# Kriging Demystified

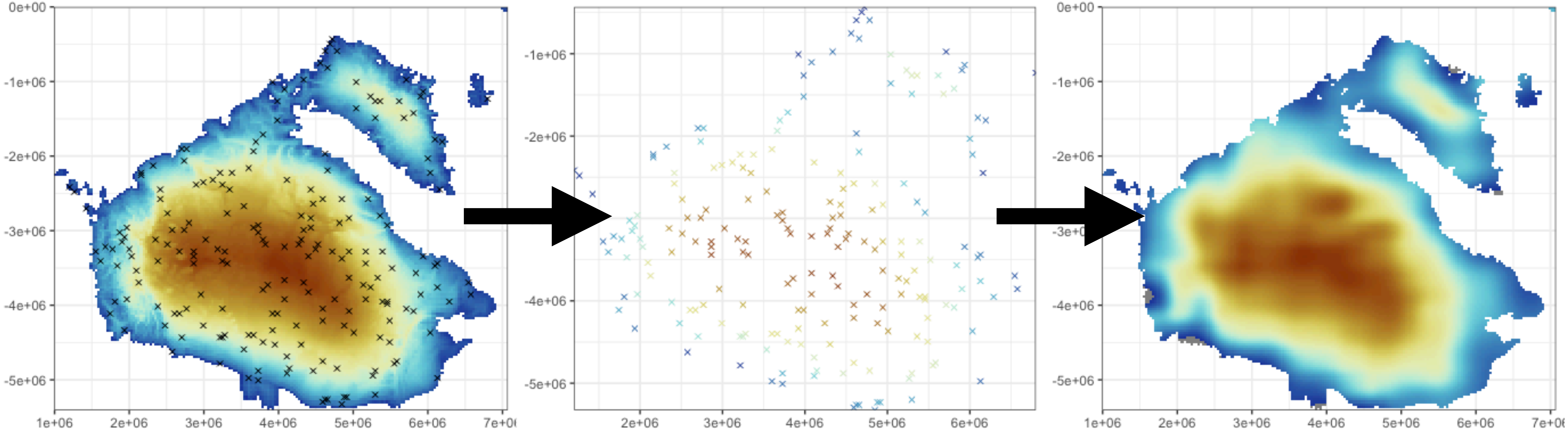
## What’s going on under the hood of our favourite interpolator?

**Lachlan Astfalck**

# What is it, why do we care?

Kriging is one of the most influential interpolation methods in statistics, geosciences, engineering, and anything else that needs interpolation.

If you’ve asked SPSS or GIS to interpolate something before it’s probably using Kriging or something related.



The multiple regression procedure for arriving at the ***best linear unbiased*** [predictor] or ***best linear weighted moving average*** [predictor] of the ore grade of an ore block (of any size) by assigning an optimum set of weights to all the available and relevant data inside and outside the ore block

Danie G. Krige

The multiple regression procedure for arriving at the ***best linear unbiased*** [predictor] or ***best linear weighted moving average*** [predictor] of the ore grade of an ore block (of any size) by assigning an optimum set of weights to all the available and relevant data inside and outside the ore block


Danie G. Krige

Given a covariance structure, Kriging estimates the unknown random-process mean with the **best linear unbiased estimator**.

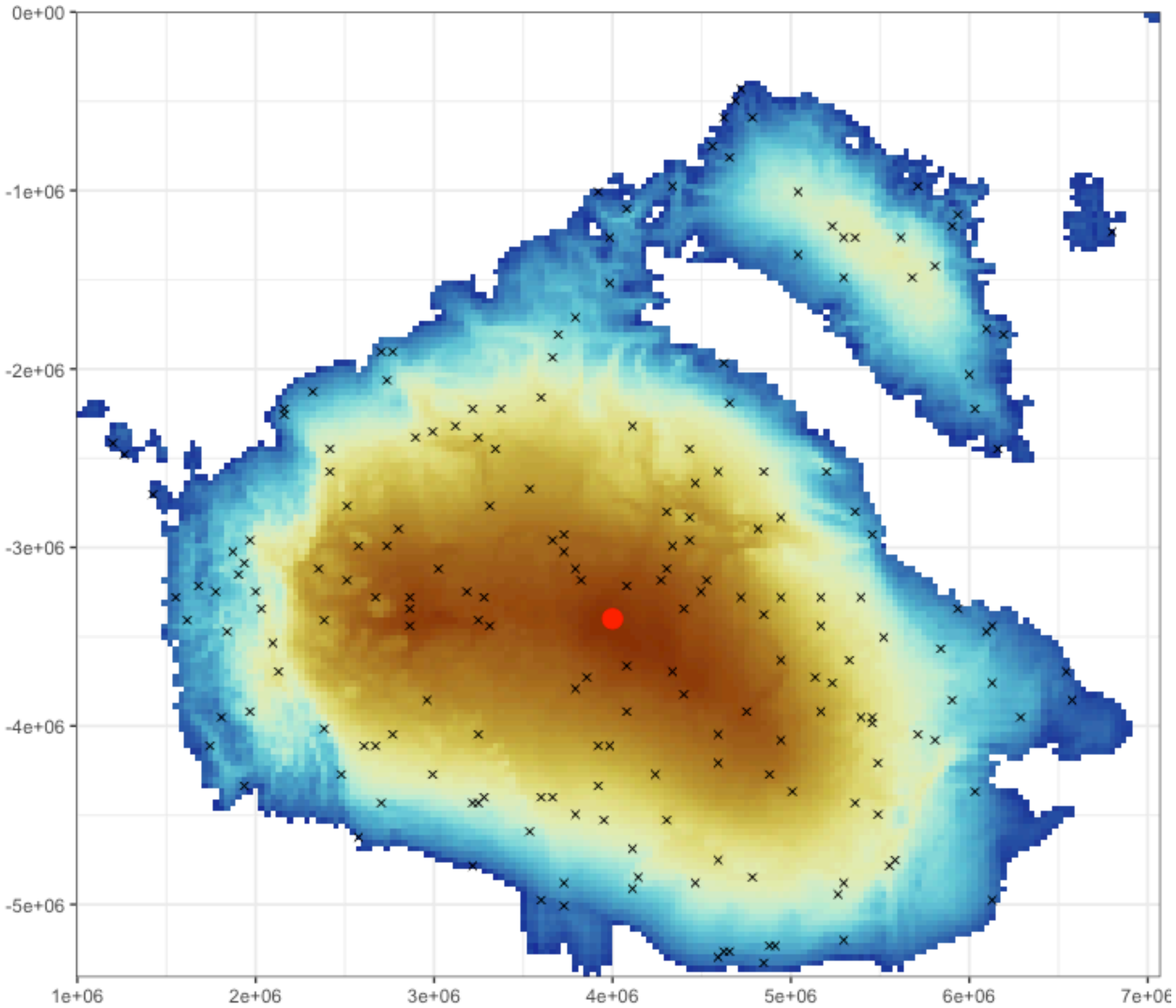
Georges Matheron

Given a covariance structure, Kriging estimates the unknown random-process mean with the **best linear unbiased estimator**.

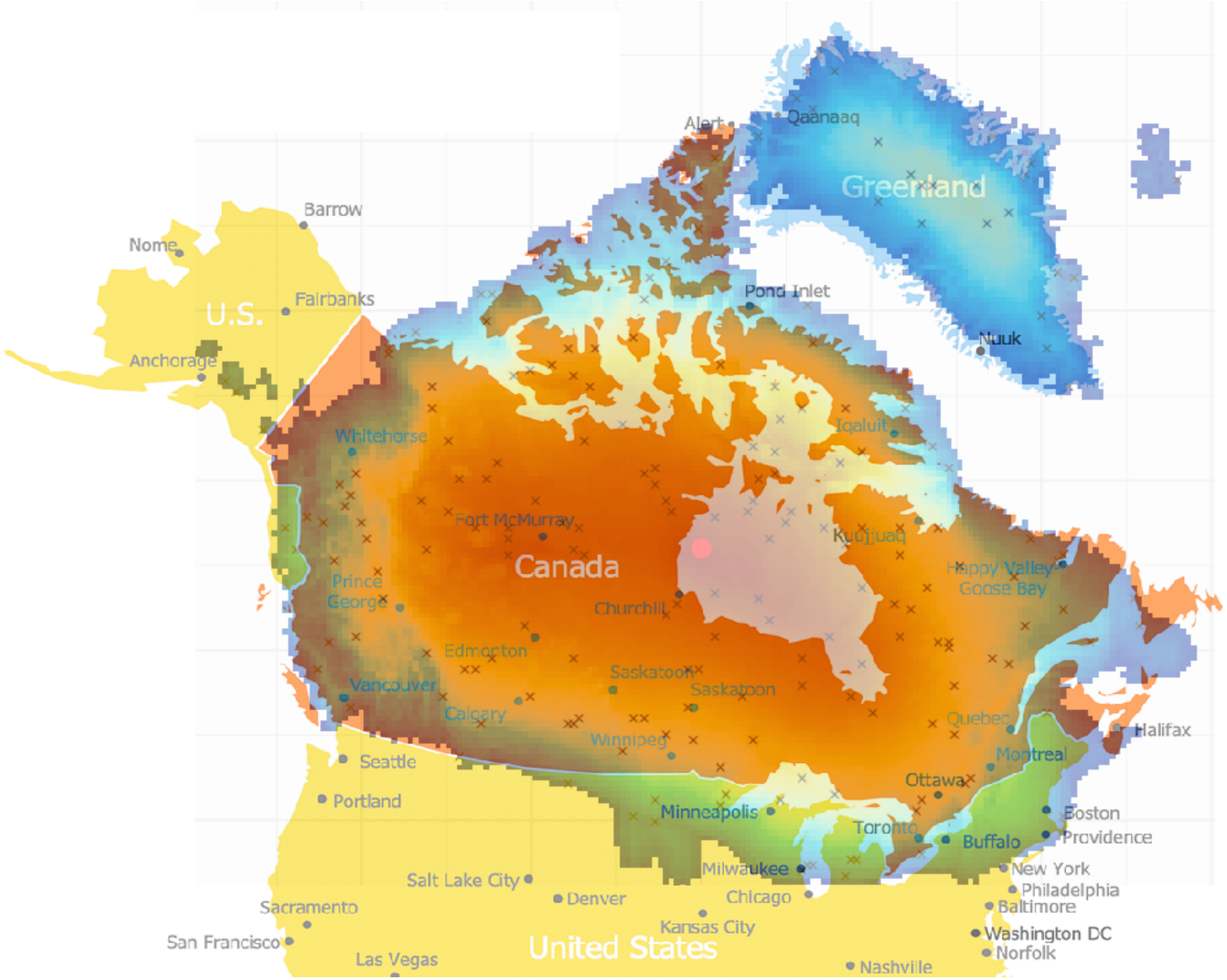

Georges Matheron

***“Everything is related to everything else, but near things are more related than distant things” -*** Tobler’s First Law of Geography

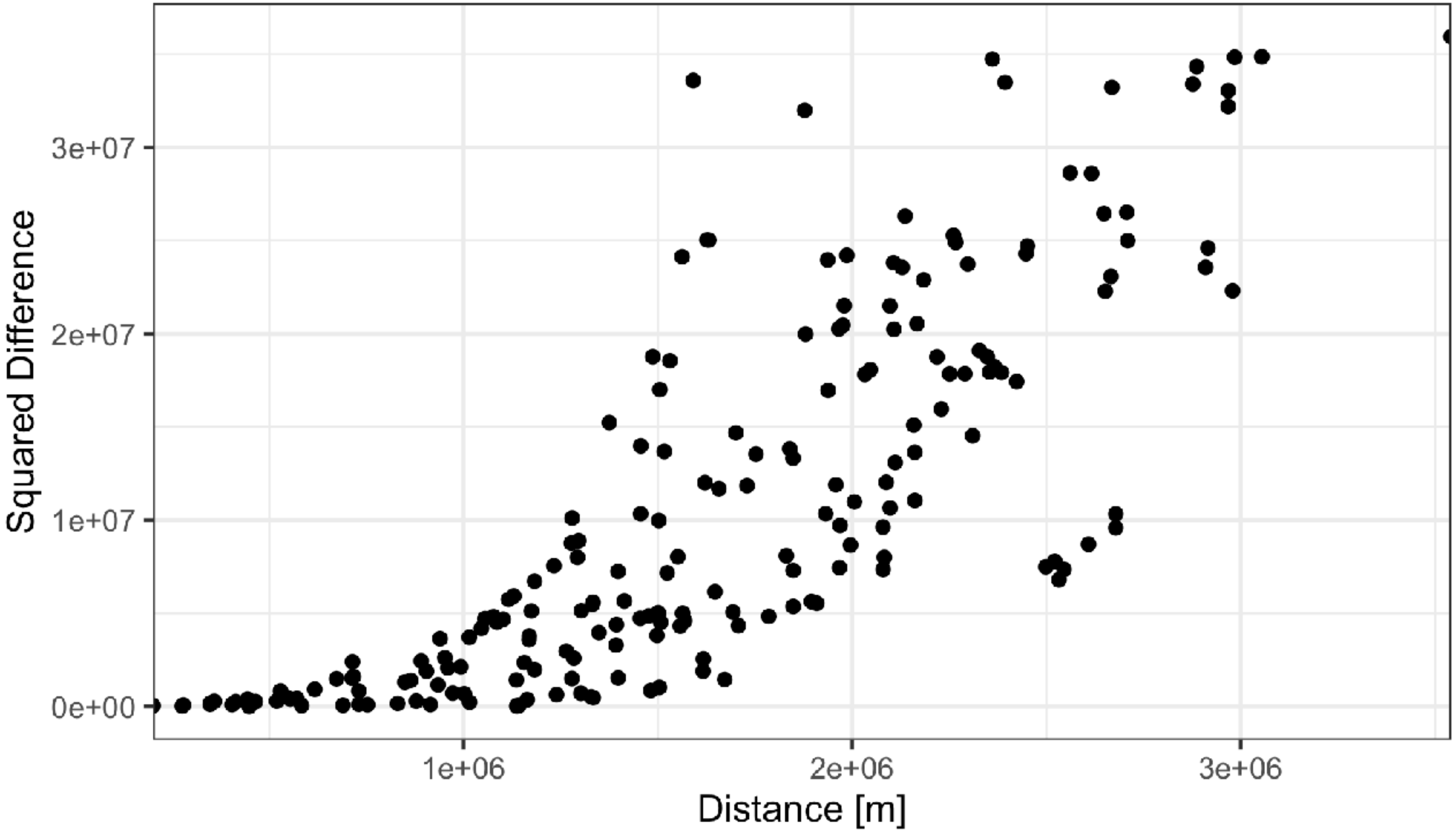
# Building up interpolation



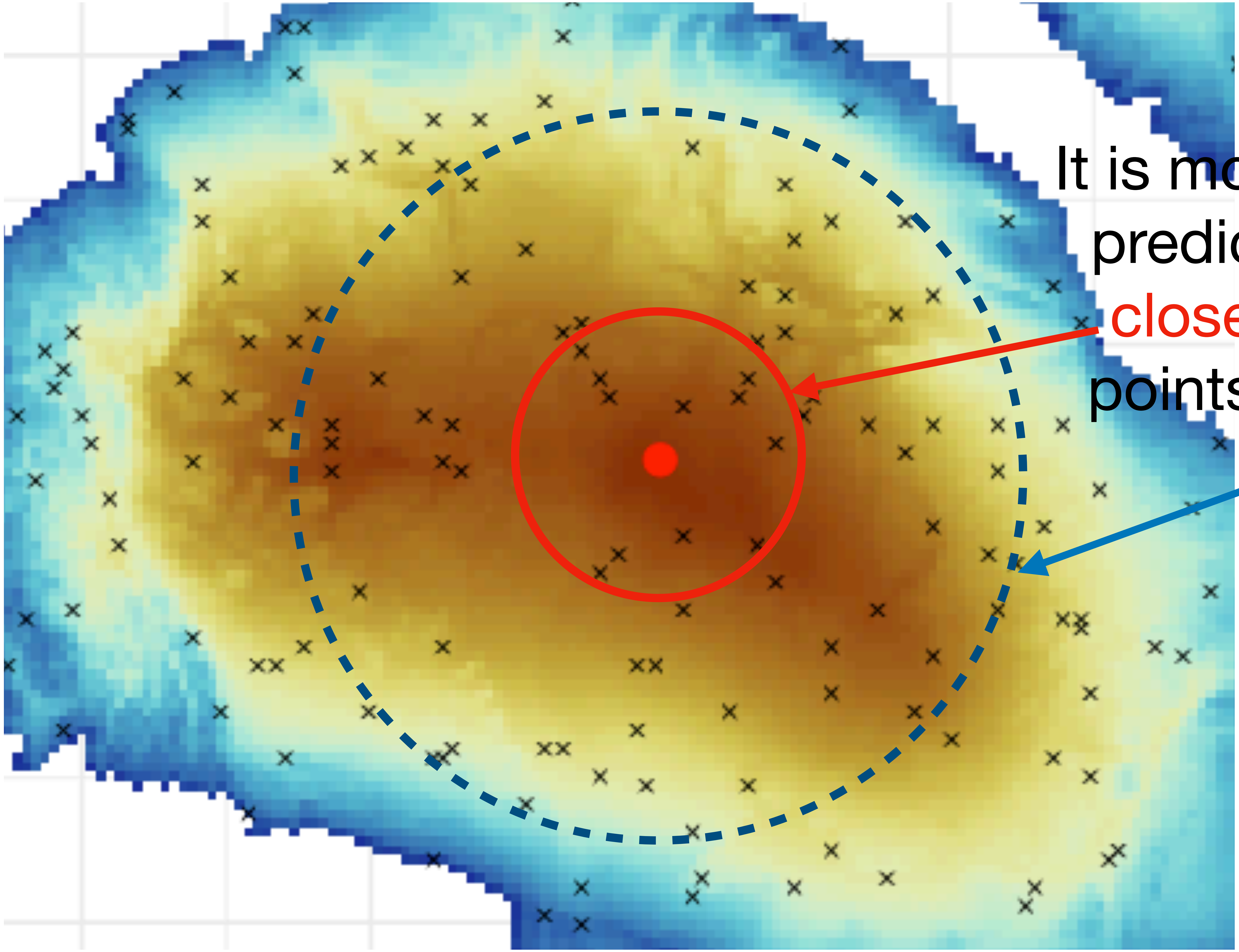
# Building up interpolation



# Building up interpolation




Tobler's First Law


It is more desirable to predict from points closer rather than points further away

# Sensible interpolation methods include…

Regression methods

Bi-linear interpolation

Inverse distance weighting

Nearest neighbour predictions

Spline interpolation

**Kriging**

# Cranking out the math

# Cranking out the math

Denote by $s_i$ the spatial locations and $Z(s_i)$ the observed data that may be thought of as realisations from a latent stochastic process.

# Cranking out the math

Denote by $s_i$ the spatial locations and $Z(s_i)$ the observed data that may be thought of as realisations from a latent stochastic process.

Denote the covariance between any two points as $\text{cov}(Z(s), Z(s')) \equiv C(s, s')$.

# Cranking out the math

Denote by $s_i$ the spatial locations and $Z(s_i)$ the observed data that may be thought of as realisations from a latent stochastic process.

Denote the covariance between any two points as $\text{cov}(Z(s), Z(s')) \equiv C(s, s')$.

We are going to focus on predicting the stochastic process, $Z(s_0)$, at a single location, $s_0$.

# Cranking out the math

Denote by $s_i$ the spatial locations and $Z(s_i)$ the observed data that may be thought of as realisations from a latent stochastic process.

Denote the covariance between any two points as $\text{cov}(Z(s), Z(s')) \equiv C(s, s')$.

We are going to focus on predicting the stochastic process, $Z(s_0)$, at a single location, $s_0$.

We're going to assume all data have point support.

# Cranking out the math

We want the ***best linear unbiased predictor*** of $Z(s_0)$. This is given by a weighted linear average of the observed data that minimises the mean-squared error

$$\mathbb{E}\left(\left(Z(s_0) - \sum_{i=1}^{n} a_i Z(s_i) - k\right)^2\right)$$

# Cranking out the math

Minimising this objective yields

$$\mathrm{a} = \mathrm{c}' C^{-1} \qquad k = E\left(Z(s_0)\right) + \mathrm{c}' C^{-1} E\left(\mathrm{Z}\right)$$

where $\mathrm{c} \equiv (C(s_0, s_1), \ldots, C(s_0, s_n))'$ and $C_{i,j} \equiv C(s_i, s_j)$.

(we're about to go through why…)

# Cranking out the math

We may thus say that the optimal predictor of $Z(s_0)$ is

$$Z^*(s_0) = E\left(Z(s_0)\right) + \mathrm{c}'C(\mathrm{Z} - E\,(\mathrm{Z}))$$

with prediction error

$$E\left(\left(Z(s_0) - Z^*(s_0)\right)^2\right) = C(s_0, s_0) - \mathrm{c}'C^{-1}\mathrm{c}$$

# Cranking out the math

First we need some basic identities:

$$E\left(x^2\right) = \text{var}(x) + \left(E(x)\right)^2$$

$$\text{var}(x + y) = \text{var}(x) + 2\text{cov}(x, y) + \text{var}(y)$$

$$\text{cov}(ax, y) = a\text{cov}(x, y)$$

$$\text{var}(ax) = a^2\text{var}(x)$$

$$\text{var}(x + c) = \text{var}(x)$$

# Cranking out the math

# Cranking out the math

$$E\left(\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)^2\right) = \mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) + \left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2$$

# Cranking out the math

$$E\left(\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)^2\right) = \mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) + \left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2$$

and so $k = E\left(Z(s_0)\right) - \mathrm{a}^T E\left(\mathrm{Z}\right)$ so that $\left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2 = 0.$

# Cranking out the math

$$E\left(\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)^2\right) = \mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) + \left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2$$

and so $k = E\left(Z(s_0)\right) - \mathrm{a}^T E\,(\mathrm{Z})$ so that $\left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2 = 0.$

$$\mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) = \mathrm{var}\left(Z(s_0)\right) - 2\mathrm{cov}\left(Z(s_0), \mathrm{Z}\right)\mathrm{a} + \mathrm{a}^T\mathrm{var}\,(\mathrm{Z})\,\mathrm{a}$$

# Cranking out the math

$$E\left(\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)^2\right) = \mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) + \left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2$$

and so $k = E\left(Z(s_0)\right) - \mathrm{a}^T E\left(\mathrm{Z}\right)$ so that $\left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2 = 0.$

$$\mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) = \mathrm{var}\left(Z(s_0)\right) - 2\mathrm{cov}\left(Z(s_0), \mathrm{Z}\right)\mathrm{a} + \mathrm{a}^T\mathrm{var}\left(\mathrm{Z}\right)\mathrm{a}$$

Differentiating with respect to a:

# Cranking out the math

$$E\left(\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)^2\right) = \mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) + \left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2$$

and so $k = E\left(Z(s_0)\right) - \mathrm{a}^T E\left(\mathrm{Z}\right)$ so that $\left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2 = 0.$

$$\mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) = \mathrm{var}\left(Z(s_0)\right) - 2\mathrm{cov}\left(Z(s_0), \mathrm{Z}\right)\mathrm{a} + \mathrm{a}^T\mathrm{var}\left(\mathrm{Z}\right)\mathrm{a}$$

Differentiating with respect to a:

$$-2\mathrm{cov}\left(Z(s_0), \mathrm{Z}\right) + 2\mathrm{a}^T\mathrm{var}\left(\mathrm{Z}\right) = 0$$

# Cranking out the math

$$E\left(\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)^2\right) = \mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) + \left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2$$

and so $k = E\left(Z(s_0)\right) - \mathrm{a}^T E\left(\mathrm{Z}\right)$ so that $\left(E\left(Z(s_0) - \mathrm{a}^T\mathrm{Z} - k\right)\right)^2 = 0.$

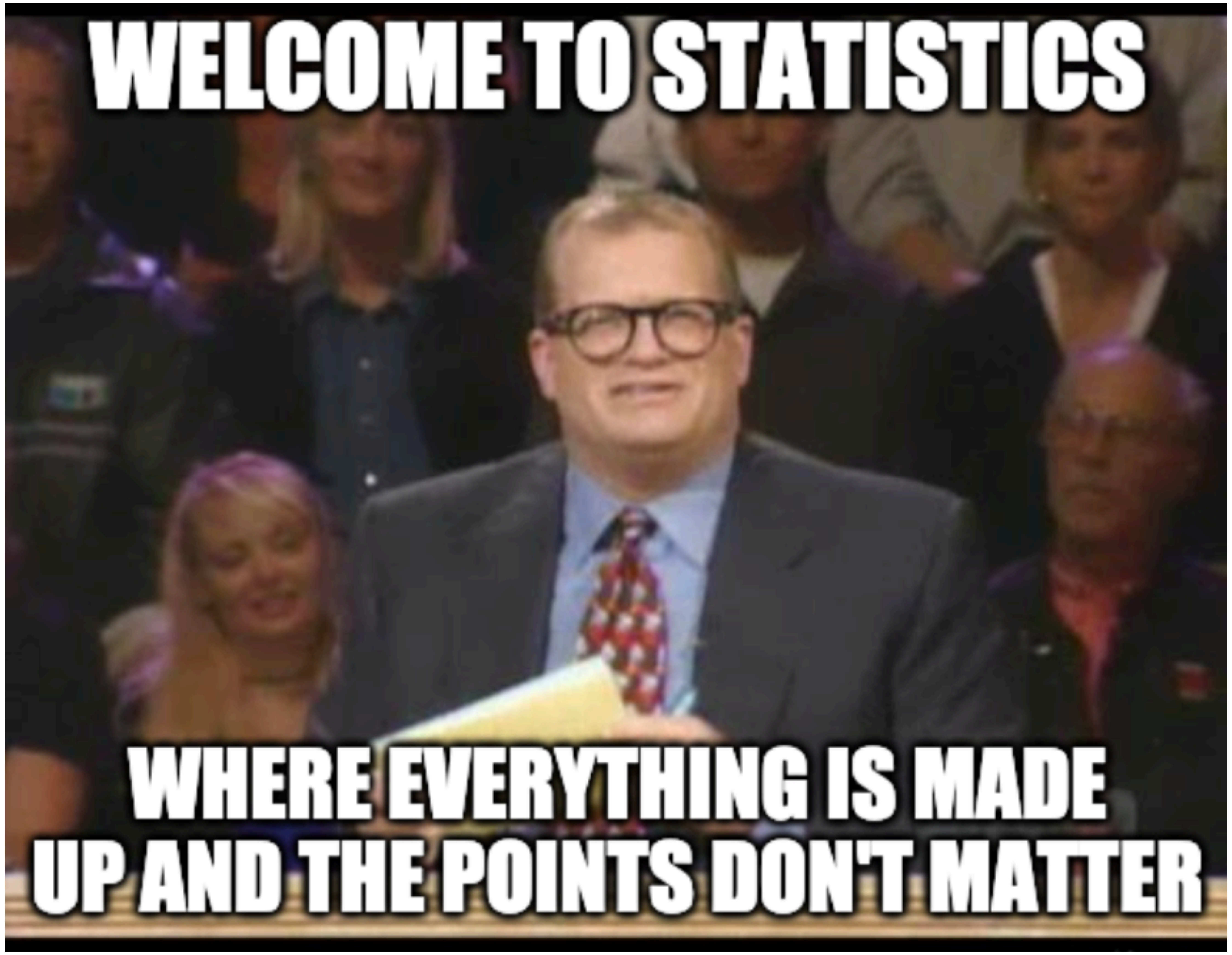
$$\mathrm{var}\left(Z(s_0) - \mathrm{a}^T\mathrm{Z}\right) = \mathrm{var}\left(Z(s_0)\right) - 2\mathrm{cov}\left(Z(s_0), \mathrm{Z}\right)\mathrm{a} + \mathrm{a}^T\mathrm{var}\left(\mathrm{Z}\right)\mathrm{a}$$

Differentiating with respect to a:

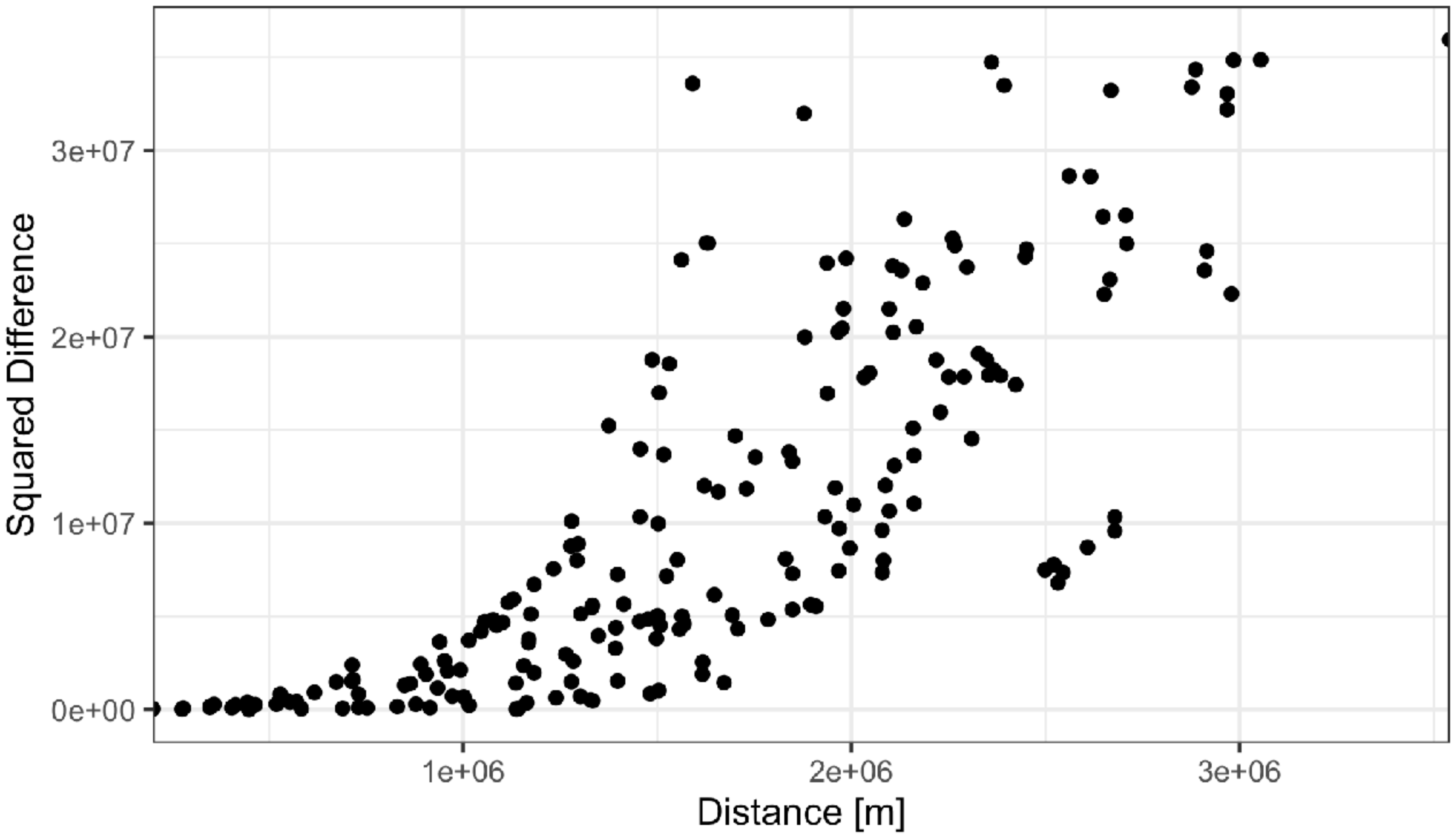
$$-2\mathrm{cov}\left(Z(s_0), \mathrm{Z}\right) + 2\mathrm{a}^T\mathrm{var}\left(\mathrm{Z}\right) = 0$$

and so $\mathrm{a}^T = \mathrm{cov}\left(Z(s_0), \mathrm{Z}\right)\mathrm{var}\left(\mathrm{Z}\right)^{-1}$

# But how do we figure out covariance?



# But how do we figure out covariance?



# The semivariogram and covariance function

# The semivariogram and covariance function

Recall that this squared distance is very reminiscent of the definition of variance: $\text{var}[x] = E\left((x - E[x])^2\right)$.

# The semivariogram and covariance function

Recall that this squared distance is very reminiscent of the definition of variance: var$[x] = E\left((x - E[x])^2\right)$.

A mathematical interpretation of Tobler's Law is that points close to each other have a high covariance (and thus observing one point resolves much of the uncertainty in another).

# The semivariogram and covariance function

Recall that this squared distance is very reminiscent of the definition of variance: var$[x] = E\left((x - E[x])^2\right)$.

A mathematical interpretation of Tobler's Law is that points close to each other have a high covariance (and thus observing one point resolves much of the uncertainty in another).

The semivariogram is a way of describing the expected variance between two points as a function of distance

# The semivariogram and covariance function

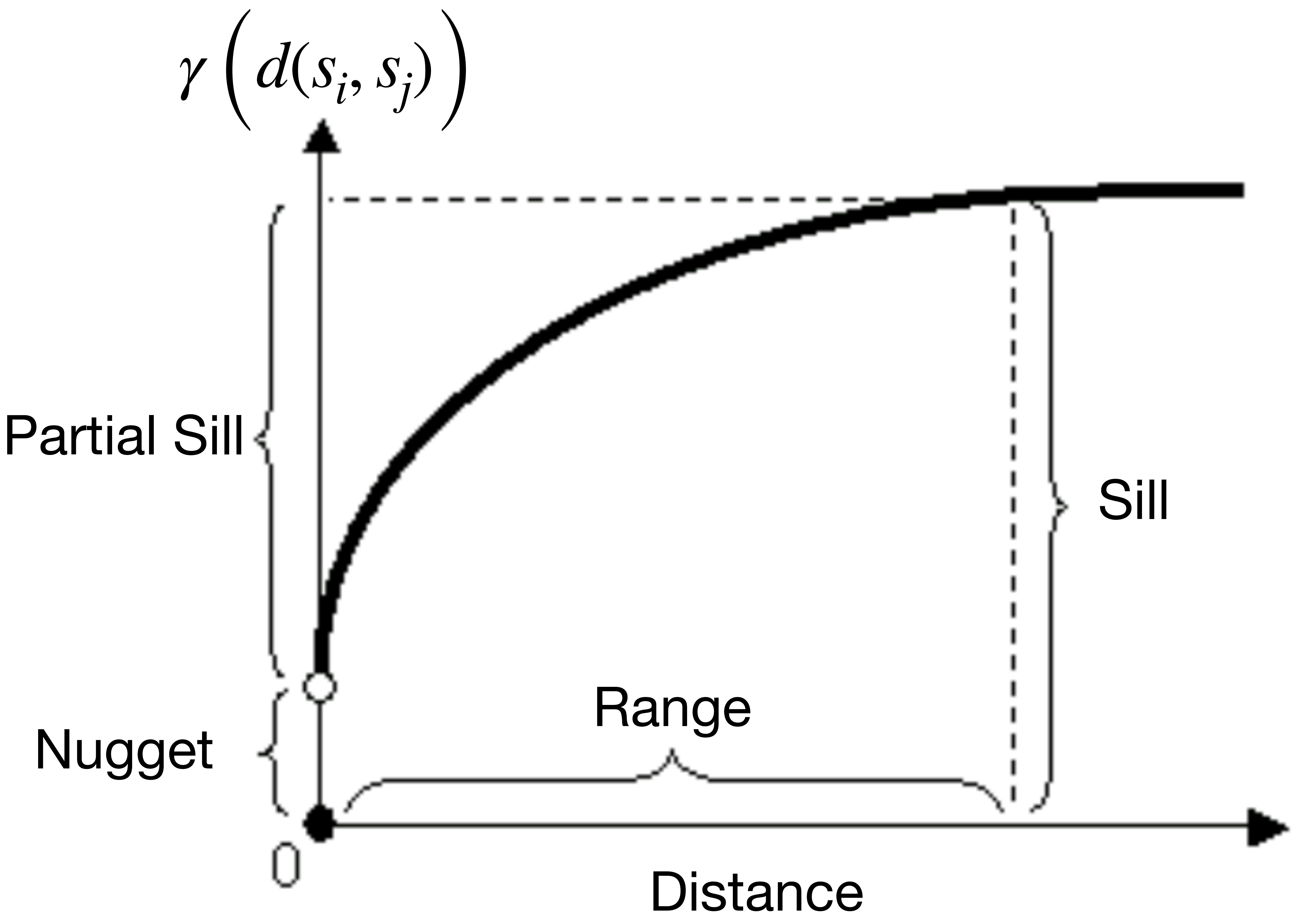
Recall that this squared distance is very reminiscent of the definition of variance: var$[x] = E\left((x - E[x])^2\right)$.

A mathematical interpretation of Tobler's Law is that points close to each other have a high covariance (and thus observing one point resolves much of the uncertainty in another).

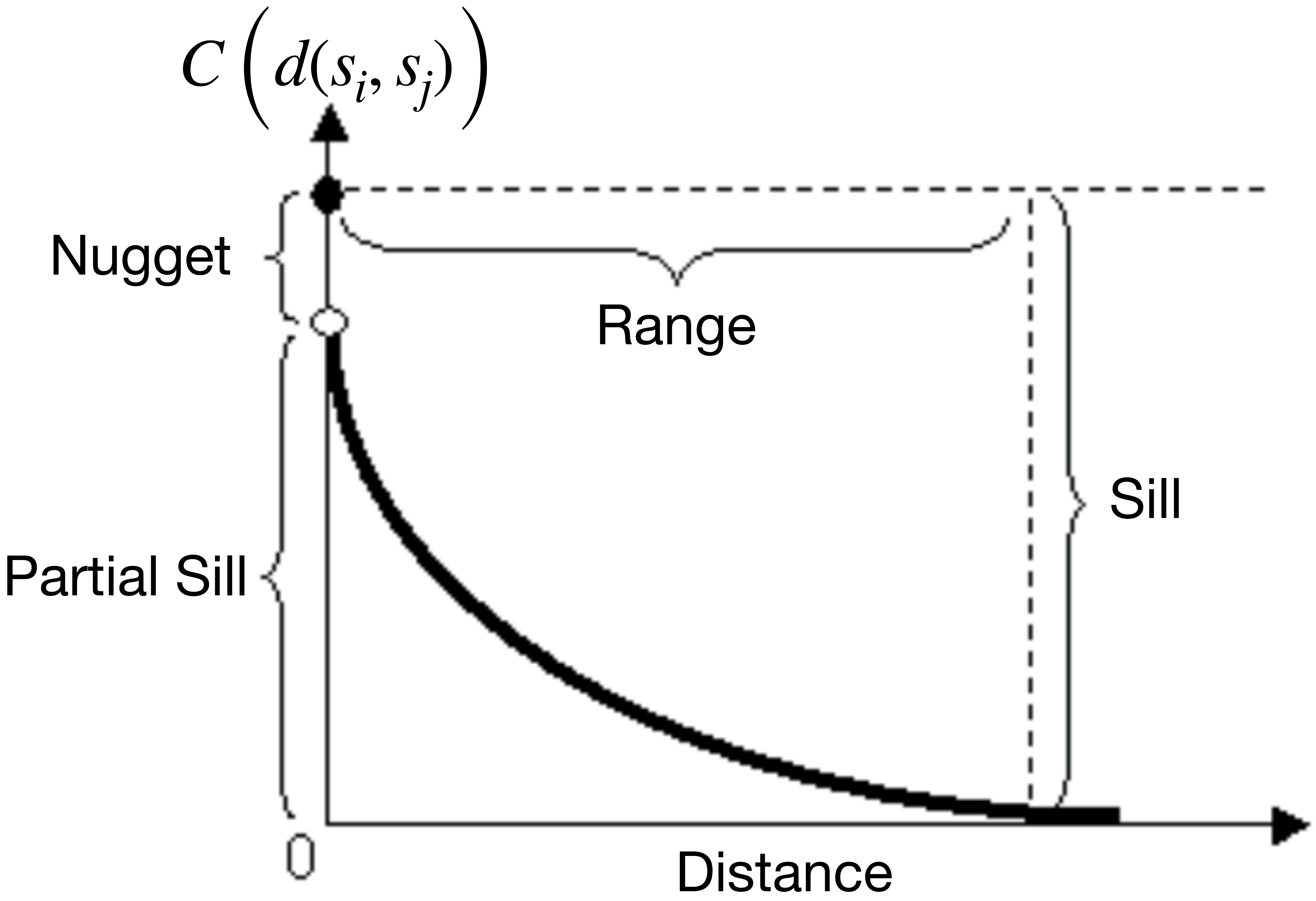
The semivariogram is a way of describing the expected variance between two points as a function of distance

The covariance function describes how two points co-vary as a function of distance

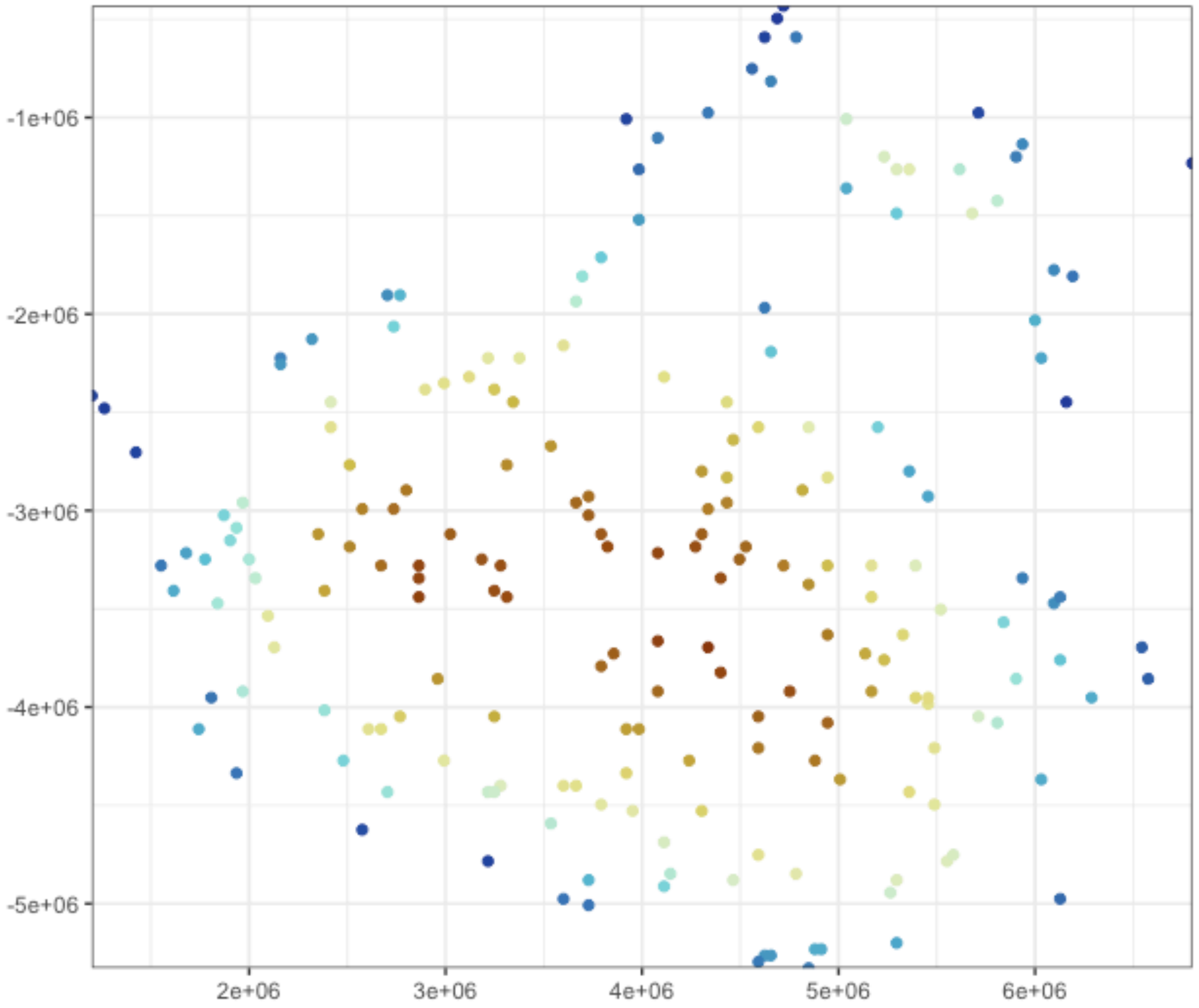
# The semivariogram



# The covariance function



# Empirical Semivariogram



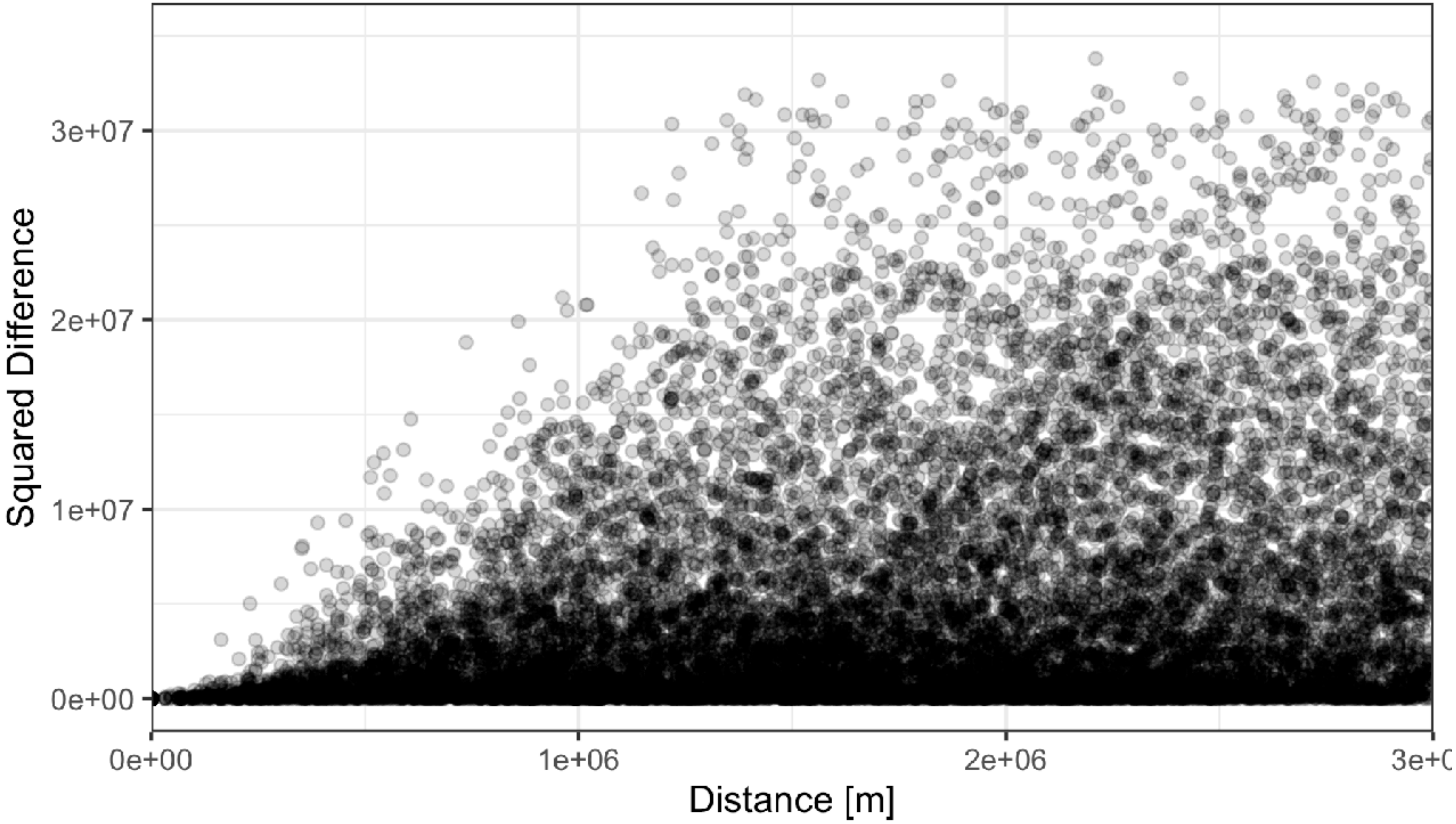

Once we have a fit semivariogram or covariance function we're good to go

# Empirical Semivariogram

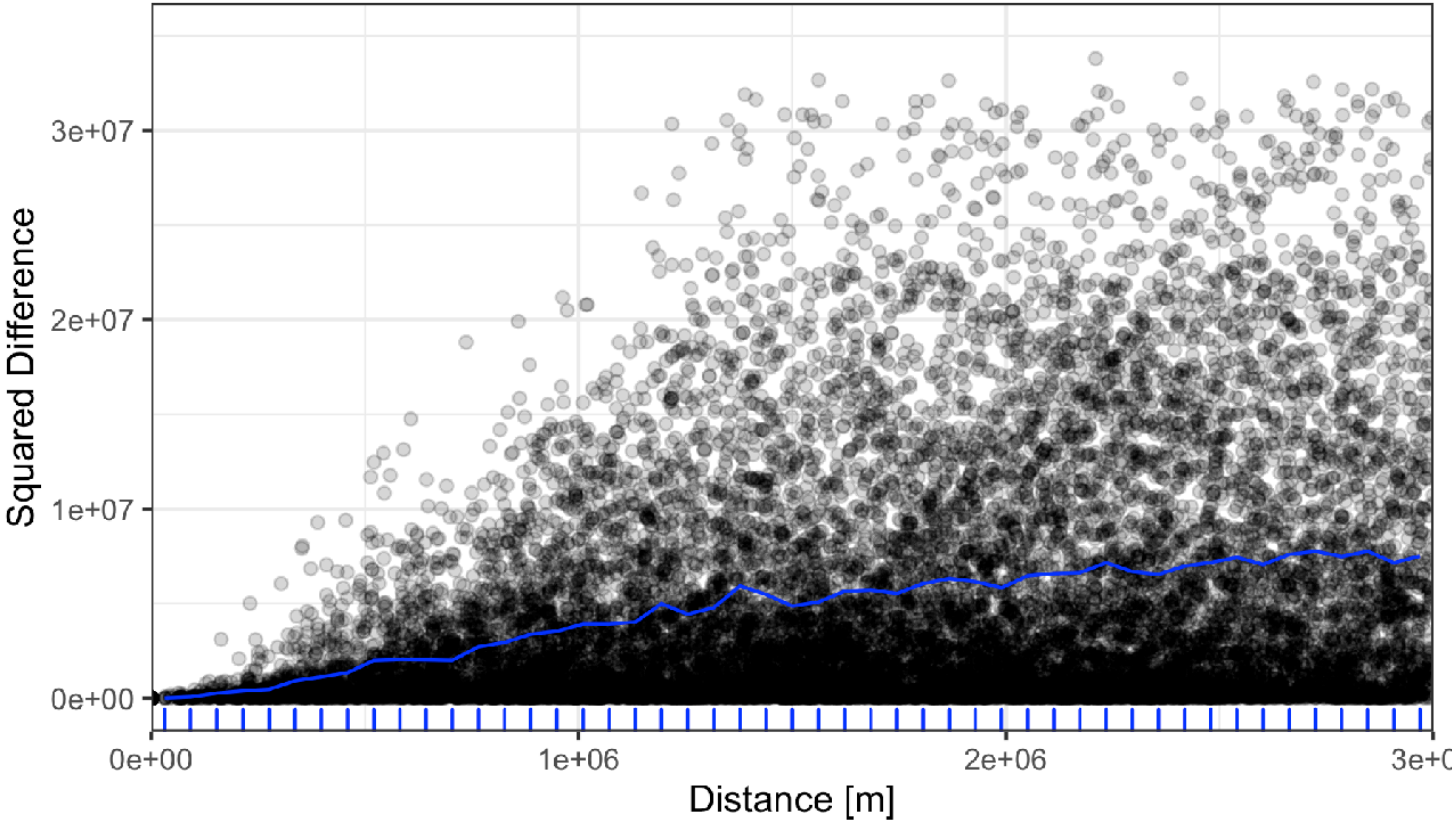
$$\hat{\gamma}(h) = \frac{1}{n_h} \sum_{s_i, s_k \in N(h)} \left( Z\left(s_i\right) - Z\left(s_k\right) \right)^2$$

for some choice of $h = | s_i - s_k |$

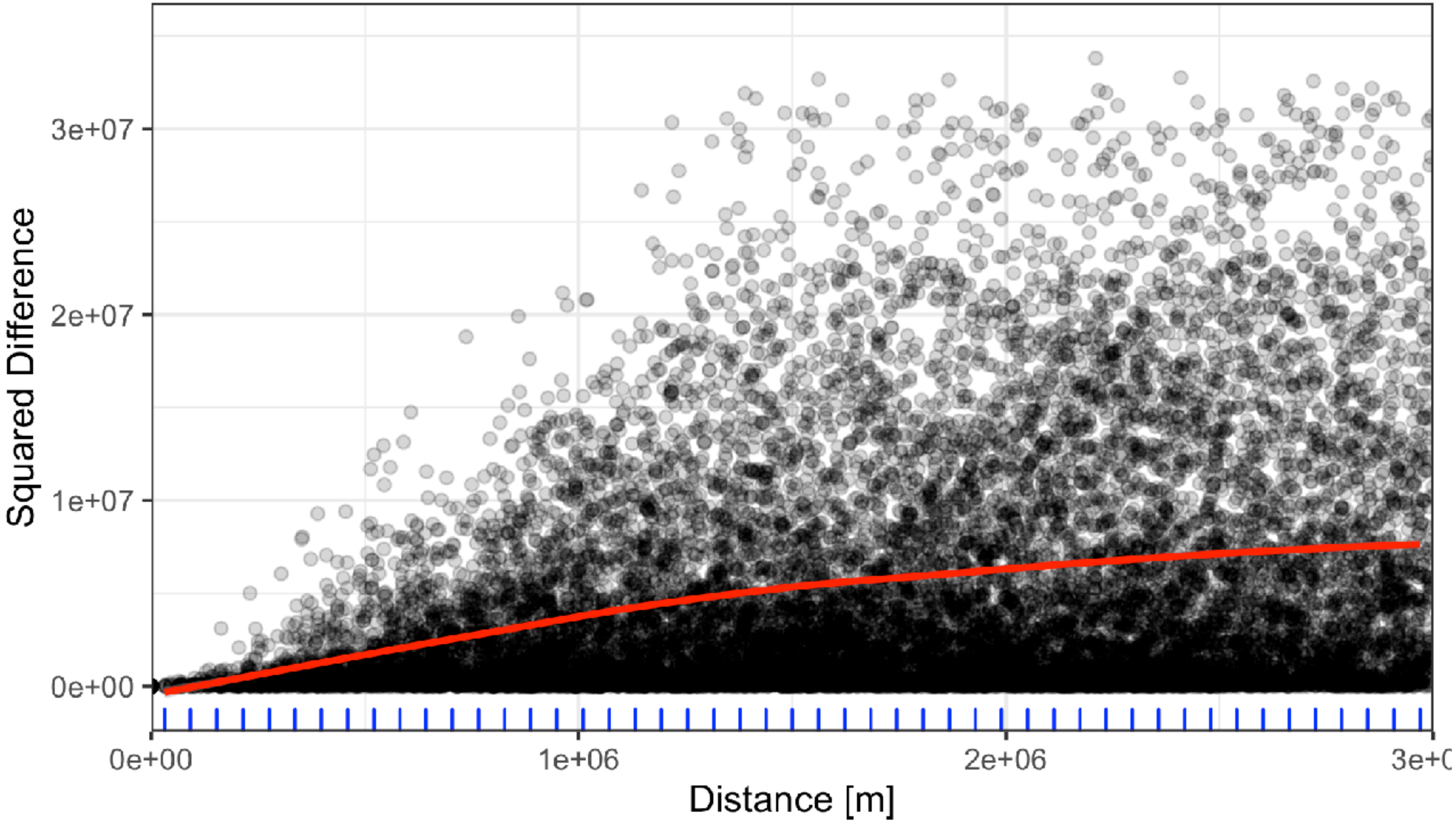
# Empirical Semivariogram



# Empirical Semivariogram



# Empirical Semivariogram



# Other methods of modelling covariance

# Other methods of modelling covariance

Empirically fitting semivariograms is appropriate for pedagogy but is quite an ‘old’ way of solving the problem.

# Other methods of modelling covariance

Empirically fitting semivariograms is appropriate for pedagogy but is quite an ‘old’ way of solving the problem.

More often covariance functions are usually specified and fit (generally via likelihood methods) manually or automatically.

# Other methods of modelling covariance

Empirically fitting semivariograms is appropriate for pedagogy but is quite an ‘old’ way of solving the problem.

More often covariance functions are usually specified and fit (generally via likelihood methods) manually or automatically.

This is most likely what is happening in the background of an kriging software that you are using.

# Other methods of modelling covariance

Empirically fitting semivariograms is appropriate for pedagogy but is quite an ‘old’ way of solving the problem.

More often covariance functions are usually specified and fit (generally via likelihood methods) manually or automatically.

This is most likely what is happening in the background of an kriging software that you are using.

Be aware of what covariance function is being assumed - different functions yield drastically different interpolations.

# Time to Predict

# Time to Predict

1. We have the data

# Time to Predict

1. We have the data
2. We have a covariance function that models the data well

# Time to Predict

1. We have the data
2. We have a covariance function that models the data well
3. We're all over the mathematical theory

# Time to Predict

1. We have the data
2. We have a covariance function that models the data well
3. We're all over the mathematical theory
4. Now we're ready to predict!

# Time to Predict

1. We have the data
2. We have a covariance function that models the data well
3. We're all over the mathematical theory
4. Now we're ready to predict!

$$E\left(Z^*(s_0)\right) = E\left(Z(s_0)\right) + \mathrm{c}'C(\mathrm{Z} - E\,(\mathrm{Z}))$$

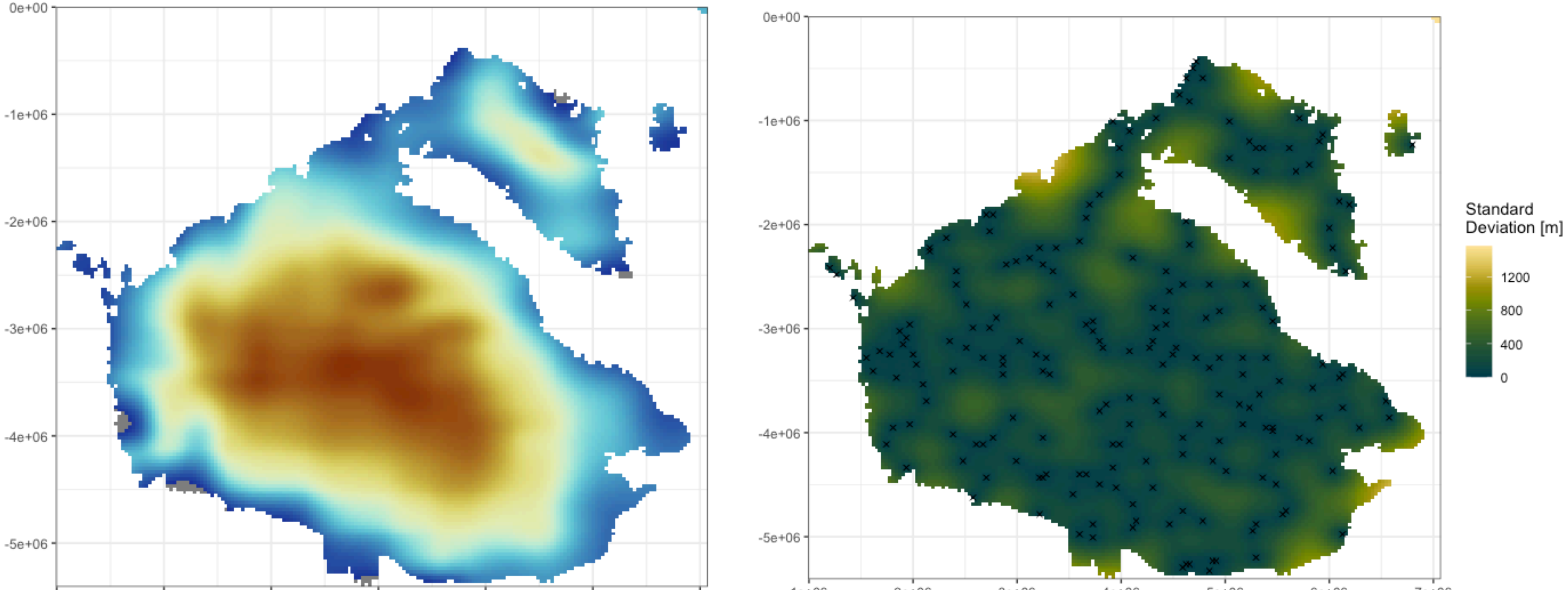
# Time to Predict

1. We have the data
2. We have a covariance function that models the data well
3. We're all over the mathematical theory
4. Now we're ready to predict!

$$E\left(Z^*(s_0)\right) = E\left(Z(s_0)\right) + \mathrm{c}' C(\mathrm{Z} - E(\mathrm{Z}))$$

$$\mathrm{var}\left(Z^*(s_0)\right) = C(s_0, s_0) - \mathrm{c}' C^{-1}\mathrm{c}$$

# Time to Predict




$E\left(Z^*(s)\right)$ $\mathrm{var}\left(Z^*(s)\right)$

# A rose by any other name…

Simple Kriging

Ordinary Kriging

Universal Kriging

Disjunctive Kriging

Linear Regression

Gaussian Processes

Kalman Filtering

Gaussian Markov Random Fields

Wiener-Kolmogorov Predictions

Spline Regression in Reproducing Kernel Hilbert Spaces

Bayes linear analysis

# A rose by any other name…

Simple Kriging

Ordinary Kriging

Universal Kriging

Disjunctive Kriging

Linear Regression

Gaussian Processes

Kalman Filtering

Gaussian Markov Random Fields

Wiener-Kolmogorov Predictions

Spline Regression in Reproducing Kernel Hilbert Spaces

Bayes linear analysis

The essence/ideas underpinning kriging are everywhere! If you’re ever interpolating or regressing, be mindful of where it may be.

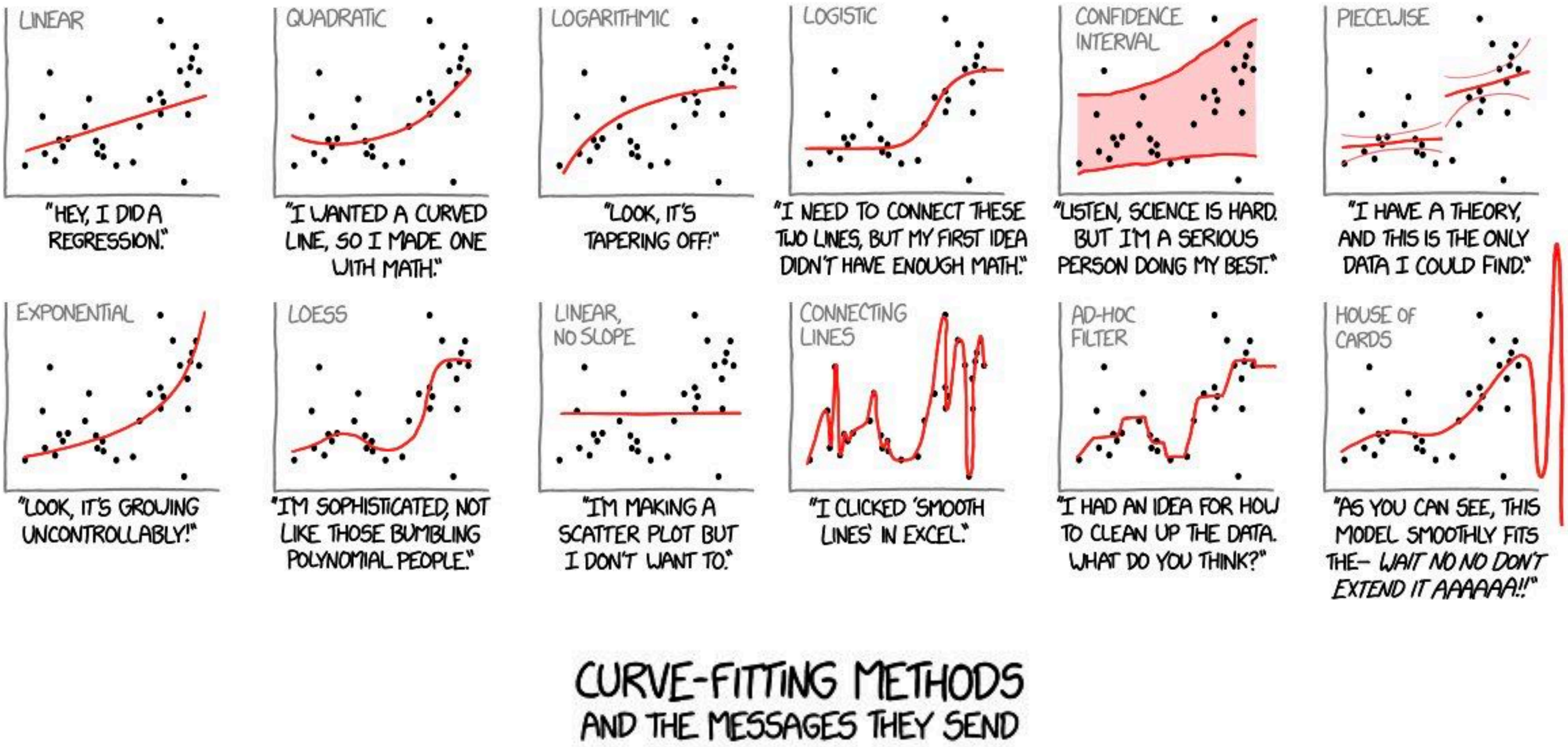
# What we haven’t talked about

Stationarity, types of covariance functions and positive definiteness

Methods of inference

Kriging with big data

Assumptions of Gaussinity


LINEAR
"HEY, I DID A REGRESSION."
QUADRATIC
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPERING OFF!"
LOGISTIC
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."
CONFIDENCE INTERVAL
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY!"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
LINEAR, NO SLOPE
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."
CONNECTING LINES
"I CLICKED 'SMOOTH LINES' IN EXCEL."
AD-HOC FILTER
"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"
HOUSE OF CARDS
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE— WAIT NO NO DON'T EXTEND IT AAAAAA!!"
CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND