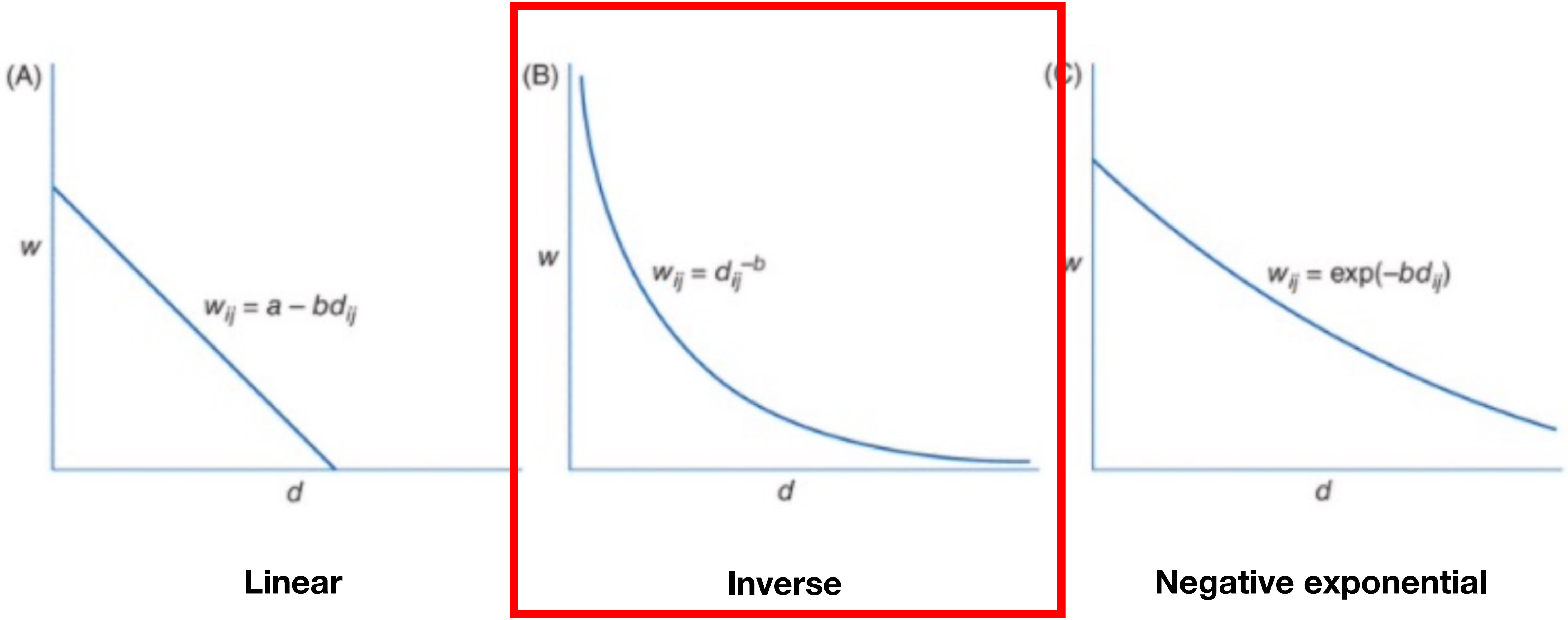
$$w_{ij} = \frac{1}{d_{ij}^{\beta}} \quad [2]$$

Negative exponential:

$$w_{ij} = \exp\left(-\frac{d_{ij}}{\beta}\right) \quad [3]$$

Note that β = 1 or β = 2 (we usually use 2, but it's really up to you which ever value you pick!)

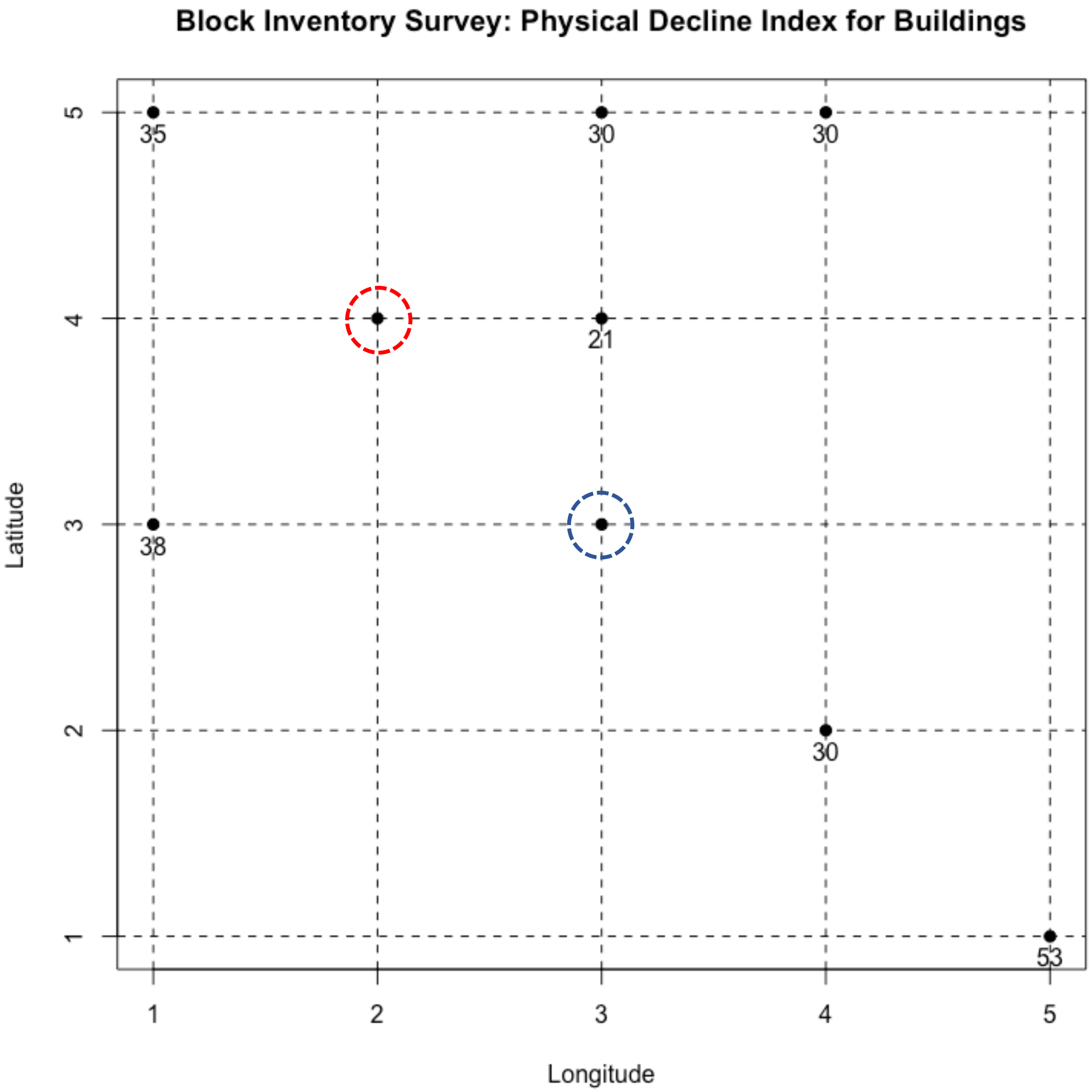
Distance and Spatial Weights [2]



Linear **Inverse** **Negative exponential**

For the IDWs we usually select this method to in computing our weights, which, in turn, are used in IDW model for spatial prediction

Example: Physical Decline Index of Buildings [1]

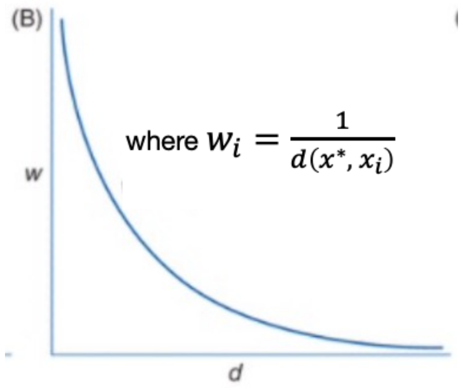


Let us demonstrate this using the site with the red dashed circle. Let us represent this as x^*. The points with the data are represented as x_i.

Here, we have a small study area with 7 sampled locations. A Block Inventory Survey (BIS) was carried out to measure the quality of buildings.

How do we use the IDW framework for predicting the two unsampled sites?

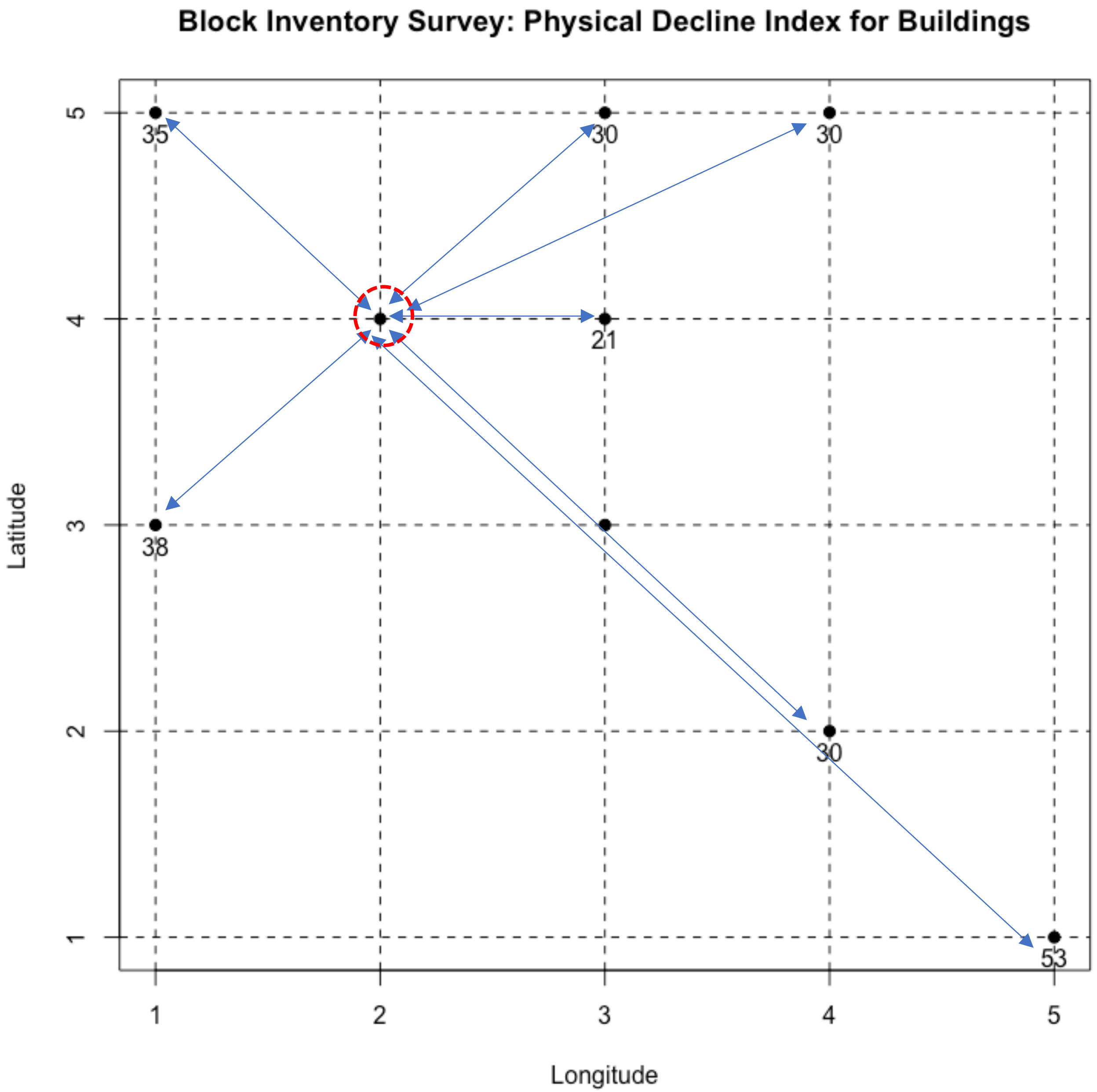
$$d = \sqrt{(x_0^* - x)^2 + (y_0^* - y)^2} \qquad [1]$$



[2]

$$x^* = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \qquad [3]$$

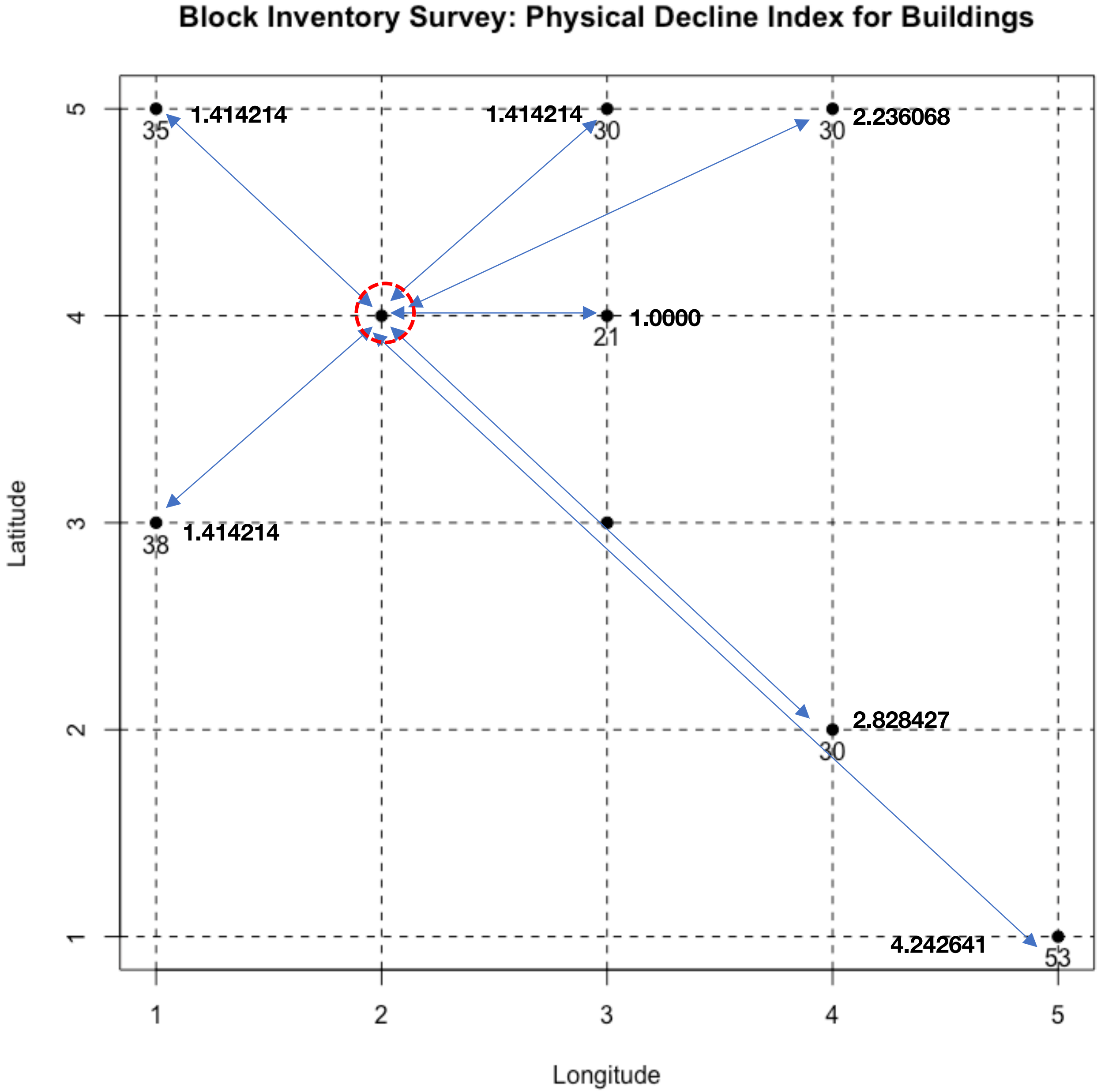
Example: Physical Decline Index of Buildings [2]



Step 1: We compute all the distance between each pair of points i.e., that is all sampled sites and unsampled site. **We use the Euclidean distance formula for the calculation.**

$$d = \sqrt{(x_0^* - x)^2 + (y_0^* - y)^2}$$

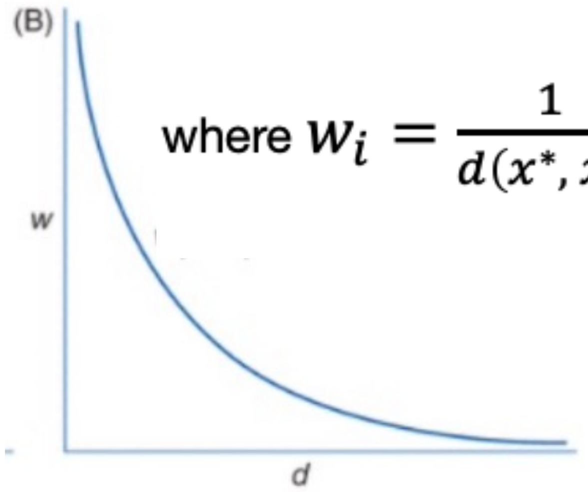
Example: Physical Decline Index of Buildings [3]



Step 1: We compute all the distance between each pair of points i.e., that is all sampled sites and unsampled site. **We use the Euclidean distance formula for the calculation.**

$$d = \sqrt{(x_0^* - x)^2 + (y_0^* - y)^2}$$ ✓

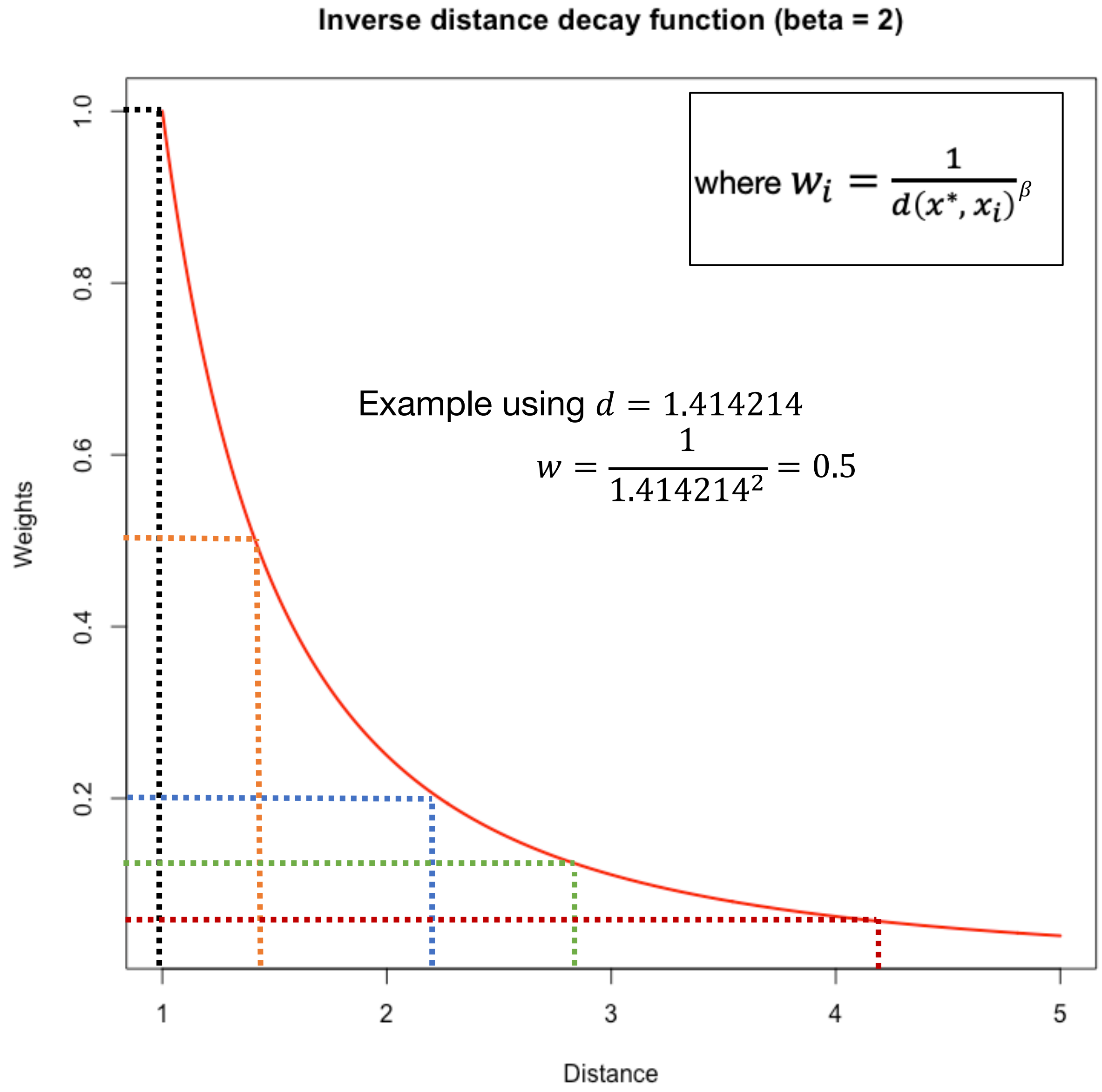
Step 2: We compute the weights using the **inverse distance formula**

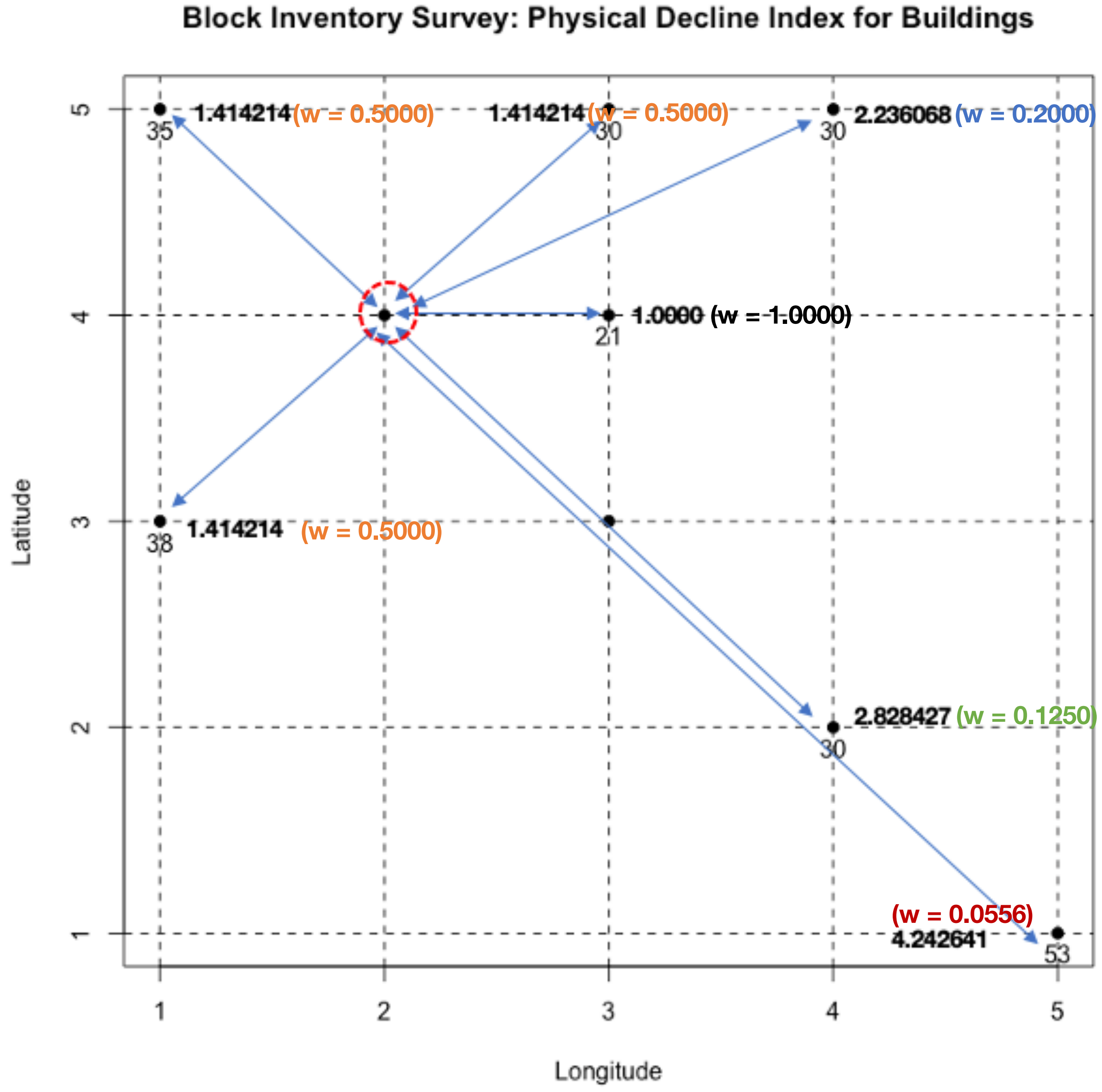


where $w_i = \dfrac{1}{d(x^*, x_i)}$

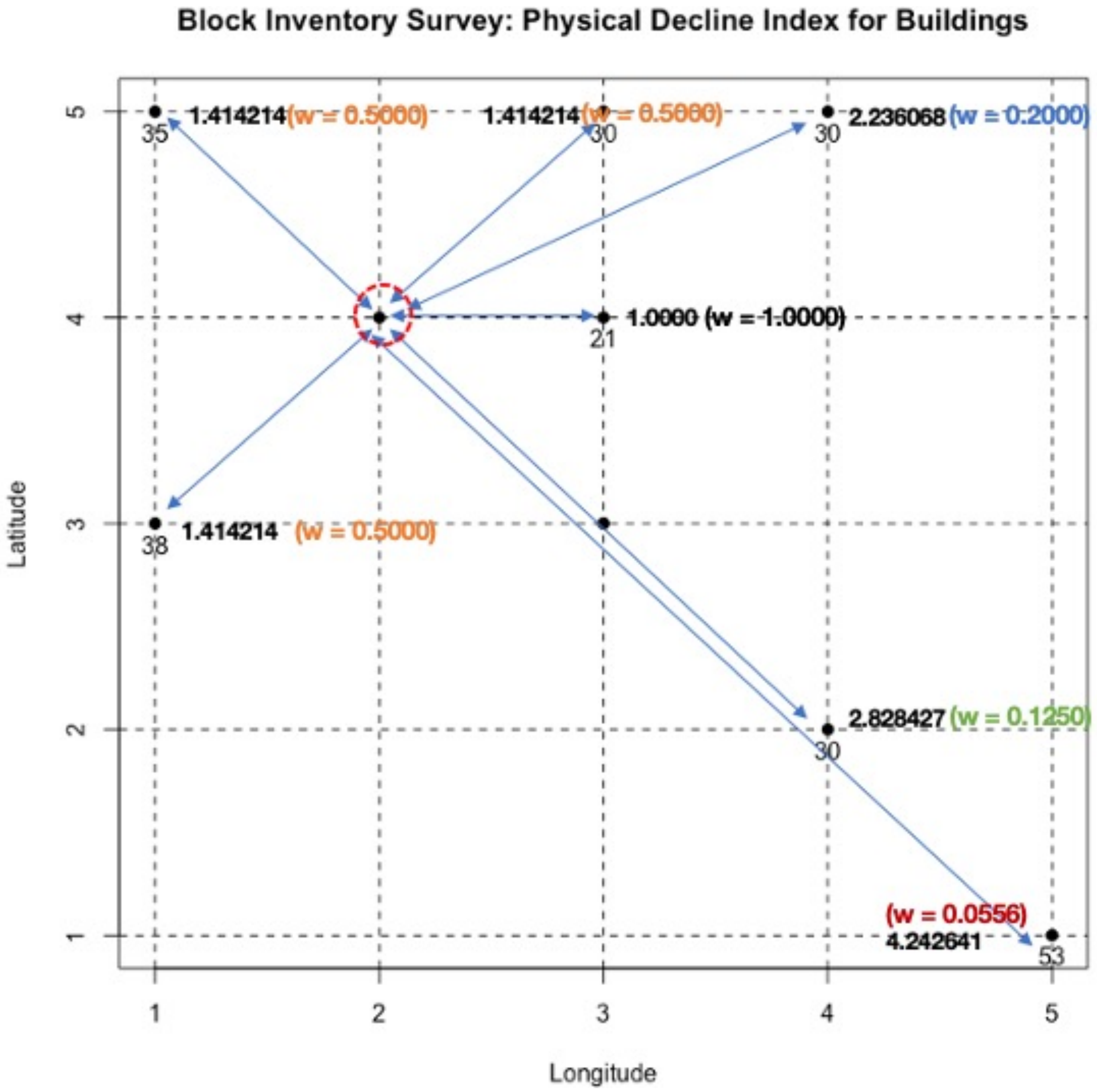
We can create this graph by using the **inverse distance** formula for all values of d between 1 to 5

How to compute the distance-based weights





Example: Physical Decline Index of Buildings [4]

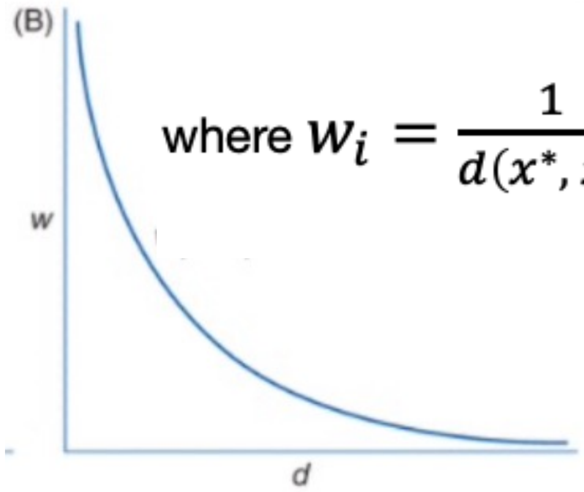


Step 1: We compute all the distance between each pair of points i.e., that is all sampled sites and unsampled site. **We use the Euclidean distance formula for the calculation.**

$$d = \sqrt{(x_0^* - x)^2 + (y_0^* - y)^2}$$

✓

Step 2: We compute the weights using the **inverse distance formula**

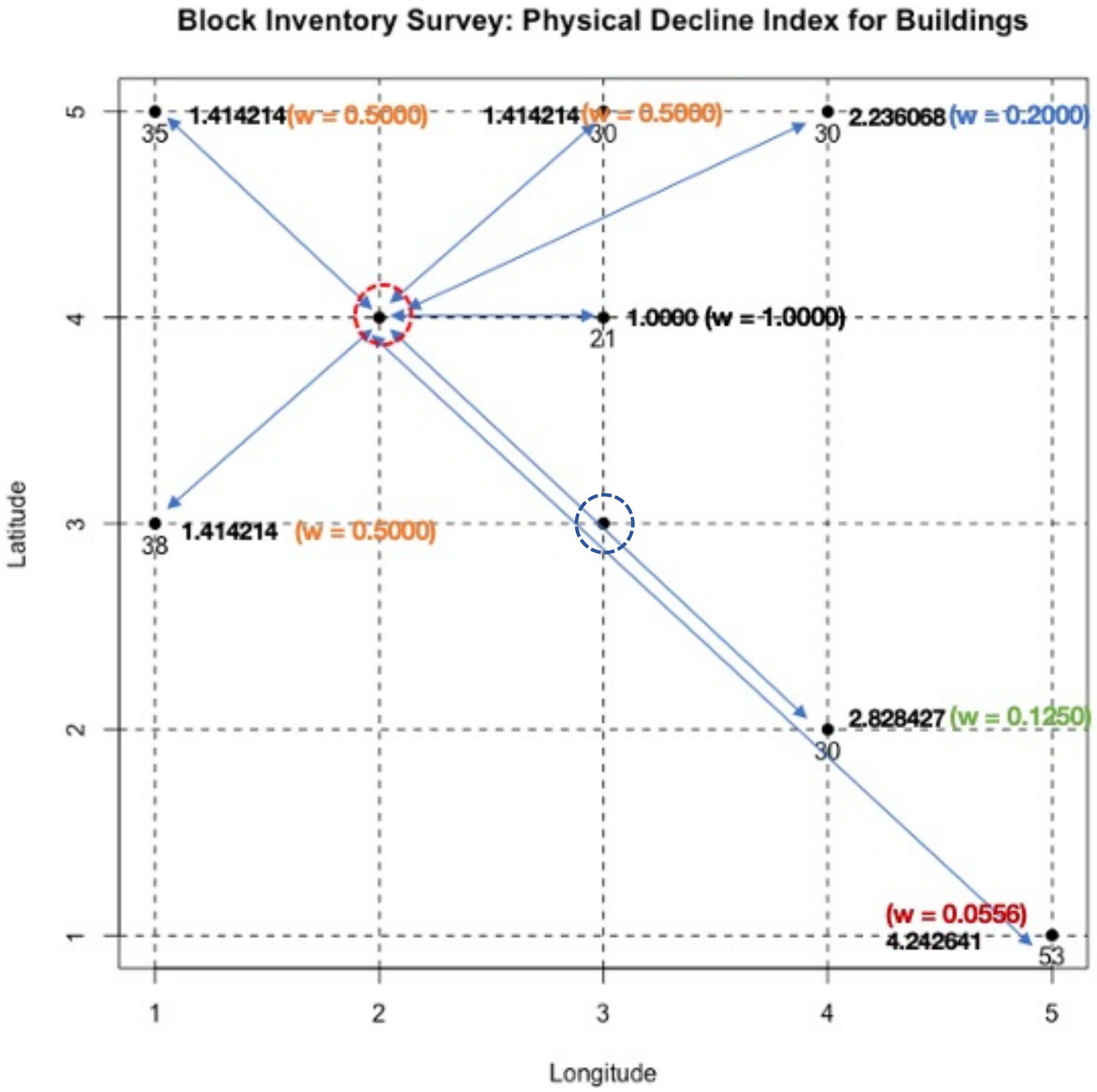


where $w_i = \dfrac{1}{d(x^*, x_i)}$

✓

We can create this graph by using the **inverse distance** formula for all values of d between 1 to 5

Example: Physical Decline Index of Buildings [5]



Step 3: Estimate x^* using that mathematical formula

$$x^* = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \equiv \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n}{w_1 + w_2 + w_3 + \cdots + w_n}$$

We multiple calculated weights with the observed values for Physical Decline Index, which in turn are summed. Next, we divide the numerator by the summed weights in the denominator to derive our interpolated point.

$$\Rightarrow \frac{1(21) + 0.5(38) + 0.5(35) + 0.5(30) + 0.2(30) + 0.125(30) + 0.0556(53)}{1 + 0.5 + 0.5 + 0.5 + 0.2 + 0.125 + 0.0556}$$

$$x^* = 29.5757$$

The estimate PDI is **29.5757**, this calculation is repeated for the other unsampled point etc.

Important Note: You can also specify some threshold distance to make the prediction on some K-number of nearest neighbours. Instead of using the full available sample of points

Kriging Modelling

What are Kriging Models [1]

Kriging is a statistical (or stochastic) approach which has a family of models that can use limited number of spatially referenced sampled data points of an continuous outcome, to make such predictions at an unsampled location.

The key characteristics of Kriging models:

1. Unlike the IDW which is purely based on distance decay between points. The kriging is more interested in both the separation distance between points as well as how these two points are spatial correlated with each other

2. Another important thing with Kriging, we are able to determine at which the **separation distance** are points no longer spatially autocorrelated with each other using a device called a **semivariogram**

3. It is through the semivariogram we can derive some important estimates called the **Semivariance**, **Nugget**, **Sill (or Partial Sill)** and **Range** to calibrate our Kriging model for spatial prediction.

Common types of Kriging models are: **Ordinary Kriging**, **Regression-based Kriging**, Universal Kriging and many more

What are Kriging Models [2]

Kriging is a statistical (or stochastic) approach which has a family of models that can use limited number of spatially referenced sampled data points of an continuous outcome, to make such predictions at an unsampled location.

Mathematical formulation for the Kriging method:

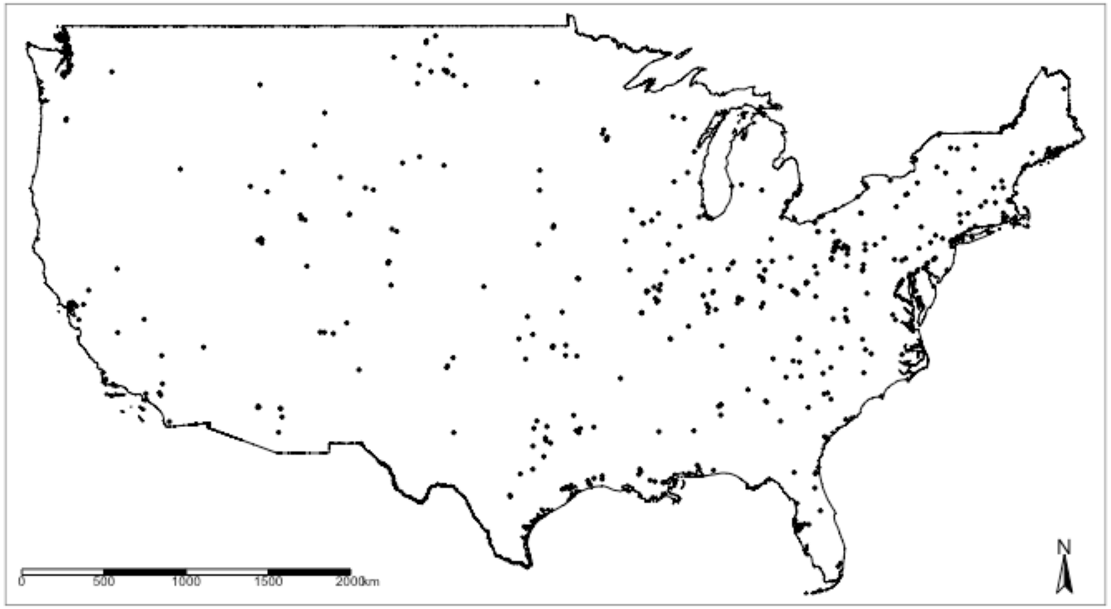
$$x^* = w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n + \varepsilon \equiv \sum_{i=1}^{n} w_i x_i + \varepsilon$$

However, the computation of these weights are an involved process

Before we can estimate our x^*

1. Calculate experimental variogram
2. Fit theoretical variogram model
3. Calculate weights (using the Lagrange multiplier method [numerical approximation technique] to the above equation to get the weights)
4. Prediction using Kriging equation above (you can obtain the predicted value and uncertainty [i.e., error])

Example: Air pollution of Sulphur Dioxide (SO_2) in USA



The points are air pollution monitoring stations which contain some measures for various air contaminants. We can use the Kriging models to predict the contamination levels where there are no stations (i.e., white spaces).

Here, we have annual measures of SO_2 estimated from 458 monitoring stations across USA. We want estimated the it concentrations were such stations are not present.

How do we use the Kriging framework for predicting the at unsampled sites?

Estimation of the **empirical variogram** to obtain **sill, Partial, sill, range** and **nugget** **[1]**

From **[1]**, specify our **theoretical variogram** in order to construct a covariance matrix **[2]**

From **[2]**, apply **matrix algebra** to estimate the weights **[3]**

From **[3]**, use result for weights and insert into our **Kriging model** to make our predictions **[4]**

$$x^* = w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n + \varepsilon$$

What is a Variogram?

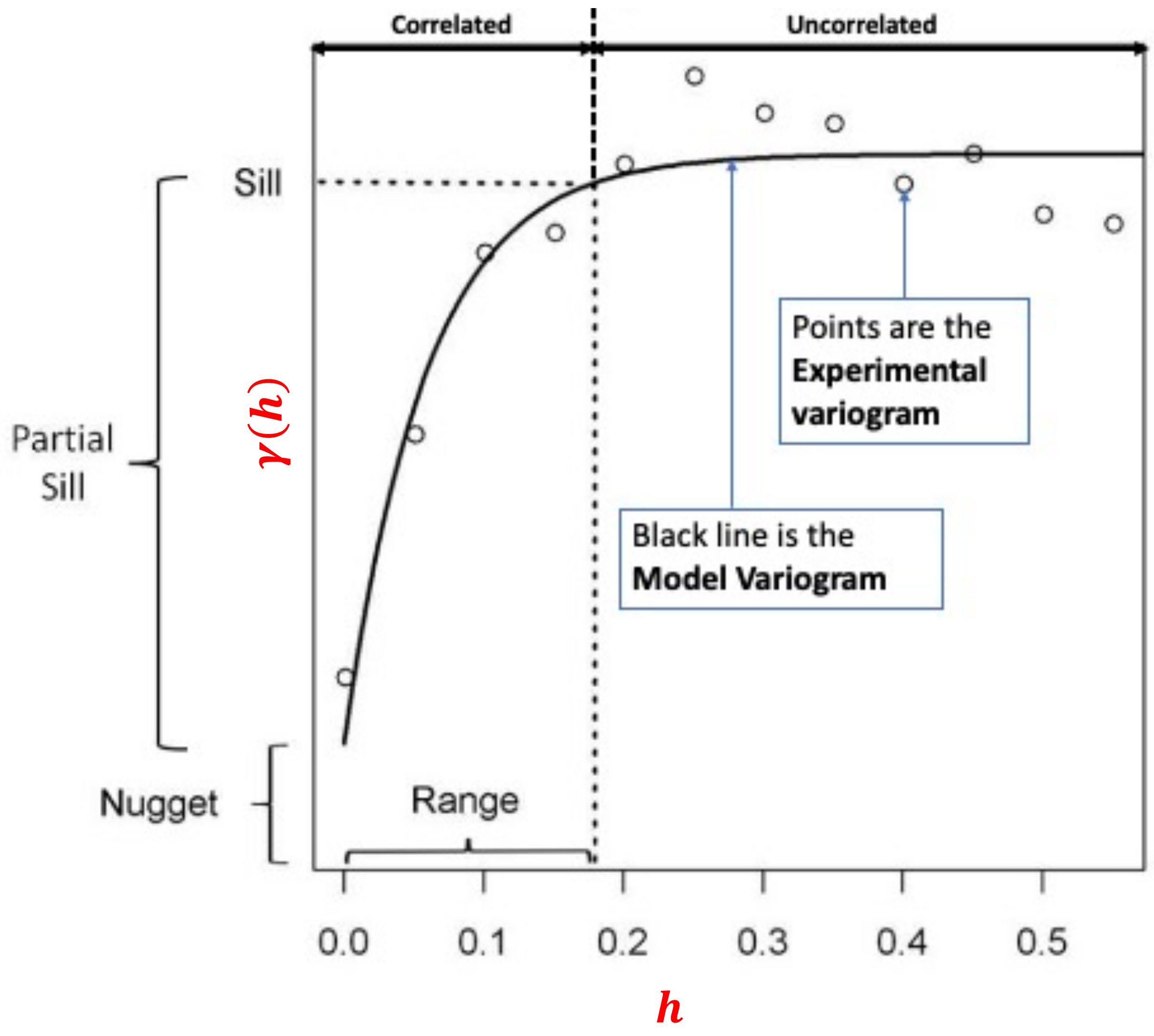
A **variogram** (or **semivariogram**) is a graphical output which allows the user to measure & plot the degree of how values differ according to how far apart they in space.

- In any kriging analysis, quantifying the **semivariogram** is the most important step.

- In order, to quantify those weights to make the prediction, there are four important quantities with must derive the following:

  1. **Semivariance**
  2. **Sill (and Partial Sill)**
  3. **Nugget**
  4. **Range**

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{(m,n)=1}^{N(h)} (x_m(h) - x_n(h))^2$$



Experimental (or Empirical) Semivariogram

Definitions:

1. **Sill:** This is the maximum Semivariance value observed, and it shows the threshold (or flatline) for which points are no longer spatially autocorrelated.

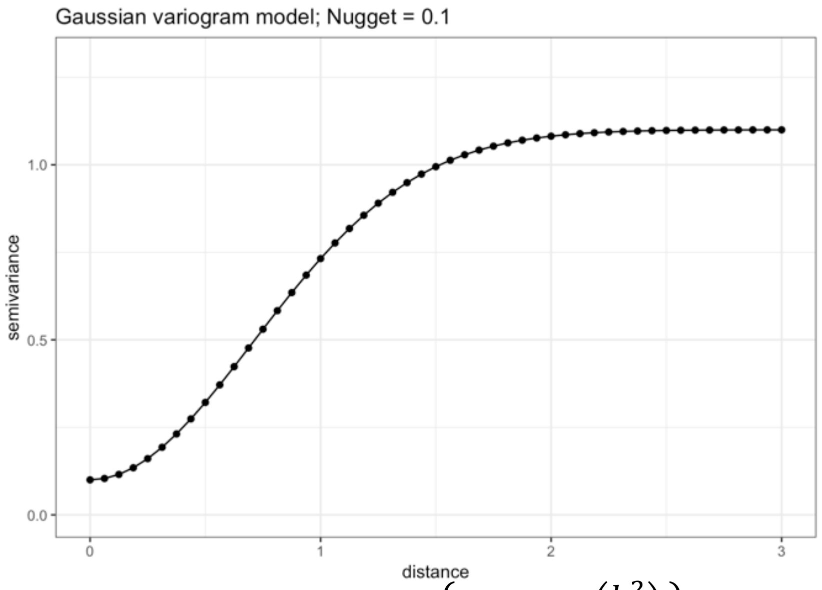
2. **Range**: Maximum separation distance for h, at which we expect our paired points to no longer be spatially autocorrelated with each other.

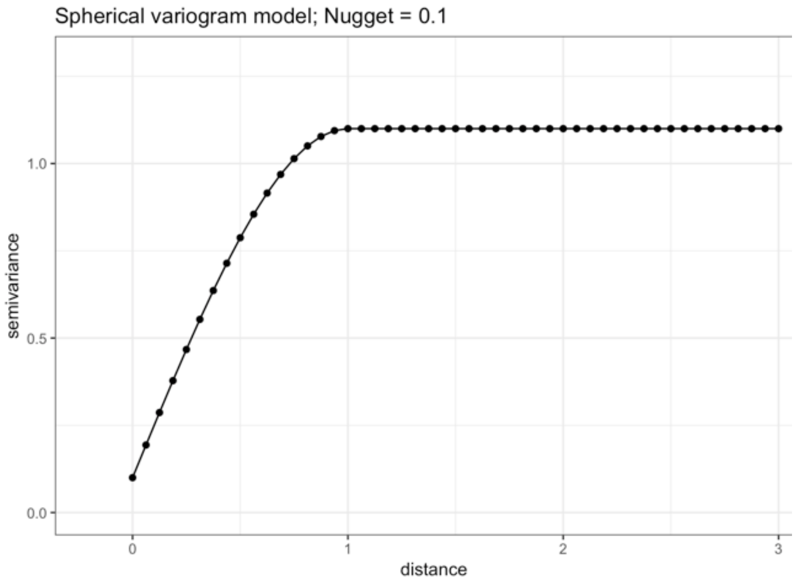
3. **Nugget**: Is a measurement error. The larger the **nugget** relative to the **sill**, the less spatial dependence there is in the data and less useful Kriging will be.

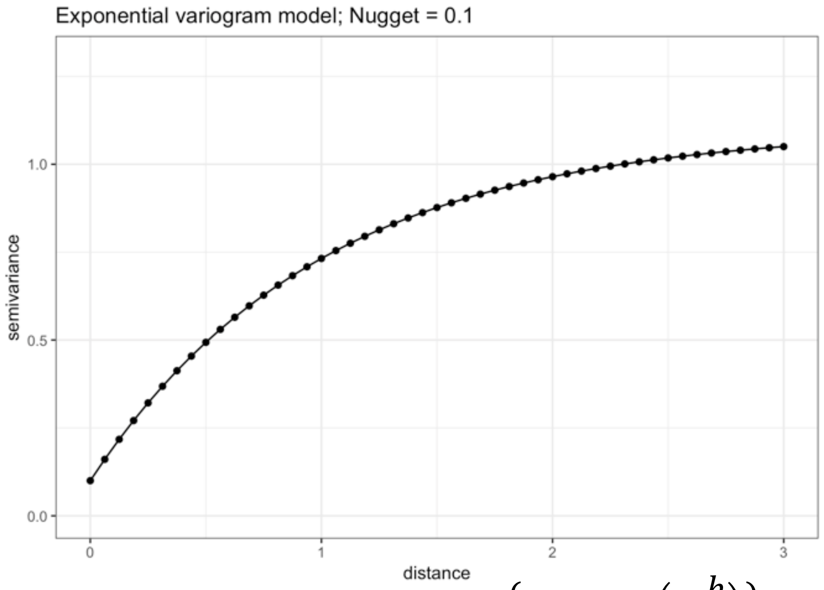
4. **Partial Sill**: Is the difference between **Sill** and **Nugget**.

We will need to extract these values to determine the best "Theoretical Semivariogram"

Common types of Theoretical Variograms







Gaussian: $C(h) = c\left\{1 - exp\left(\frac{h^2}{r^2}\right)\right\}$

Spherical: $C(h) = c\left\{1.5\,\frac{h}{r} - 0.5\left(\frac{h}{r}\right)^3\right\}$ if $h \leq r$

$\gamma(h) = c$ for $h > r$

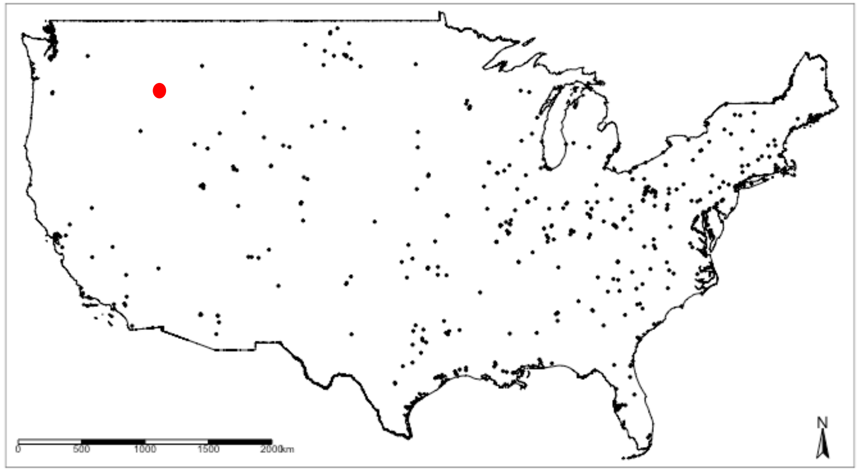
Exponential: $C(h) = c\left\{1 - exp\left(-\frac{h}{r}\right)\right\}$

$\gamma(h) =$ c if $h = 0$

Note that $c =$ partial sill, $h =$ separation distance and $r =$ range. $C(h)$ is the covariance

Fitted Semivariance: $\ddot{\gamma}(h) = c\ - C(h)$

Step 1: Variogram analysis [1]



The **458 points** are air pollution monitoring stations which contain some measures for various air contaminants. We can use the Kriging model to predict the contamination levels where there are no stations (i.e., white spaces).

Red point is a location without a pollution monitor. Here, is an example of location where we can use the Kriging.

- We need to list every possible pair of points. **So from the 458 sample locations, there are 104,653 pairs**. A distance between each is calculated – this know as a **separation distance (*h*)**
- At the red point, a **separation distance** is computed elsewhere. Note that this place has no value but we want to make a prediction at the point.

We need to use this formula to compute Semivariance from each pair of points

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{(m \neq n)=1}^{N(h)} (x_m(h) - x_n(h))^2 \quad [1]$$

- $\hat{\boldsymbol{\gamma}}(\boldsymbol{h})$ is the Semivariance, which measures how 2 points are spatially autocorrelated with respect to their differences in distance (i.e., separation distance ***h***)
- ***h*** is the separation distance (**see equation [1] in slide 11**)
- ***N(h)*** is the total number of paired points with the same separation distance value of ***h*** (if there are not set of points with the same ***h***, then we do not need to use the summation part of the equation, nor the ***1/2N(h)*** in the formula will not be needed as well).
- This means that we only concern ourselves with using this equation:

$$\hat{\gamma}(h) = (x_m(h) - x_n(h))^2 \quad [2]$$

- $\boldsymbol{x_m(h)}$ and $\boldsymbol{x_n(h)}$ are two pairs of observations which has a separation distance of ***h***

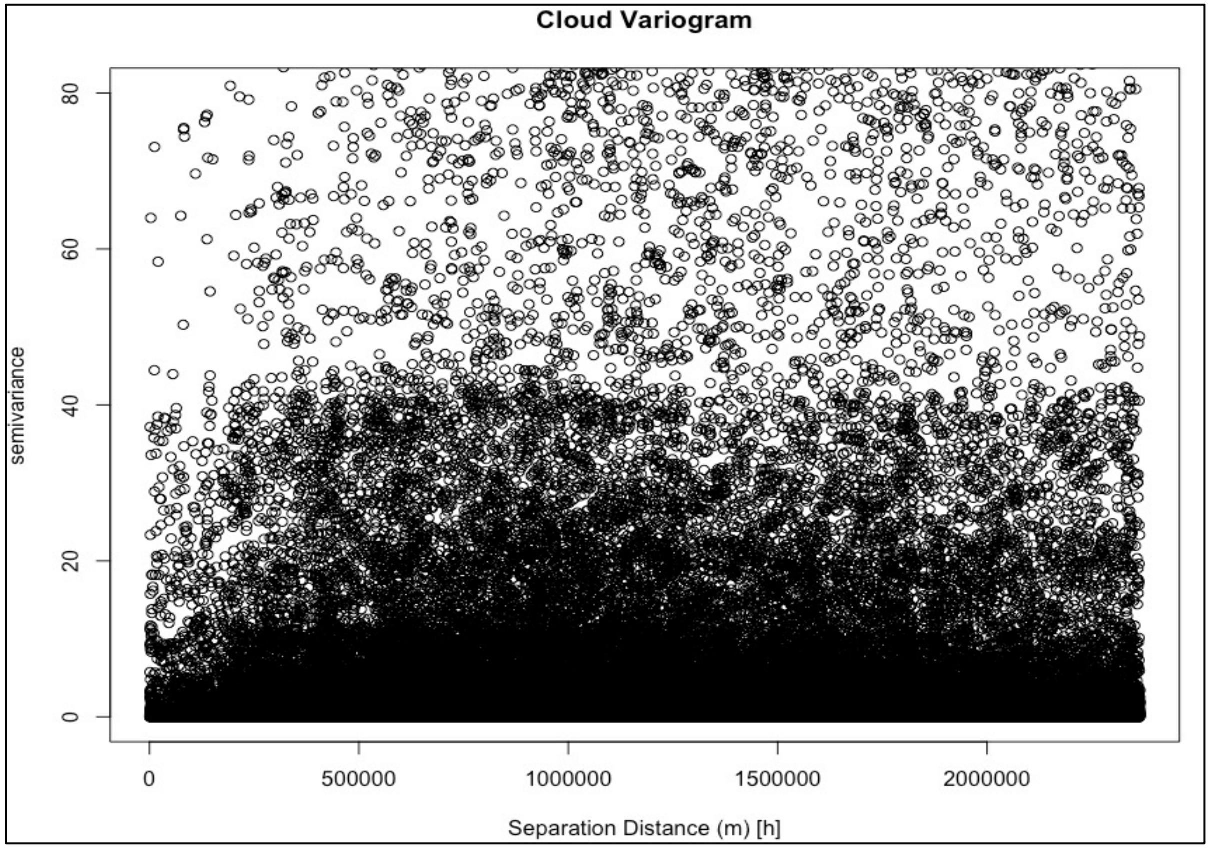
Example:

Suppose the separation distance for two stations was *h* = 125000m and SO_2 levels were 67 ppb at $x_m(h)$ and 61 ppb at $x_n(h)$.

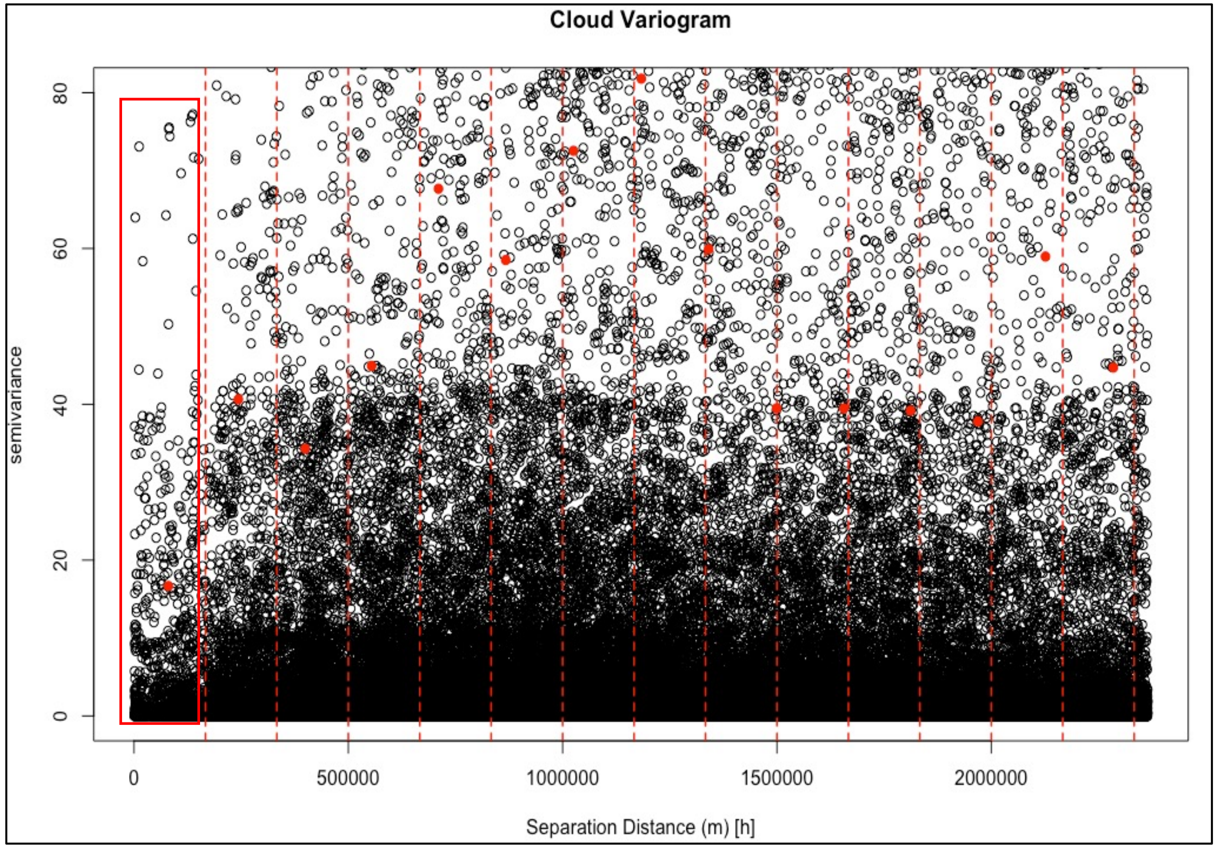
The $\gamma(h) = (67 - 61)^2$ = 36ppb (plot 36ppb against *h* (125000m))

Repeat for all remaining 104,653 points

Step 1: Variogram analysis [2]

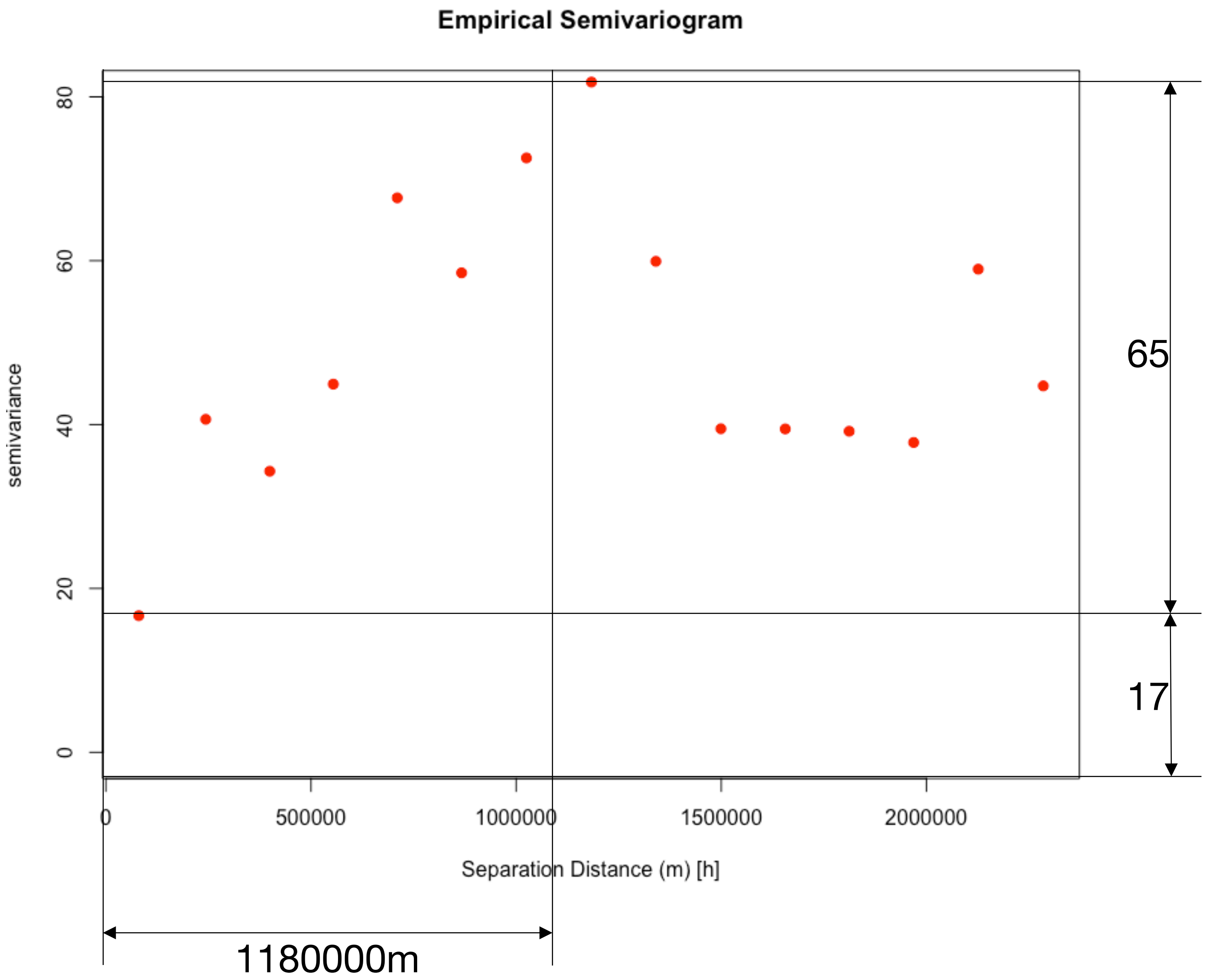


This plot is known as a **Cloud Variogram**, which contains a cloud of points.



We create bins (red dashed lines) and compute an average using all points within a bin. For example, all these points within the first bin (0m to 166666.7m) (i.e., rectangle block) are averaged to produce the red dot.

Step 1: Variogram analysis [3]



From the output, we should note the approximate values for the **partial sill, nugget and range**.

Here, we are eyeball here!

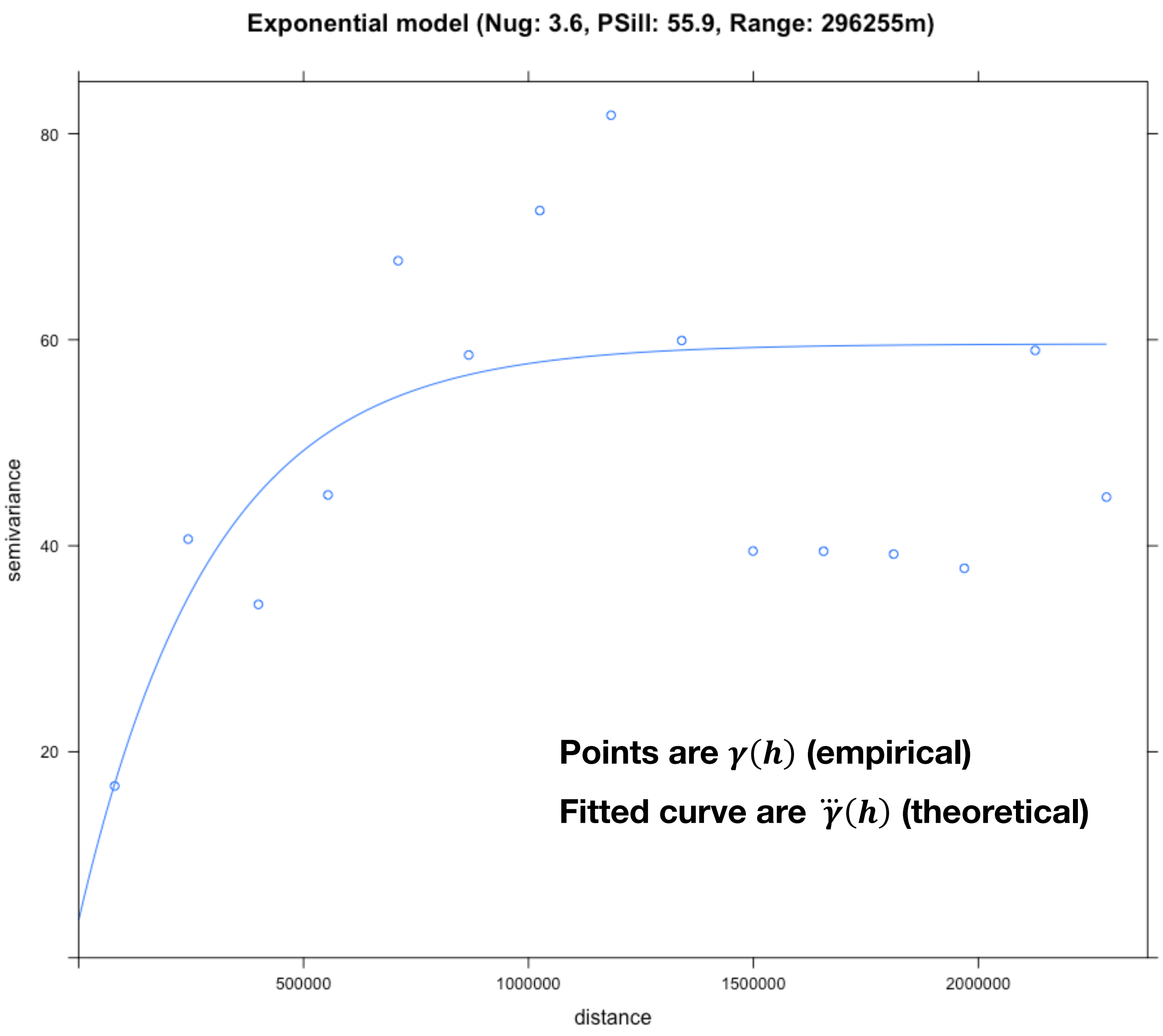
- The nugget is roughly 17
- The range is roughly 1180000 meters
- The partial sill is 65. This is derived from the peak value for gamma subtracted by the nugget (82 - 17 = 65).

These serve as initial values to give us an idea of what to expect when we proceed to fit a theoretical semivariogram

These help us generate some function, which in turn, helps us to determine whether this pattern is either **Exponential, Spherical or Gaussian.**

We need to select function that's appropriate for our Kriging model.

Step 2: Fitting the theoretical variogram



Difficult step is the determination for the appropriate theoretical variogram. It is basically up to the user in terms of which function is used…

An informed approach is allowing the software to make the selection for you:

- Provide a set of initial values and modelled result from **Empirical Semivariogram**. What happens is that the software will converge to the optimal **nugget, Partial sill** and **range** value
- Based on the optimal values from point [1] and shape of the empirical variogram, the best model is selection (here – we specify all 3 models as part of the model selection process).

Exponential:

$$C(h) = c\left\{1 - exp\left(-\frac{h}{r}\right)\right\}$$

where $\gamma(h) = \text{c if } h = 0$

Parameters:

$C(h)$ is estimated covariance (error)
c = 55.9 (partial sill)
h = Separation distance (use all values of h)
r = 296255m (range)

Fitted Semivariance: $\ddot{\gamma}(h) = c - C(h)$

Step 3: Estimation of spatial weights

Our statistical model is: $x^* = w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n + \ \varepsilon$

(m+1) by (n+1) K matrix
(m = n making it a square matrix)

$$K = \begin{bmatrix} \gamma(h)_{1,1} & \cdots & \gamma(h)_{1,n} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(h)_{m,1} & \cdots & \gamma(h)_{m,n} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \qquad k = \begin{bmatrix} \dddot{\gamma}(h)_1 \\ \vdots \\ \dddot{\gamma}(h)_n \\ 1 \end{bmatrix} \qquad w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \varepsilon \end{bmatrix}$$

K, Semivariance computed from our empirical variogram

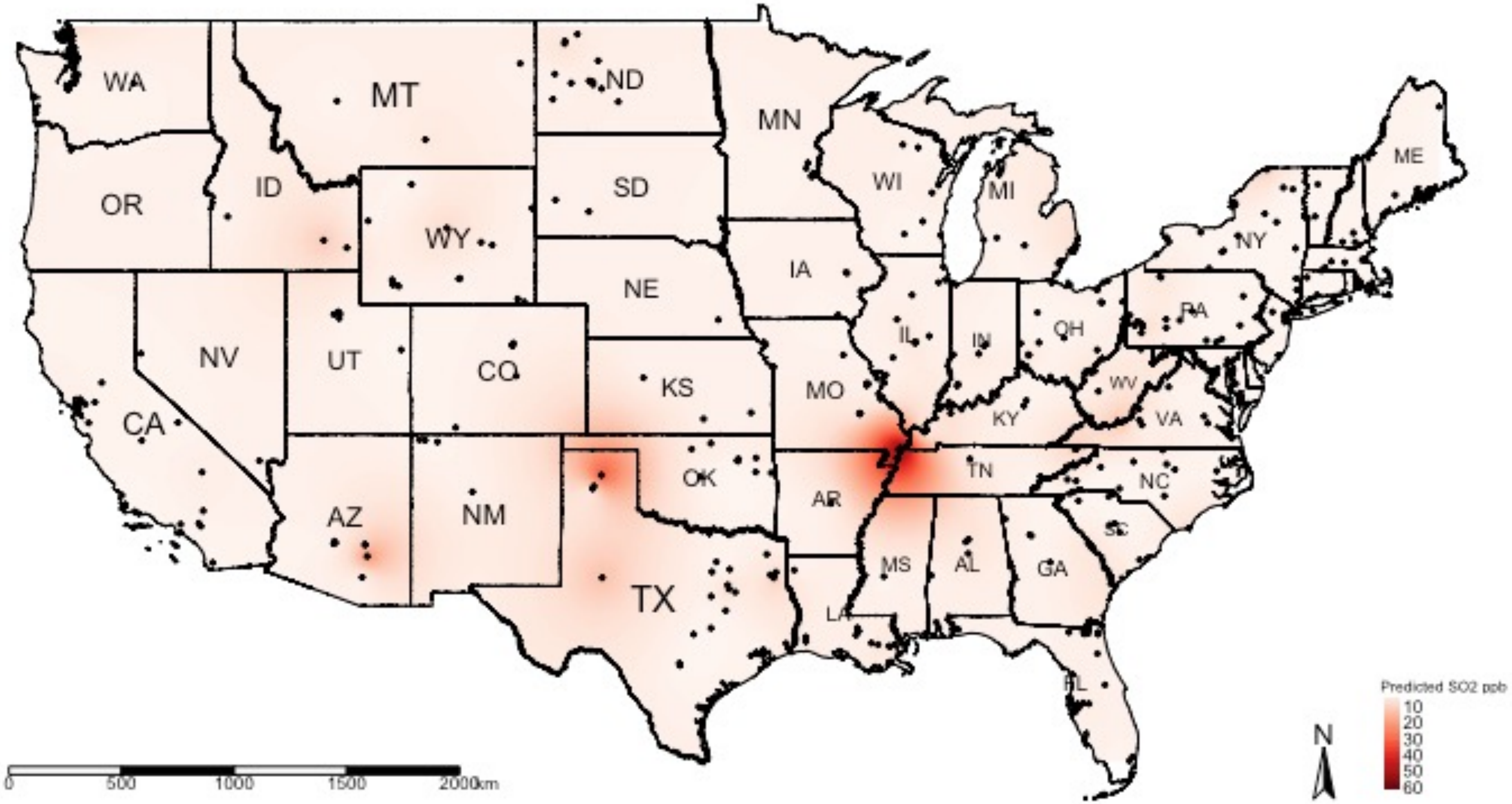
k, These are the corresponding Fitted semivariance computed from our theoretical variogram. The correspond to each other based on the separation distance *h*

w, these are the weight coefficients, along with the error we need to estimate.

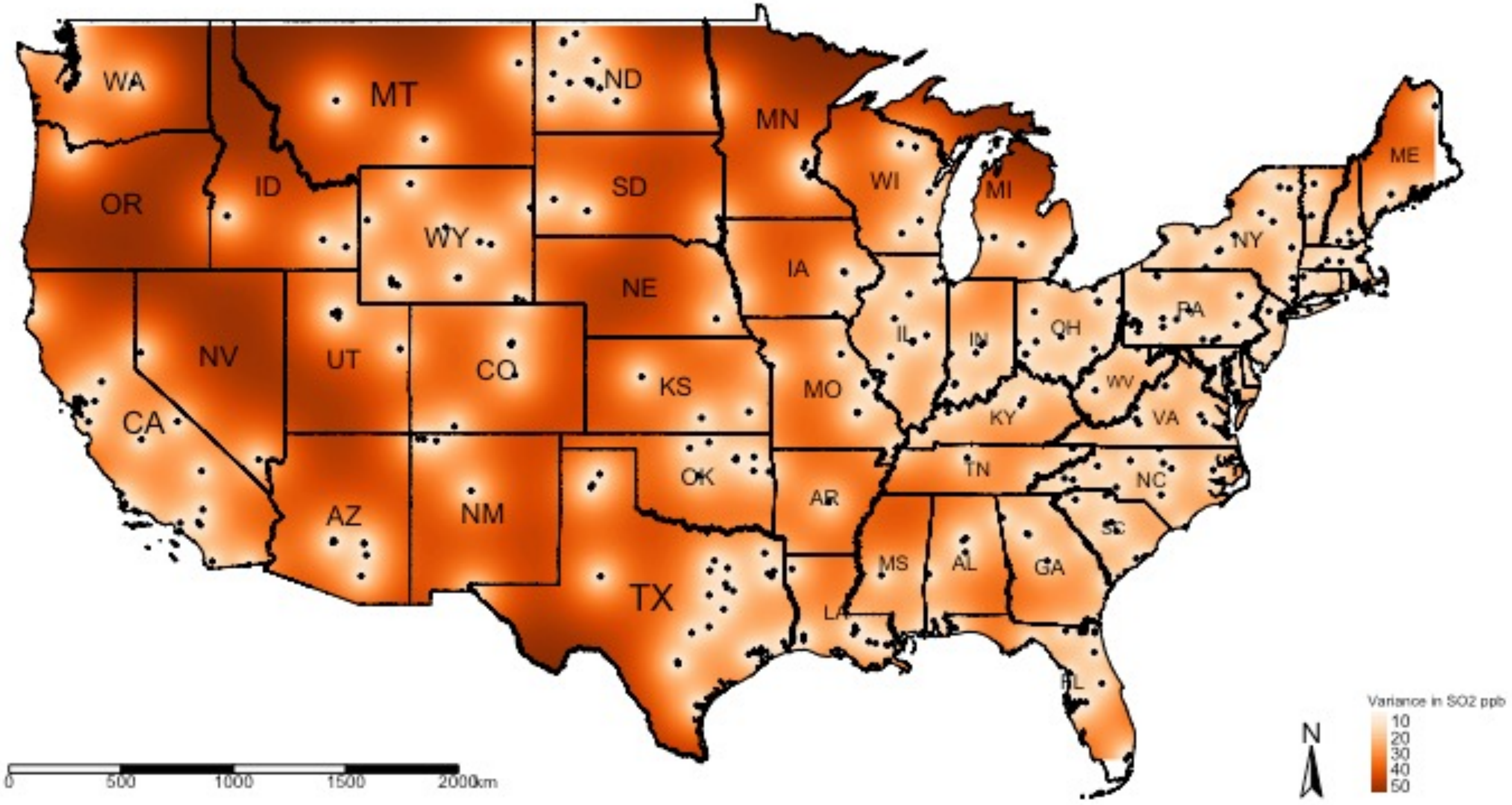
To estimate *w*:

$$\boldsymbol{w} = \boldsymbol{K}^{-1}\boldsymbol{k} \qquad \equiv \qquad \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \varepsilon \end{bmatrix} = \frac{1}{DET} \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & \gamma(h)_{m,n} & \cdots & \gamma(h)_{m,1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \gamma(h)_{1,n} & \cdots & \gamma(h)_{1,1} \end{bmatrix} \times \begin{bmatrix} \dddot{\gamma}(h)_1 \\ \vdots \\ \dddot{\gamma}(h)_n \\ 1 \end{bmatrix}$$

Predicted air SO_2 level from the Kriging model



Uncertainty for the predicted air SO_2 levels from the Kriging model



Summary [1]: Work flow for using a Kriging in R

When you want to conduct predictive inference with a geostatistical model – you might want to follow these steps:

- **STEP 1**: Carry some descriptive analysis to understand the underlying spatial distribution. Make sure the underlying spatial process is from a continuous distribution (e.g., concentrations of air particulates, climate-related variables like rainfall, temperature etc., land surface elevation.

- **STEP 2**: Standardisation of spatial data to a single CRS. If the data is in decimal degrees (aka WGS84: 4326) – I highly recommend to transform them to a distance that's understandable (i.e., meters, kilometres etc.,); Use either Spherical Mercator (EPSG: 3587) (or if dealing with UK data – British National Grid [EPSG: 27700])

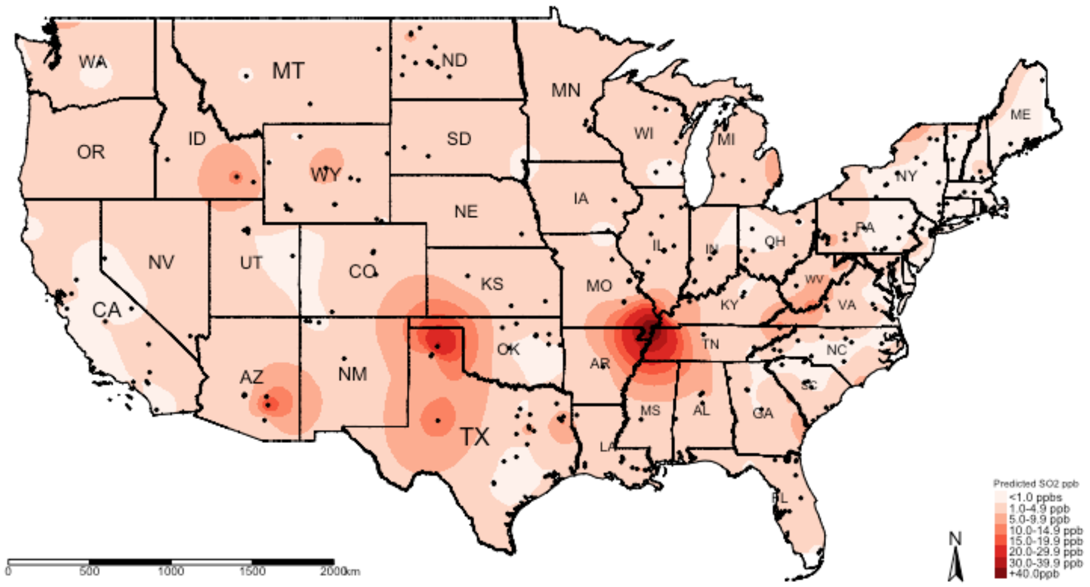
- **STEP 3**: Construction an empirical semivariogram, to estimate the values for nugget, sill and range. Next, use the initial values to determine the best fitted "Theoretical Semivariance"; here, we will use the best fitted models' nugget, sill and range to set the Kriging model.

- **STEP 4**: Setting up the raster template for Kriging. Again, make sure that the raster's CRS matches that of the point data and shapefile.

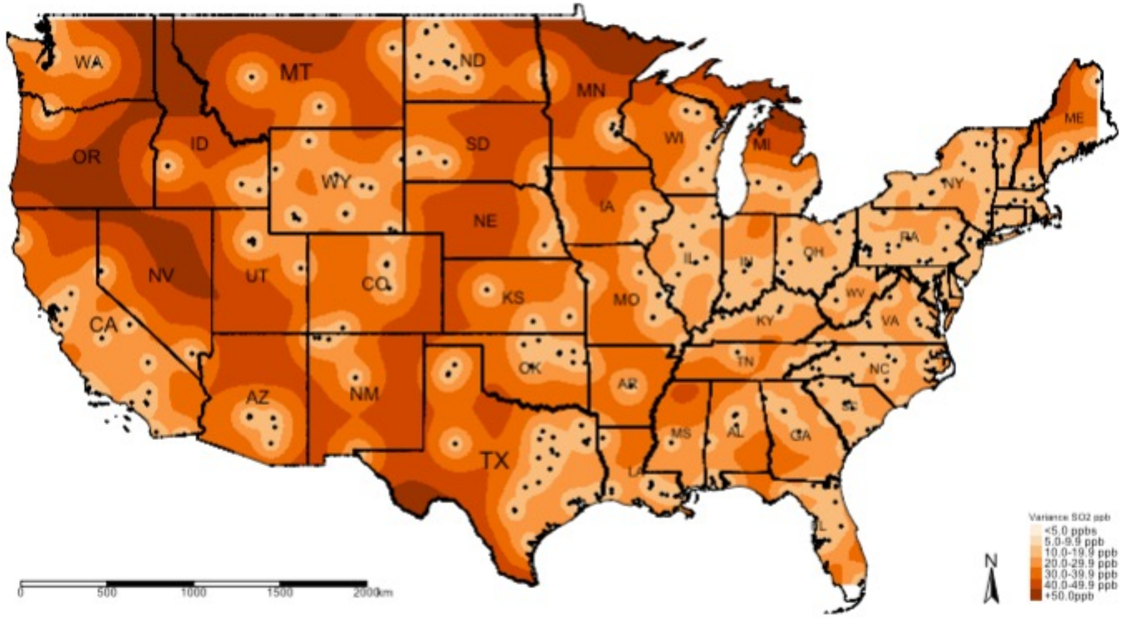
- **STEP 5**: Apply Kriging on to the raster template.

Summary [2]: Best practice for visualisation and interpretation

Predicted air SO_2 level from the Kriging model



Uncertainty for the predicted air SO_2 levels from the Kriging model



Interpretation: The **nugget** is a small value of 3.6, which is a strong indication of larger spatial autocorrelation in the concentrations for SO_2 across sampling sites in USA. The **range** is 296,255m which indicates that any separation distance with values of 296,255m and above means that spatial autocorrelation in the observed levels of SO_2 between points are no longer similar. For the **partial sill**, within this range for the Semivariance i.e., 3.6 and 55.9 – is the values are spatially autocorrelated

Any questions?

Important information

Not important information, but you should have some awareness though