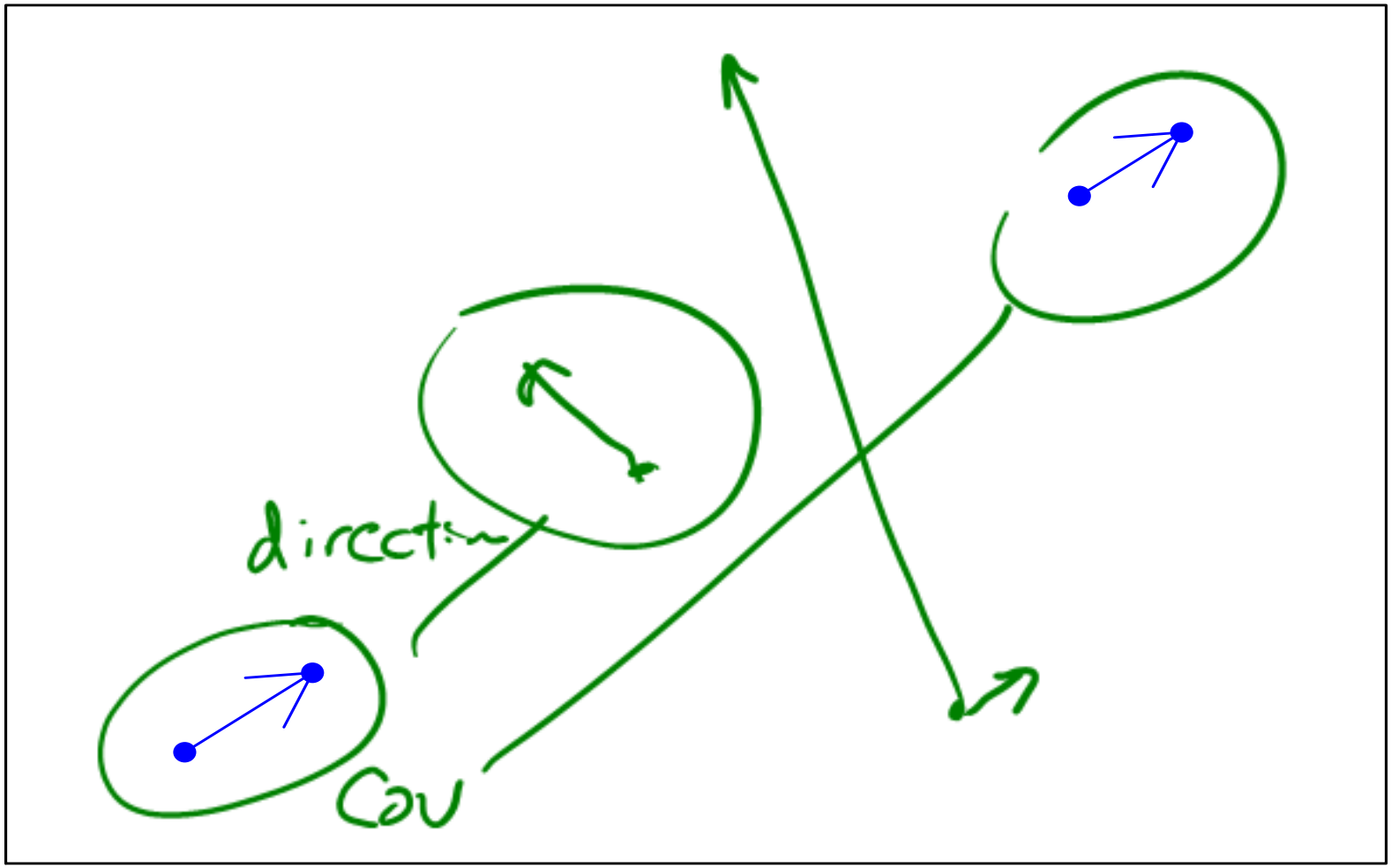
Semivariograms

- All forms of kriging assume you know Cov $(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$
- Usually, need to estimate this
  - Big problem: only have one observation at $\boldsymbol{s}_i$ and one at $\boldsymbol{s}_j$
  - usual data-based estimate won't work!
  - Also, need vector of Cov $(Z(\boldsymbol{s}_0), Z(\boldsymbol{s}_i))$ when haven't observed $Z(\boldsymbol{s}_0)$
- Need a model! How does Cov $(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$ depend on:
  - distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$
  - direction from $\boldsymbol{s}_i$ to $\boldsymbol{s}_j$
  - location of $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ in the study area

- Two pairs of points, same direction, same distance, different parts of study area
- Same covariance?

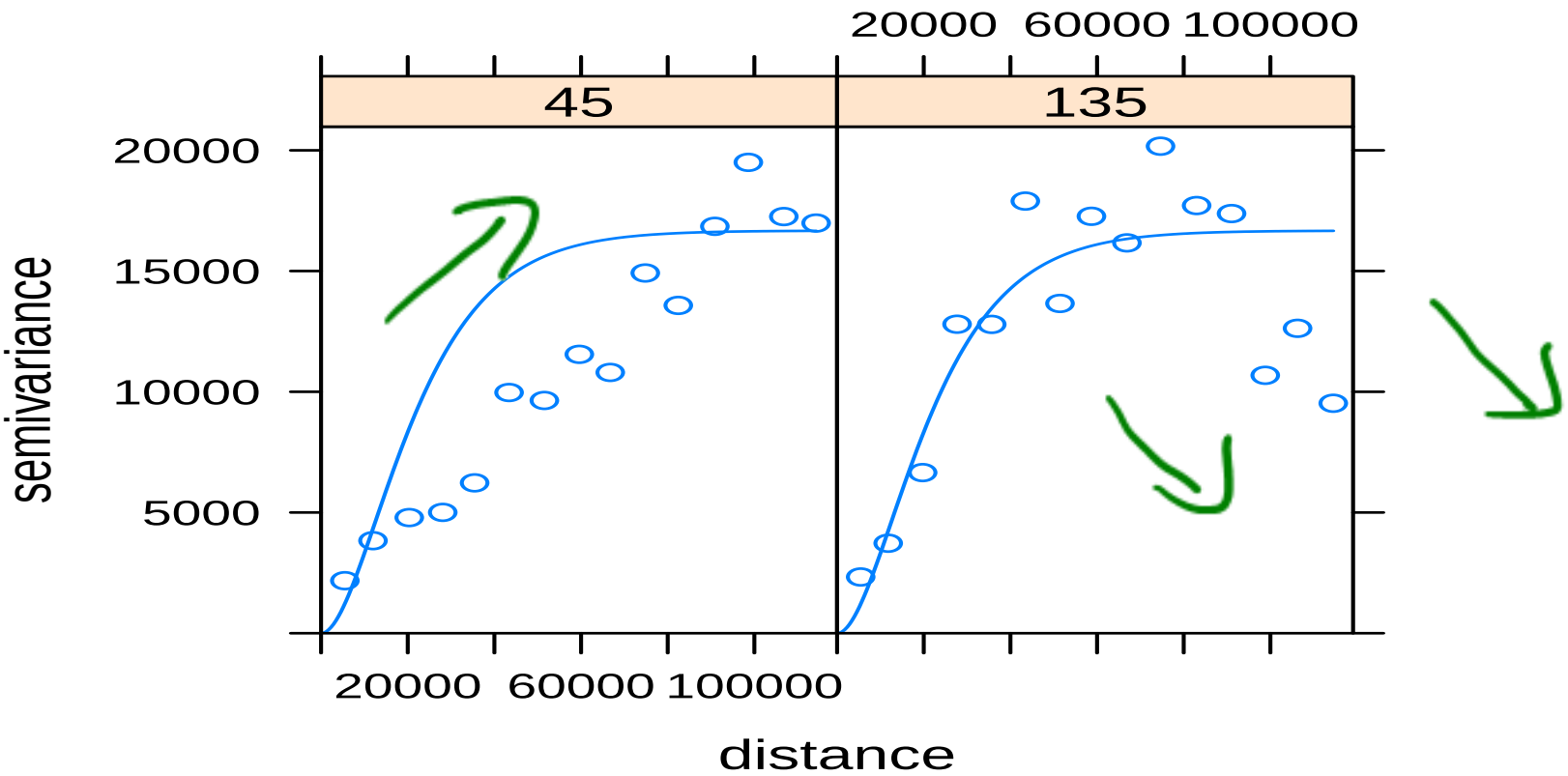


Stationary spatial processes

- 2nd order stationary
  - $\mu(s)$ constant across study area
  - Cov $(Z(s), Z(s+h))$ same across study area
    h specifies a particular distance and direction
  - So in previous picture, the two pairs of points would have same Cov
- instrinsic stationarity
  - Var $(Z(s) - Z(s+h))$ same everywhere
  - Slightly weaker assumption
  - Some really care about the difference. I don't.
- We'll assume 2nd order stationarity

Isotropic spatial process

- Cov $(Z(s), Z(s+h))$ same in all directions
  Only depends on distance between two points, i.e. $||h||$
- Anisotropic: Cov $(Z(s), Z(s+h))$ depends on direction



Anisotropic spatial process

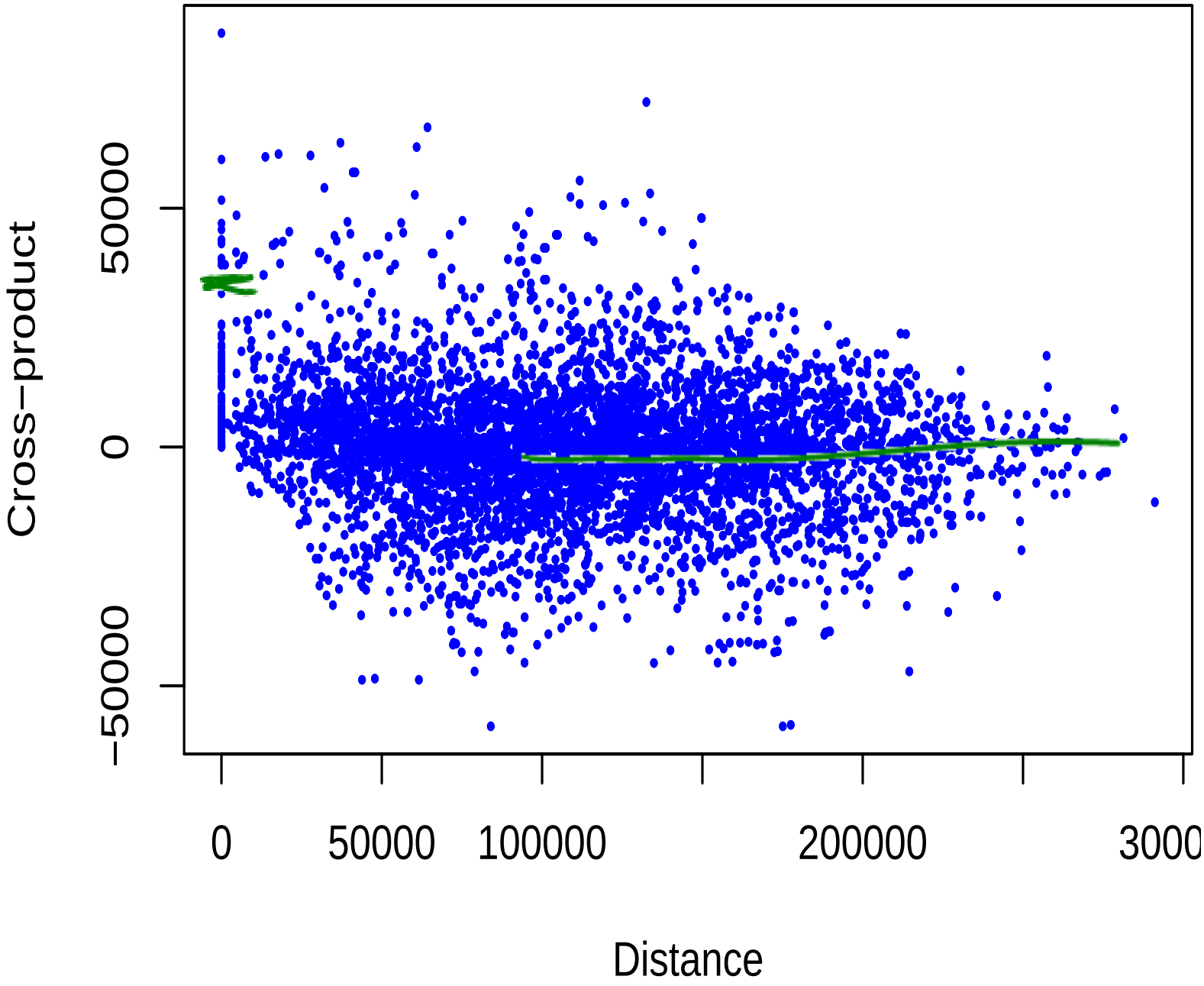
- Geometric anisotropy: simple scaling & rotation $\rightarrow$ isotropy
  draw picture on board
- For the most part, we will assume isotropy
  - If geometric anisotropy, can transform coordinates to make isotropic
  - If general anisotropy, can repeat what we're about to do in different directions
    more complications, more details, no change in concept
- So, Cov $(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$ depends on $||\boldsymbol{s}_i - \boldsymbol{s}_j||$, i.e. Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$
  - If working over large areas, should use great circle distance instead

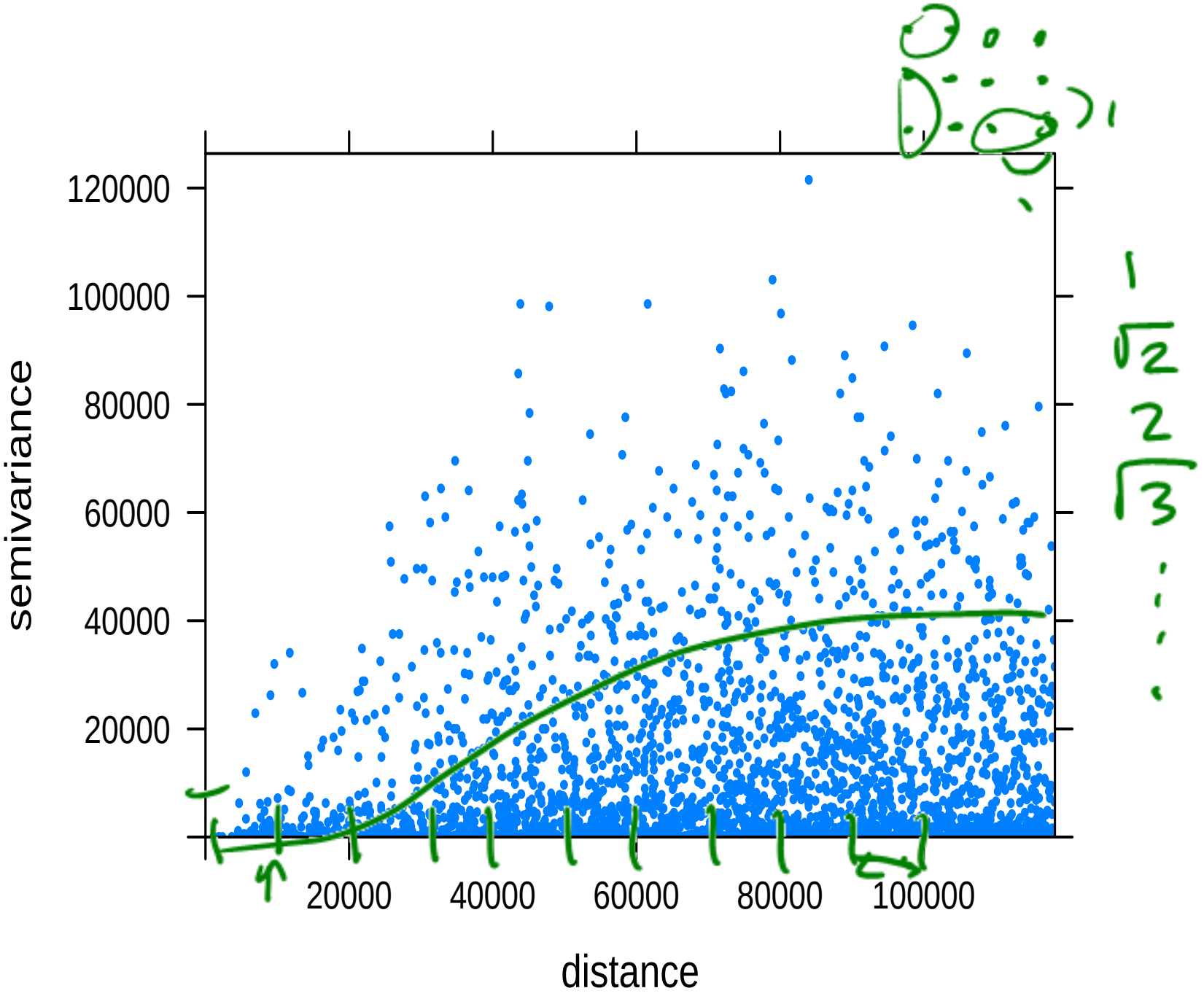
Semivariogram cloud

- Assuming 2nd order stationarity, so μ constant
- so can calculate $\left(Z(\boldsymbol{s}_i) - \bar{Z}\right)\left(Z(\boldsymbol{s}_j) - \bar{Z}\right)$
  - for each pair of obs.
  - Plot vs. distance
  - **Empirical covariogram cloud**
- Or, can calculate $\frac{1}{2}\left(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\right)^2$
  - **Empirical semivariance cloud**
- Notice that don't have to calculate $\bar{Z}$ to estimate the semivariance
- These are related: When σ^2 constant,

$$\frac{1}{2}\left(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\right)^2 = \sigma^2 - \left(Z(\boldsymbol{s}_i) - \bar{Z}\right)\left(Z(\boldsymbol{s}_j) - \bar{Z}\right)$$

- Example: covariogram and semivariogram clouds for the Swiss rain data


Cross−product
50000
0
−50000
0
50000
100000
200000
3000
Distance


semivariance
120000
100000
80000
60000
40000
20000
20000
40000
60000
80000
100000
distance
1
√2
2
√3

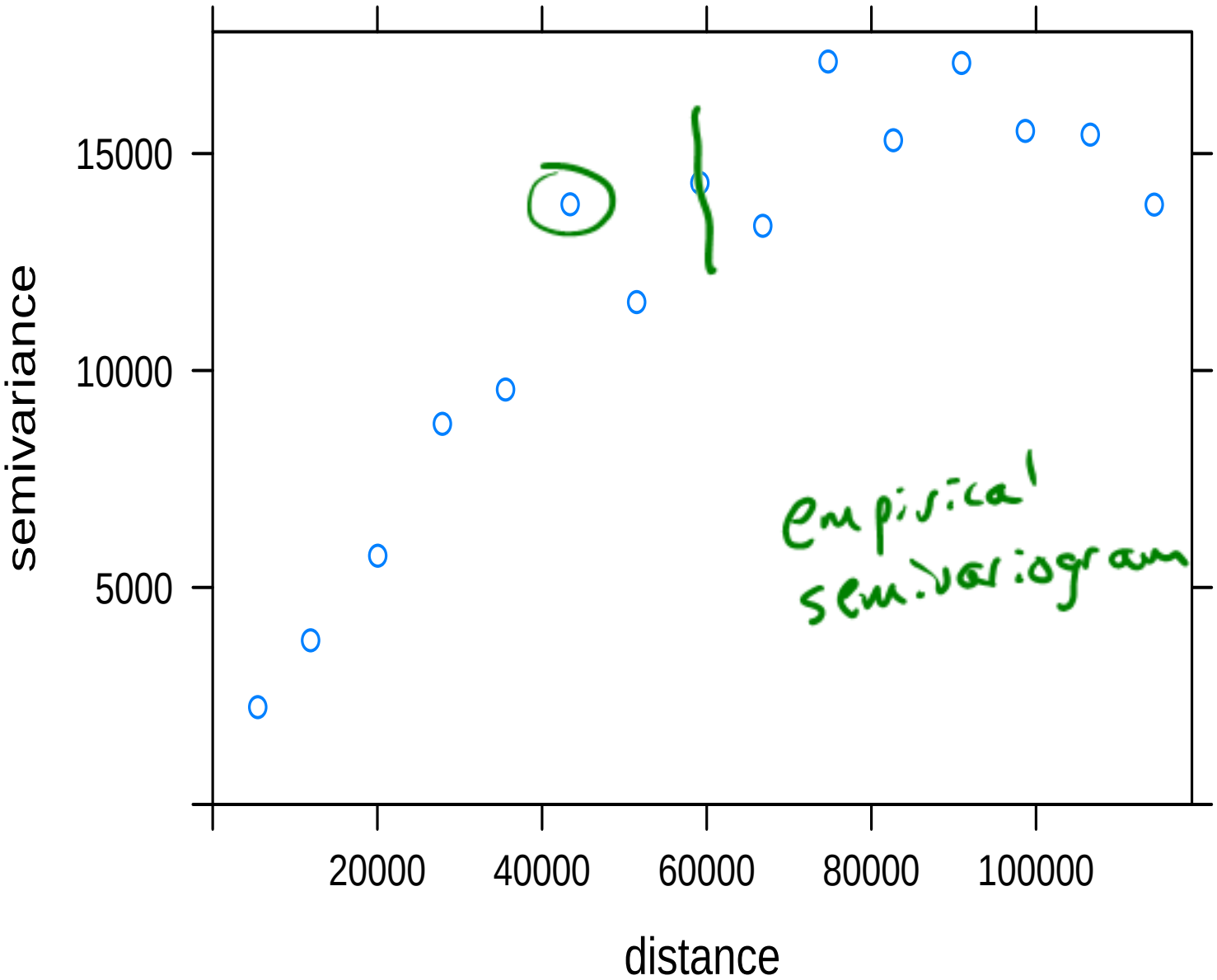
Semivariogram plot

- Smooth semivariogram cloud by averaging
- "Classical" or Matheron estimator

$$\hat{\gamma}(h) = \frac{1}{2\,N(h)} \Sigma_{N(h)} [Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)]^2$$

gamma semivariance at distance h

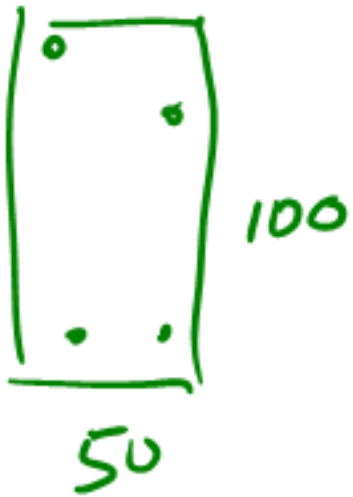
- Sum over pairs of points separated by distance h
- $N(h)$ is number of pairs
- Easy to do above on a grid.
- When locations are irregular, have to create "distance bins"
  - Define a range of distances = a bin, e.g. 0-25000m
  - Calculate mean distance and mean semivariance for the bin
  - Repeat for rest of bins
- Plot X = mean distance vs. Y= mean semivariance


semivariance
15000
10000
5000
20000
40000
60000
80000
100000
distance
empirical
sem.variogram

Choice of bin size/number

- How precise is the estimated semivariance?

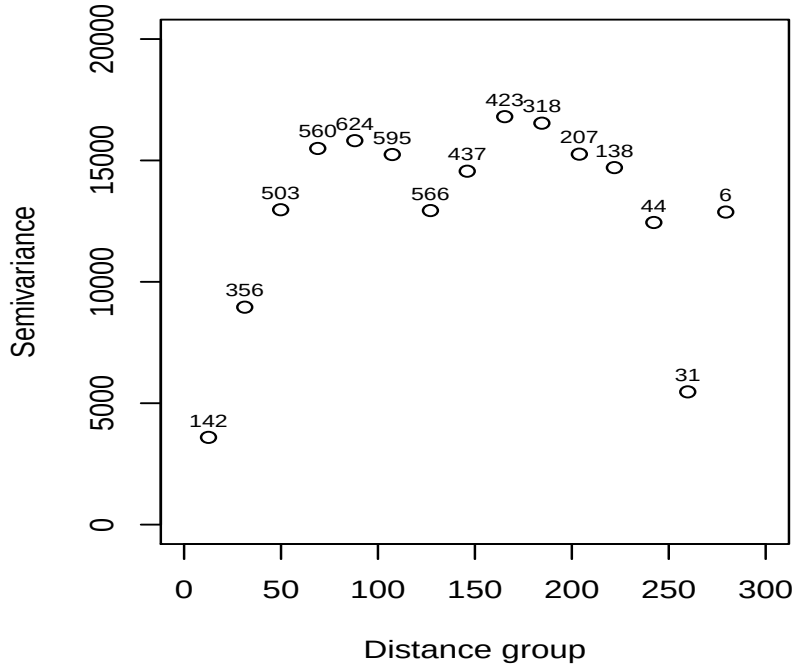
$$\text{Var } \hat{\gamma}(h) \approx \frac{2\gamma(h)^2}{N(h)}$$

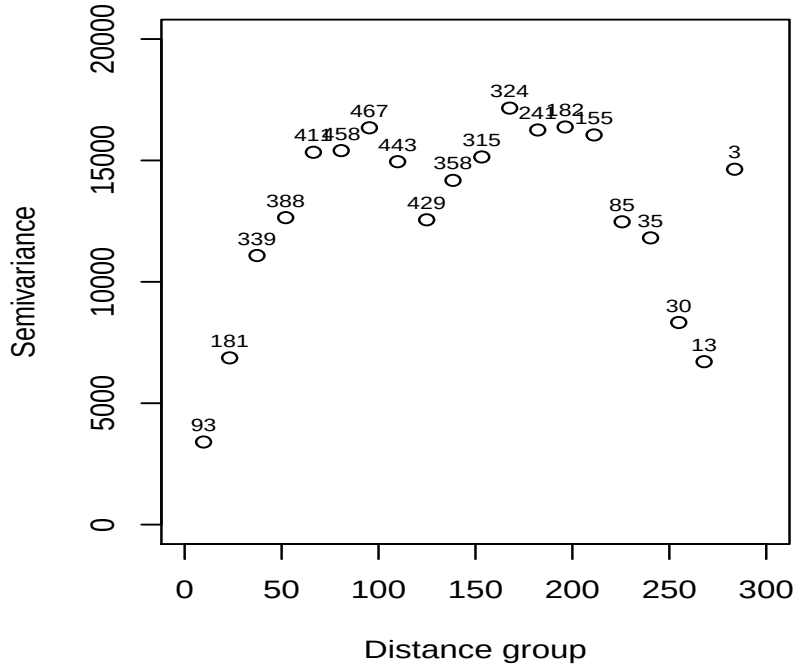


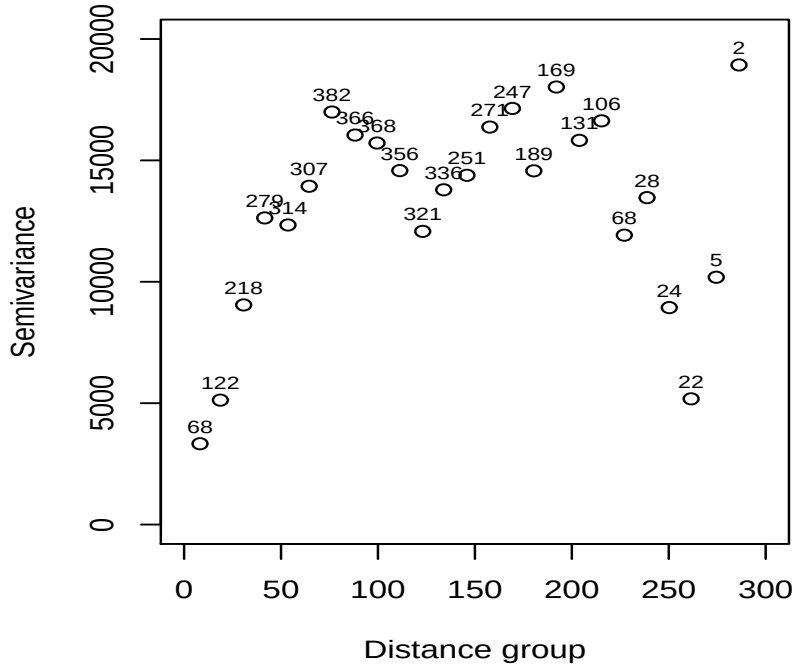
- How big should the bin be?
  - Want at least 30, preferably 50 or more, pairs per bin
- Choice of binning matters
  - very common to make all bins equally wide
  - but, $N(h)$ often small for large h
  - $\hat{\gamma}(h)$ not very precise, semivariogram is erratic
- Solution: don't calculate $\hat{\gamma}(h)$ for large distances
  - one recc.: calculate $\hat{\gamma}(h)$ to 1/2 max distance
  - Swiss rain: max distance is 291 km, so calculate to $h = 150$ km.
  - Notice default max distance in R is less
- Would like to have 10-15 bins, but # pairs more important

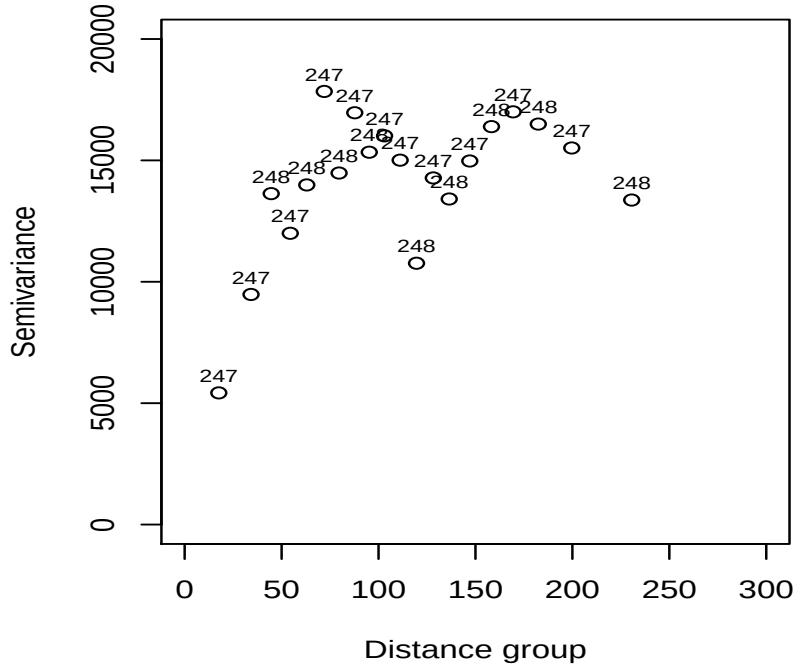
Choice of bin size/number

- $N(h)$ often small at short distances, when locations irregular
  - less of a problem because $\gamma(h)$ small, so Var $\hat{\gamma}(h)$ small
- Alternative is to have equal # pairs per bin
  - $\rightarrow$ wide bins for short distances and very long distances
  - concern is loss of information about $\hat{\gamma}(h)$ at small h
  - that info. is crucial for fitting models to semivariograms
- I tend to use equi-distant bins without large lag distances


Semivariance
Distance group
0
5000
10000
15000
20000
0
50
100
150
200
250
300
423
318
560
624
595
207
437
138
503
566
44
6
356
31
142


Semivariance
Distance group
0
5000
10000
15000
20000
0
50
100
150
200
250
300
324
467
241
182
155
414
458
443
315
358
3
388
429
85
35
339
30
181
13
93


Semivariance
Distance group
0
5000
10000
15000
20000
0
50
100
150
200
250
300
2
169
382
247
106
271
366
368
131
356
251
189
307
336
28
279
314
321
68
5
218
24
122
22
68


Semivariance
Distance group
0
5000
10000
15000
20000
0
50
100
150
200
250
300
247
247
247
248
248
247
247
247
248
247
248
248
247
248
247
248
247
247

Cressie-Hawkins estimator

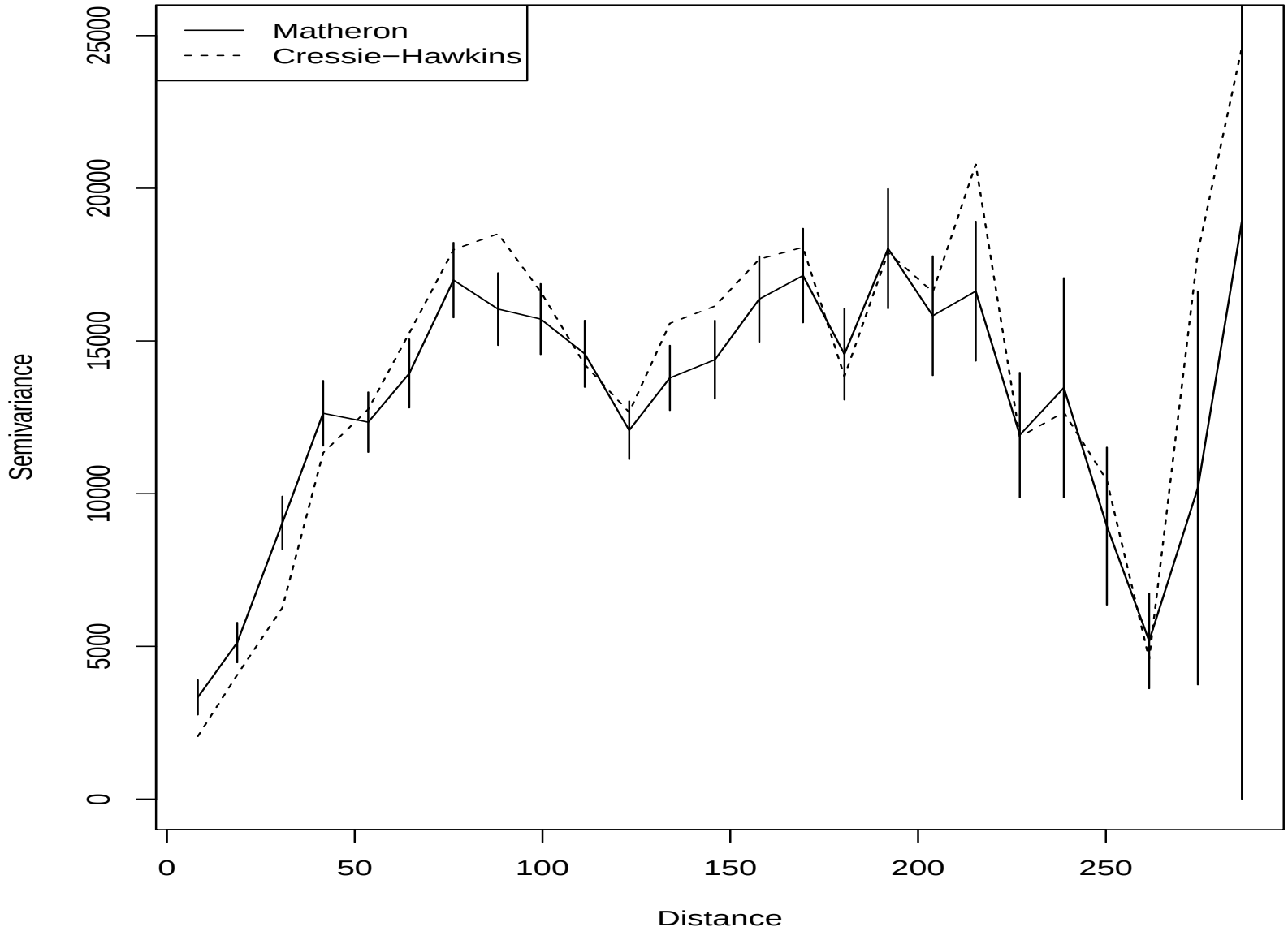
- Classical estimator is very sensitive to outliers
  - One unusual value can really mess up $\hat{\gamma}(h)$ because used in N-1 pairs.
- Cressie-Hawkins estimator is more robust to outliers

$$\hat{\gamma}(h)_{CH} = \frac{0.5\left[\Sigma_{N(h)} \mid Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j) \mid^{1/2}\right]^4}{0.457 + \frac{0.494}{N(h)} + \frac{0.045}{N(h)^2}}$$

- Where does this come from?
  - $\mid Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j) \mid^{1/2}$ is not dominated by a single large squared difference
  - That makes C-H more robust to outliers
  - When $Z(s) \sim N(0,1)$,

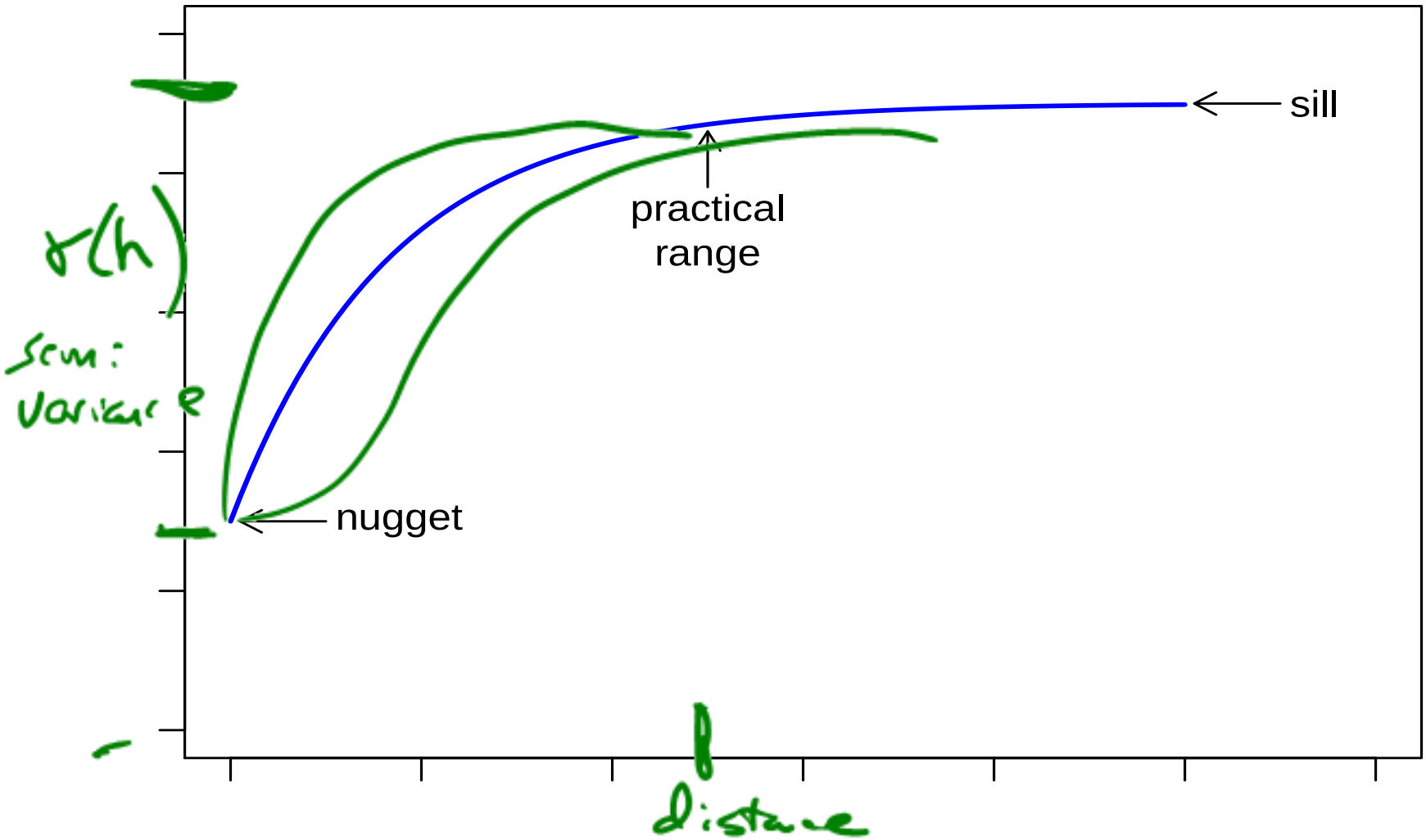
$$\mathrm{E}\ \frac{1}{N(h)}\left[\Sigma_{N(h)} \mid Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j) \mid^{1/2}\right]^4 \approx 2\gamma(h)\left[0.457 + \frac{0.494}{N(h)} + \frac{0.045}{N(h)^2}\right]$$

  - $\Rightarrow$ C-H estimator (also called robust SV estimator)
- With either estimator, plot of X=h vs. Y=$\hat{\gamma}(h)$ describes the pattern of spatial correlation
- Often (but not always) CH and Matheron estimates are similar


Matheron
Cressie−Hawkins
Semivariance
0
5000
10000
15000
20000
25000
0
50
100
150
200
250
Distance

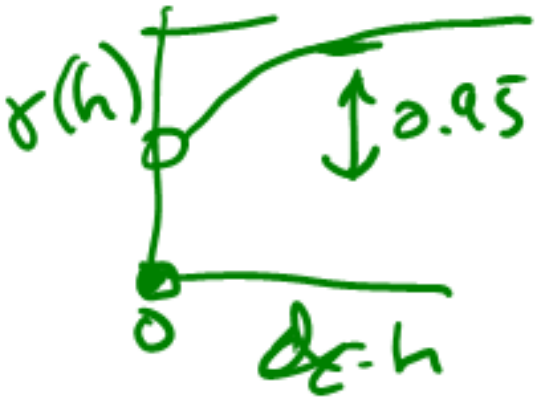
Semivariogram parameters

- Three important numbers that describe the semivariogram: nugget, sill, and practical range

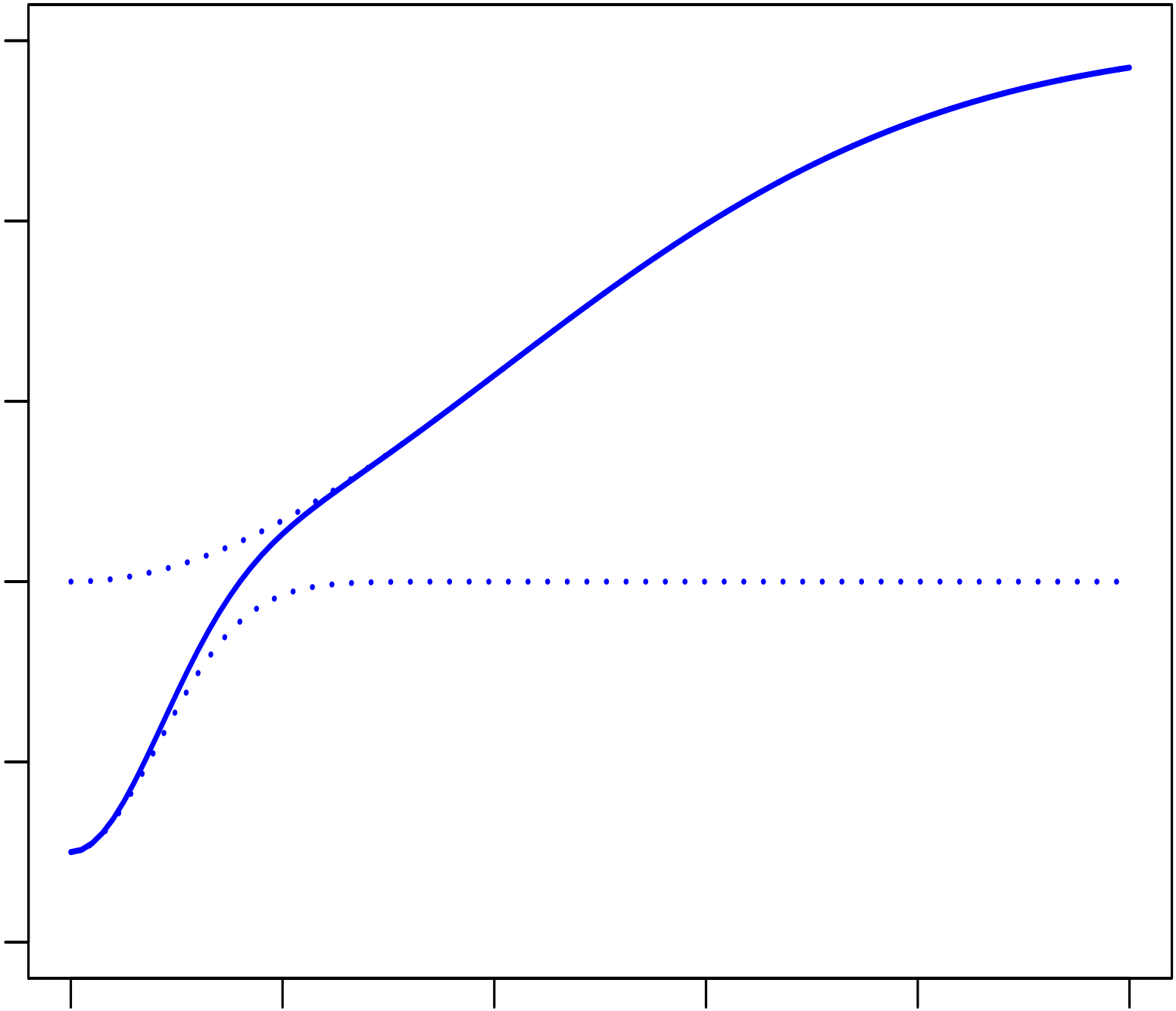


Semivariogram parameters

- Nugget: $\gamma(0)$, semivariance at 0 distance.
  - Replicate observations at a single location assumed to be identical (variance=0)
  - nugget = micro-scale variation
  - term from mining geology:
    nuggets of metal in ore $\rightarrow$ micro-scale variation



- Sill: $\gamma(h)$, for large h.
  - If σ^2 constant, sill $= \sigma^2$
  - Sill is an asymptote, technically, $\gamma(h)$ never equals sill.
- Practical range: where $\gamma(h) = 0.95(\text{sill} - \text{nugget}) + \text{nugget}$
  - obs. separated by more than the practical range are essentially uncorrelated.
  - 0.95 is traditional, other fractions sometimes used.
- All above assumes one spatial process.
- Can have multiple sills and practical ranges if multiple processes



Semivariogram models

- If field is second order stationary,
  - $\gamma(h) = \sigma^2 - \text{Cov}(h) = \text{Cov}(0) - \text{Cov}(h)$, so
  - $\text{Cov}(h) = \text{Cov}(0) - \gamma(h)$
- Now all need to do is estimate $\gamma(h)$ for any h

- A: Model $\gamma(h)$. We will focus on three parametric models:
- Spherical:

$$\gamma(h) = \sigma^2 \left[\frac{3}{2}\frac{h}{\alpha} - \frac{1}{2}\left(\frac{h}{\alpha}\right)^3 \right], \text{ for } h \leq \alpha$$

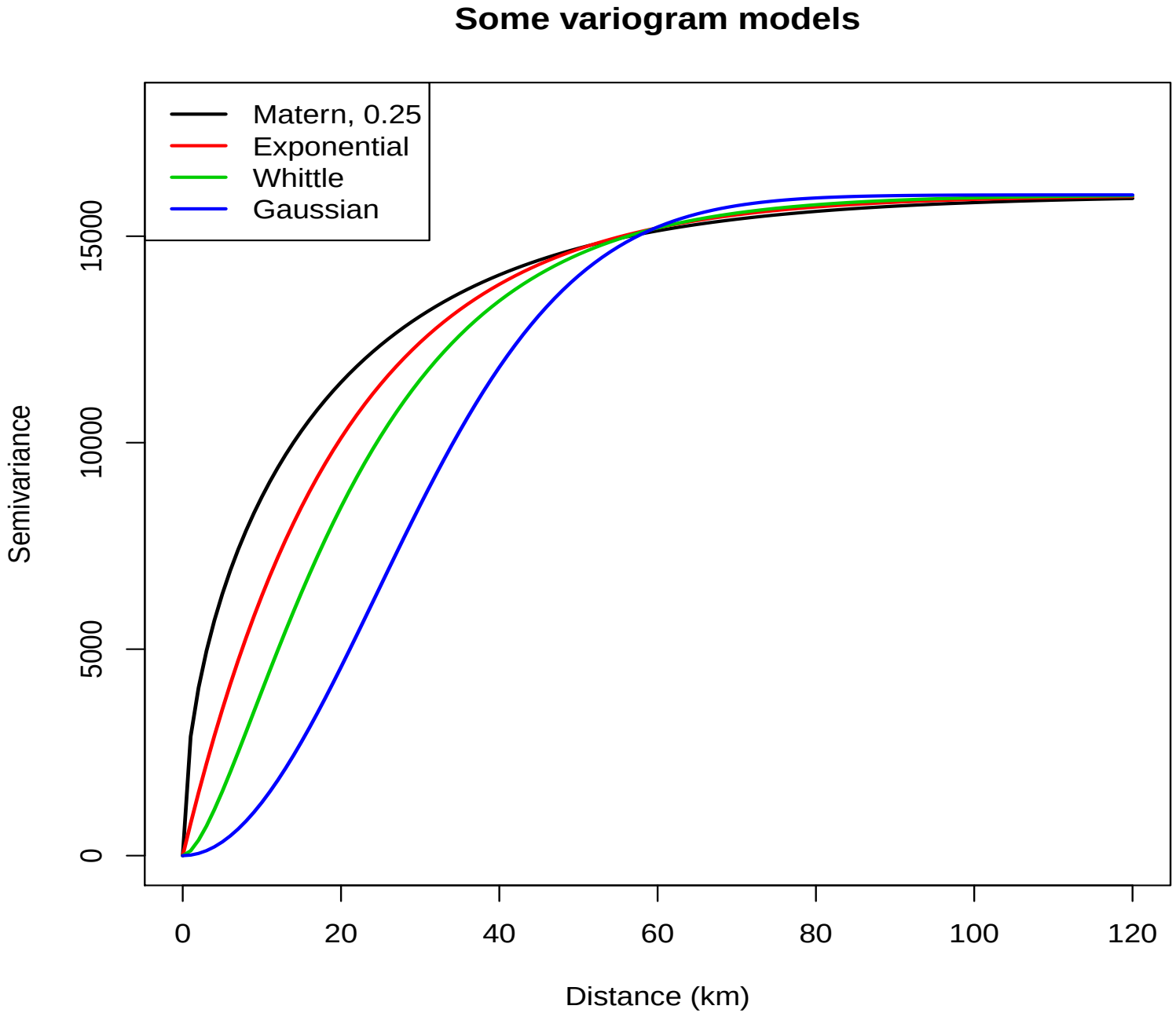
  - Cov exactly 0 for $h > \alpha$
- Exponential:

$$\gamma(h) = \sigma^2[1 - \exp(-3h/\alpha)]$$

- Gaussian:

$$\gamma(h) = \sigma^2[1 - \exp(-3\left(\frac{h}{\alpha}\right)^2)]$$

- In all, α is the practical range, σ^2 is the sill


Some variogram models
Matern, 0.25
Exponential
Whittle
Gaussian
Semivariance
0
5000
10000
15000
0
20
40
60
80
100
120
Distance (km)

Semivariogram models

- Exponential and Gaussian are two specific examples of models in the Matern class, a large family of semivariogram models

$$\gamma(h) = \sigma^2 \left[1 - \frac{1}{\Gamma(k)} \left(\frac{\theta h}{2} \right)^k 2K_k(\theta h) \right]$$

  - $\Gamma(k)$ and $K(\theta h)$ are math special functions (Gamma and modified Bessel fn of 2nd kind, order k)
  - θ controls the range, $= 1/\alpha$, σ^2 controls the sill
  - k controls the shape of the variogram
  - Gaussian is $k = 2$, exponential is $k = 0.5$.
- $k = 1$ is the Whittle model. I have found it quite useful

$$\gamma(h) = \sigma^2 \left[1 - \frac{h}{\alpha} K_1 \left(\frac{h}{\alpha} \right) \right]$$

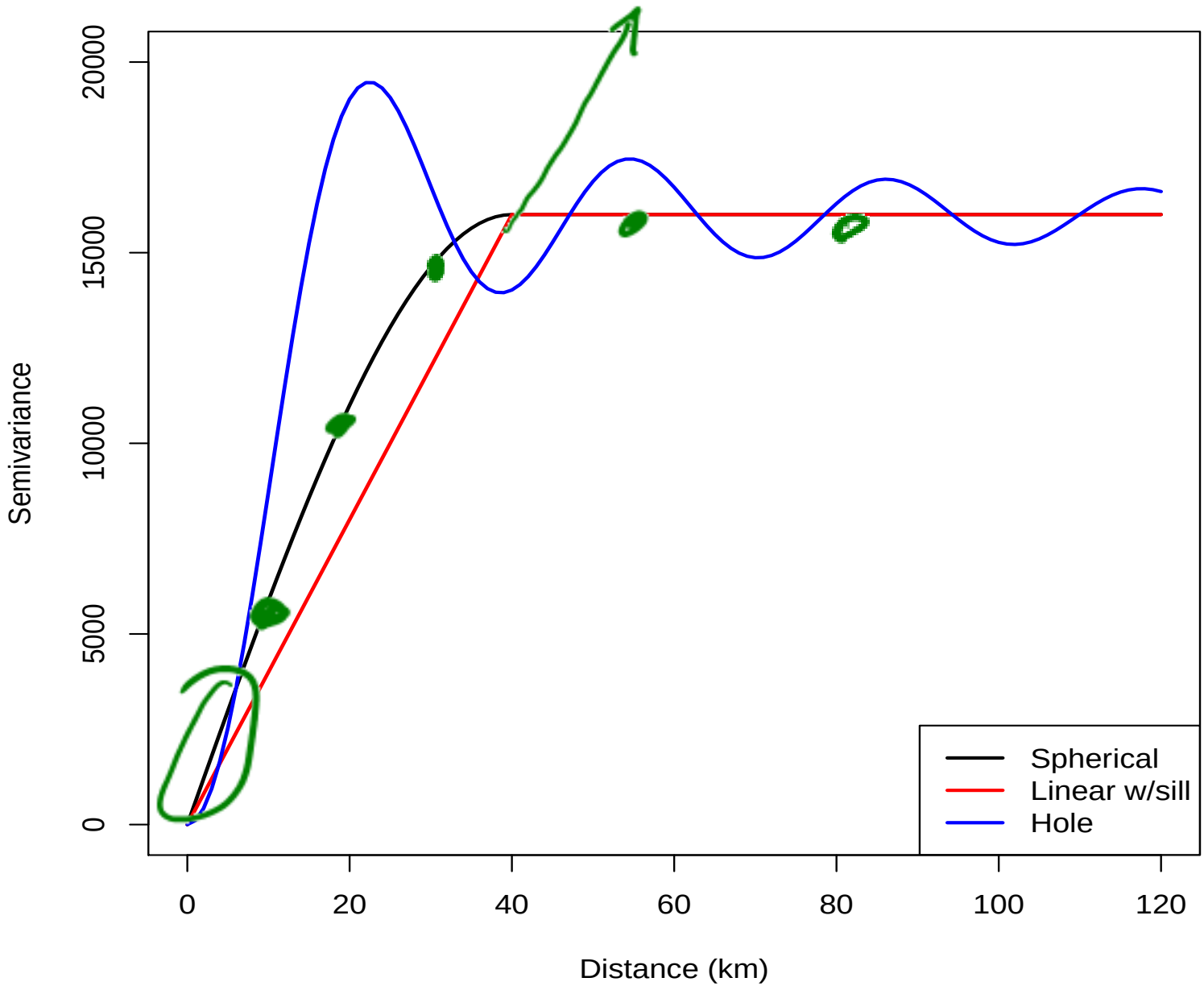
- The Gaussian is one of the historical, traditional models
  - Current opinion is to avoid it, corr ≈ 1 for nearby locations.
  - Wackernagel, 2003. The Gaussian model is "pathological".

More variogram models

- linear: $\gamma(h) = \theta h$
  - no sill!, unless you force one.
  - Not second order stationary, $C(0) = \sigma^2$ doesn't exist unless you force a sill
  - But is intrinsic stationary:
    Var $(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))$ is constant for any distance h
  - So semivariogram is well defined
  - My experience is that a linear trend semivariogram usually indicates trend across the study area
  - I suggest modeling the trend, then computing a semivariogram from the residuals from that trend
- wave or hole-effect:

$$\gamma(h) = \sigma^2 \left[1 - \left(\frac{h}{\alpha}\right) \sin\left(\frac{h}{\alpha}\right)\right]$$

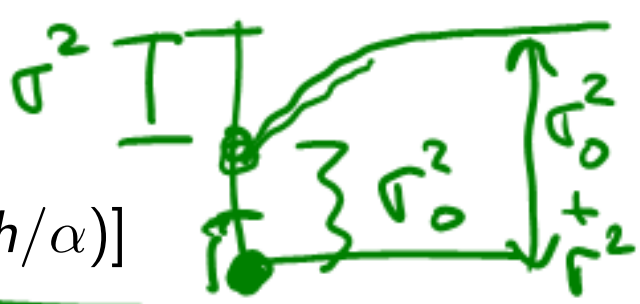
  - models periodic spatial patterns
  - picture on next slide


More Semivariogram models
Semivariance
0
5000
10000
15000
20000
0
20
40
60
80
100
120
Distance (km)
Spherical
Linear w/sill
Hole

Nuggets

- Can add to any model
- e.g. Exponential with nugget

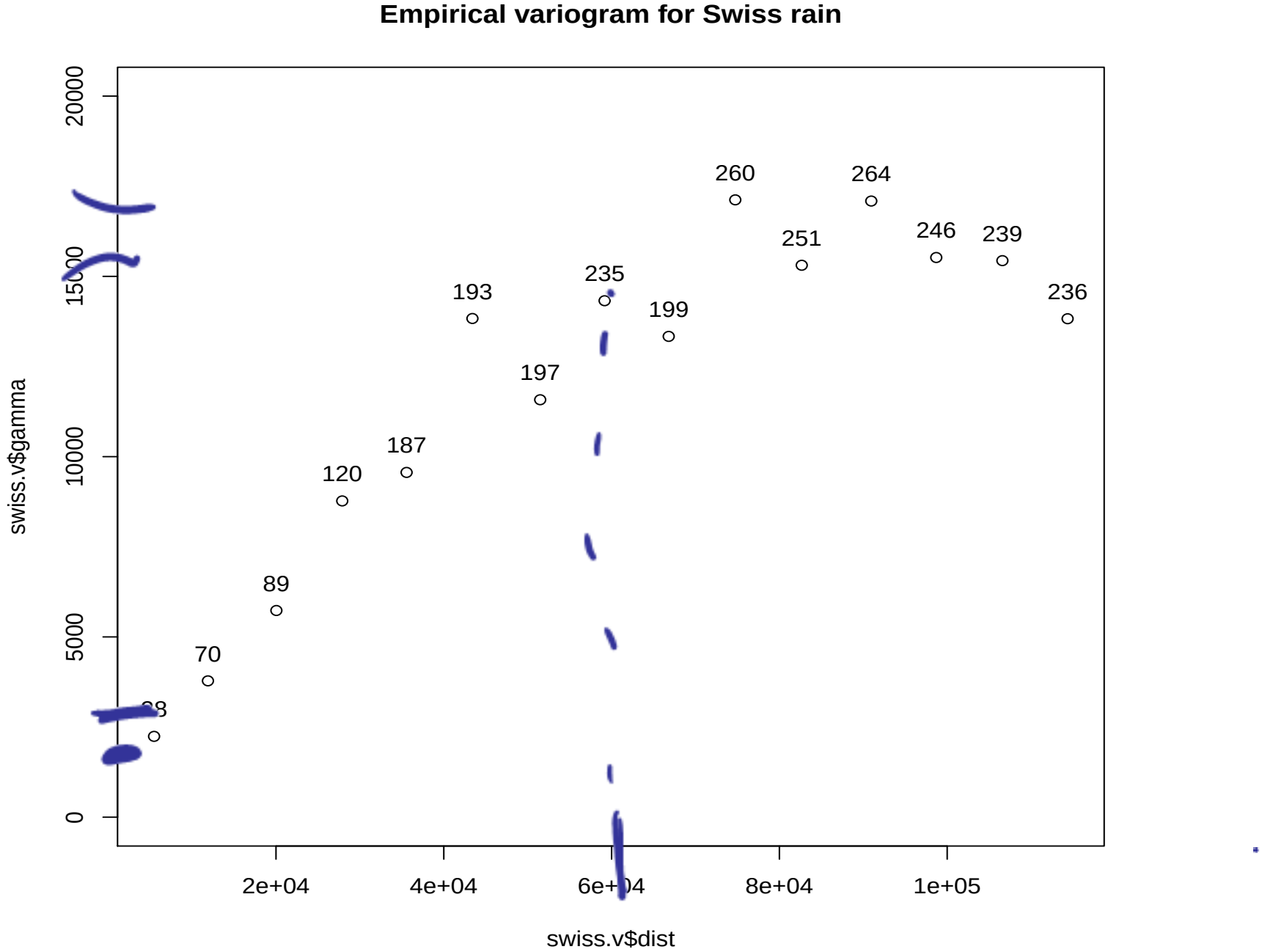
$$\gamma(h) = \sigma_0^2 + \sigma^2 \left[1 - \exp(-3h/\alpha)\right]$$



  - σ^2 now called partial sill, is the spatially-associated variation
  - σ_0^2 is the micro-scale variation
- In practice, very few pairs of observations are really close to each other, i.e. with $h \approx 0$
  - So, the estimated nugget is an extrapolation down to $h = 0$
  - nugget can be very sensitive to the choice of semivariogram model.
- Even if there is a nugget, $\gamma(0)$ is still defined as 0
  - The Kriging prediction at an obs. location is the obs. value
  - But, $\gamma(\epsilon) =$ nugget, where ϵ is a very small number
  - Which means that the prediction of $Z(s)$ can be quite different a very small distance away from that observed location
- Will talk later about "measurement error Kriging" which is not a perfect predictor at observed locations
- And redefines what "nugget" means.

Estimating a SV model

- Given an empirical average SV or SV cloud and a model, how do you estimate the SV model parameters (e.g. σ^2, α, and perhaps σ_0^2)?
- All SV models are non-linear functions of their parameters
- Not $X_{1i}\beta_1 + X_{2i}\beta_2$
- Still use least-squares
  - Could fit to cloud, but tradition is to fit to binned averages
  - Given $\hat{\gamma}(h_i, \theta)$ for a set of bins, a specific model, and parameters θ
  - Find $\theta = (\sigma^2, \alpha, \sigma_0^2)$ for which $\Sigma\left[\gamma(h_i) - \hat{\gamma}(h_i, \theta)\right]^2$ is as small as possible
  - Same criterion as linear regression
- No closed-form solution - need iterative algorithm
  - must provide starting values
  - bad start $\rightarrow$ big trouble
  - generally well-behaved if start is reasonable
- I (and everyone else) uses eyeball guess as starting values

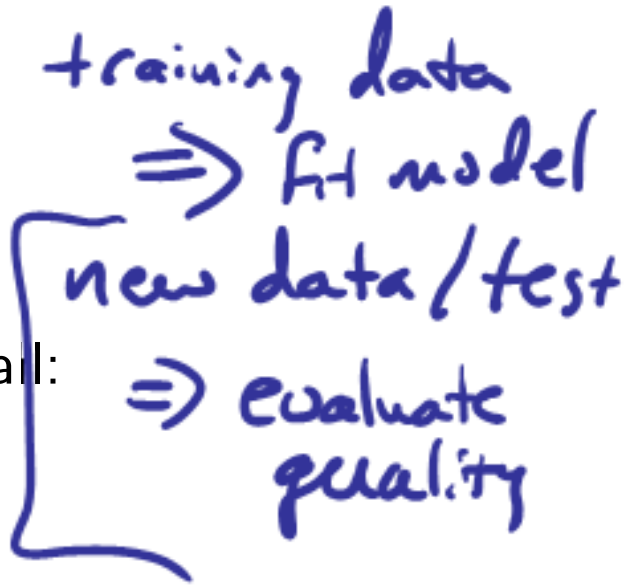

Empirical variogram for Swiss rain
swiss.v$gamma
swiss.v$dist
0
5000
10000
15000
20000
2e+04
4e+04
6e+04
8e+04
1e+05
260
264
251
246
239
236
235
193
199
197
187
120
89
70

Fitting SV models

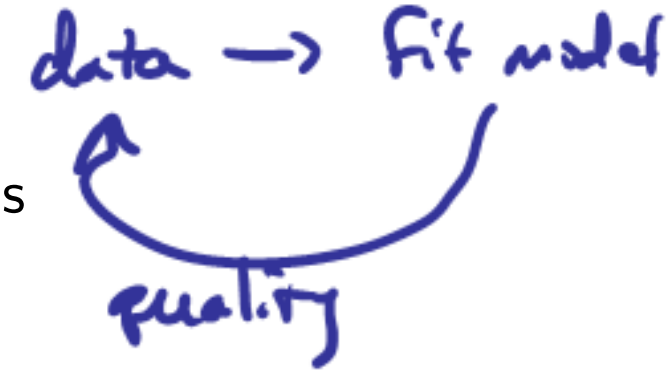
- One issue:
  - LS assumes all obs equally important
  - But, if Var $Y_2 >$ Var Y_1, want to give more attention to fitting Y_1 more closely
  - Remember Var $(\hat{\gamma}(h) \approx 2\frac{\gamma(h)^2}{\#pts}$
    - more pts $\rightarrow\downarrow$ Var
    - $\uparrow \hat{\gamma}(h) \rightarrow\uparrow$ Var
- Used weighted LS to put more emphasis on obs with lower variance
- Minimize $\Sigma w_i \left[\gamma(h_i, \theta) - \gamma(h_i)\right]^2$
  - statistically optimal weights are $w_i = 1/\text{Var } \gamma(h_i)$
  - Problem: Var depends on $\gamma(h)$, which is what we are trying to estimate!
  - Solution (not the only one): use $w_i = \frac{\#pts}{h^2}$ as the weight
  - assumes $\gamma(h) \approx h$

Which SV model to use?

- Which model to use? Two approaches:
- 1) which model best fits semivariogram
- use wt SS as measure of fit. For Swiss rainfall:
  - Exponential: SS = 1.608
  - Spherical: SS=1.184
  - Gaussian: SS=1.258
  - Matern, k=0.25: won't fit
  - Matern, k=5: 1.153
  - Matern, k=4: 1.140
  - Suggests Spherical or Matern with k=4
- 2) which model gives most accurate predictions?



Which SV model to use?



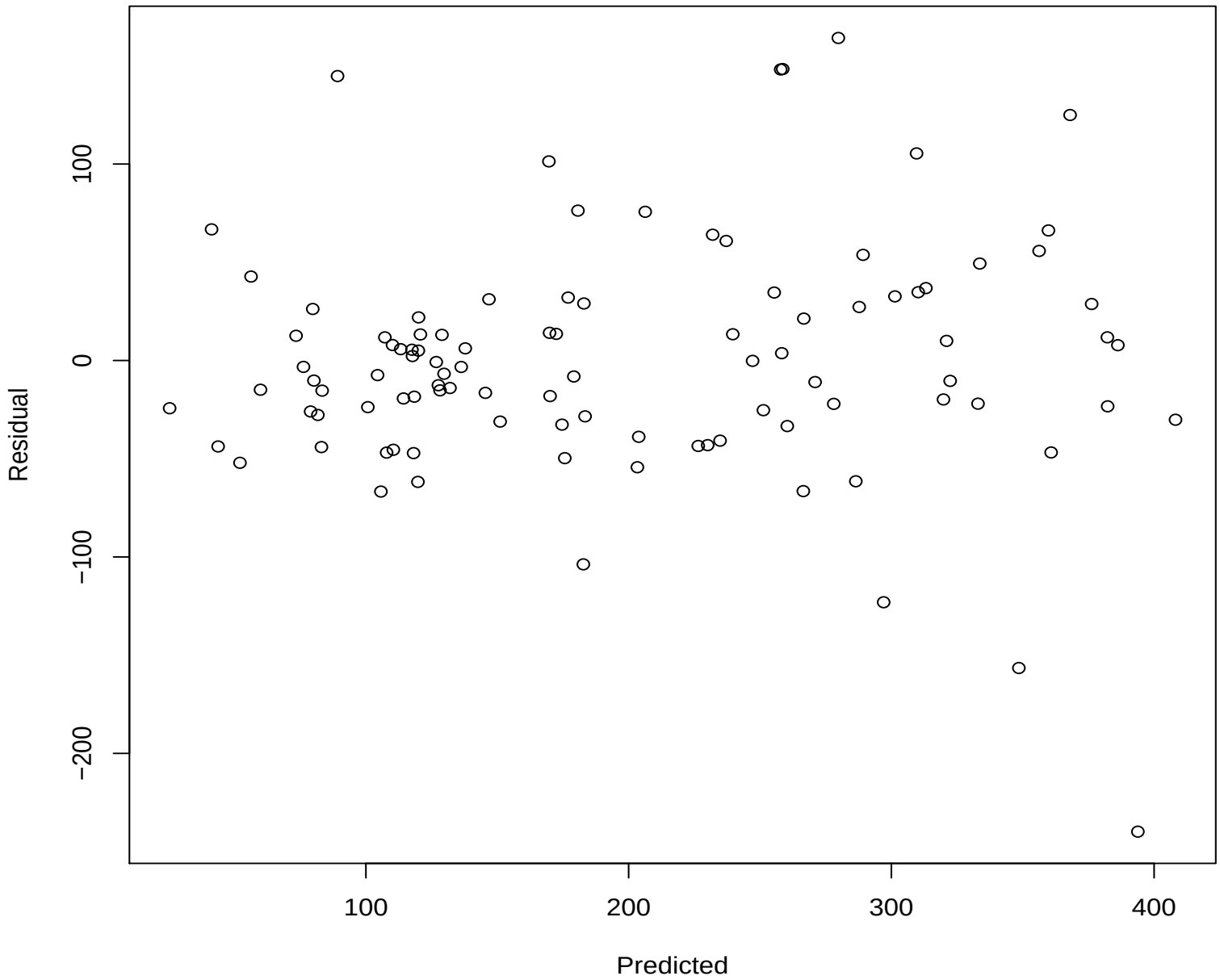
- One small problem with assessing predictions
- Using data twice:
  - Once to fit SV model
  - Again, to assess precision of predictions
- known to be overly optimistic and favor models with more parameters
- leads to overfitting data at hand
- Three solutions, all often used:
- 1) Training/test set
  - fit model to subset of data (training set)
  - evaluate on remaining data (test set)
  - requires arbitrary division into training and test
  - both are subsets of full data set

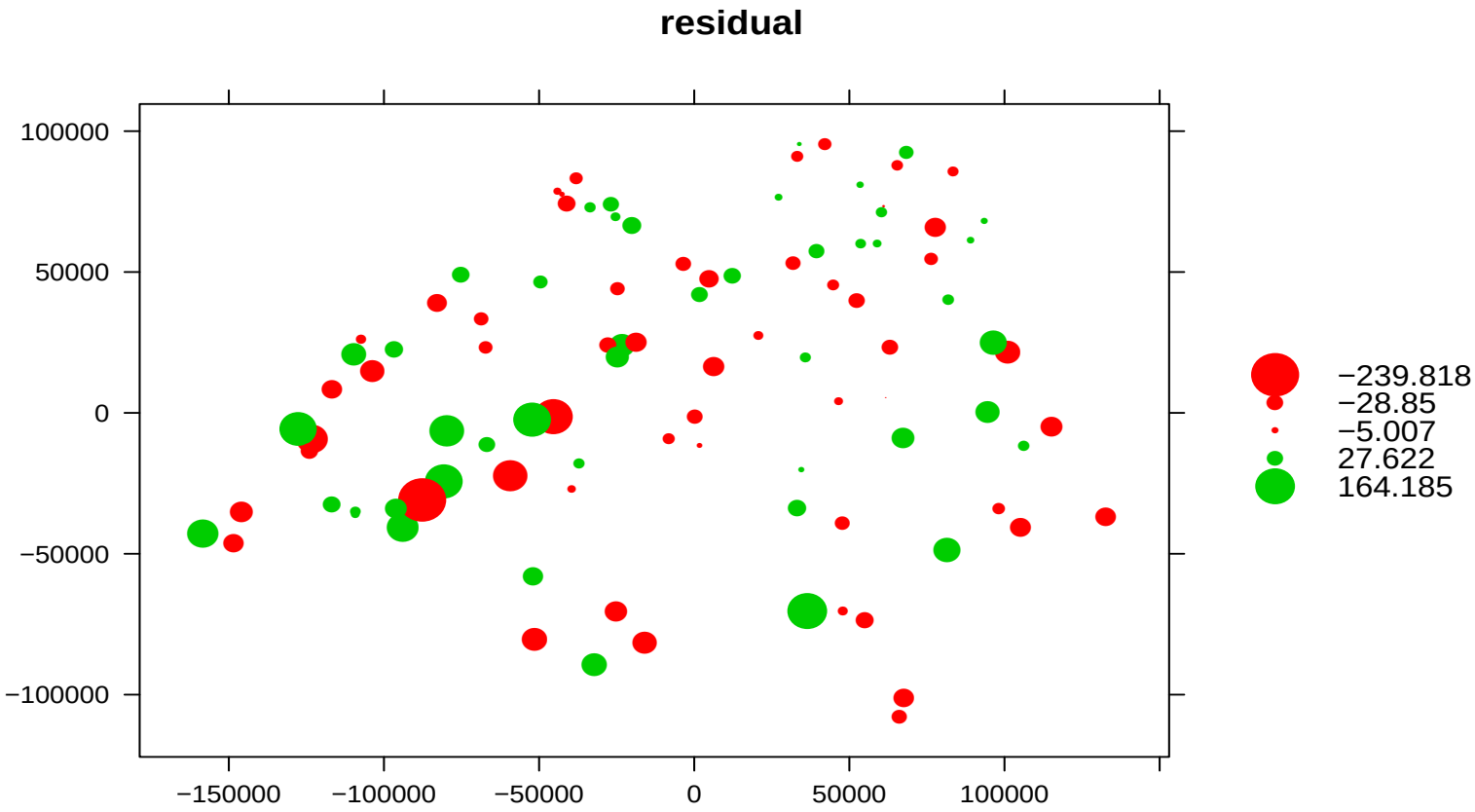
Cross-validation

- 2) cross-validation
  - remove 1st obs, using $N-1$ obs. to fit SV and predict $Z(\boldsymbol{s}_1)$
  - calculate squared error for 1st obs: $\left[Z(\boldsymbol{s}_1) - \hat{Z}(\boldsymbol{s}_1)\right]^2$
  - $\hat{Z}(\boldsymbol{s}_1)$ NOT based on $Z(\boldsymbol{s}_1)$, so valid "out-of-sample" prediction
  - return 1st obs. to data set, remove 2nd obs., repeat above
  - repeat for all obs., average squared errors
  - gives mean squared error of prediction: $\frac{1}{N}\left[(Z(\boldsymbol{s}_i) - \hat{Z}(\boldsymbol{s}_i)\right]^2$
  - exact same concept as PRESS statistic in linear regression
- 3) k-fold CV
  - If many points, leave-one-out can be slow
  - Divide data into k "folds"=sets of obs., k=5 and k=10 are common
  - remove entire set of obs., fit model on 90% (for k=10) of data
  - predicte omitted points, calculate rMSEP
  - repeat for all other parts

Cross-validation

- For Swiss rainfall data:
  - Exponential: MSEP = 3615
  - Spherical: MSEP = 3447
  - Gaussian: MSEP = 3725
  - Matern, k=5: 3610
  - Matern, k=4: 3591
- Suggests Spherical (one of two with good fit)
- But, none substantially better than any other
- CV also provides a way to assess overall quality of the model
  - plot predicted values vs residuals - should be flat sausage, just like for linear regression
  - spatial (or bubble) plot of residuals - should be no big clusters
    plots on next page
  - Compare range of residuals to range of obs. values
    obs. values: 0 to 493, residuals: -240 to 160
    more variability than probably would like

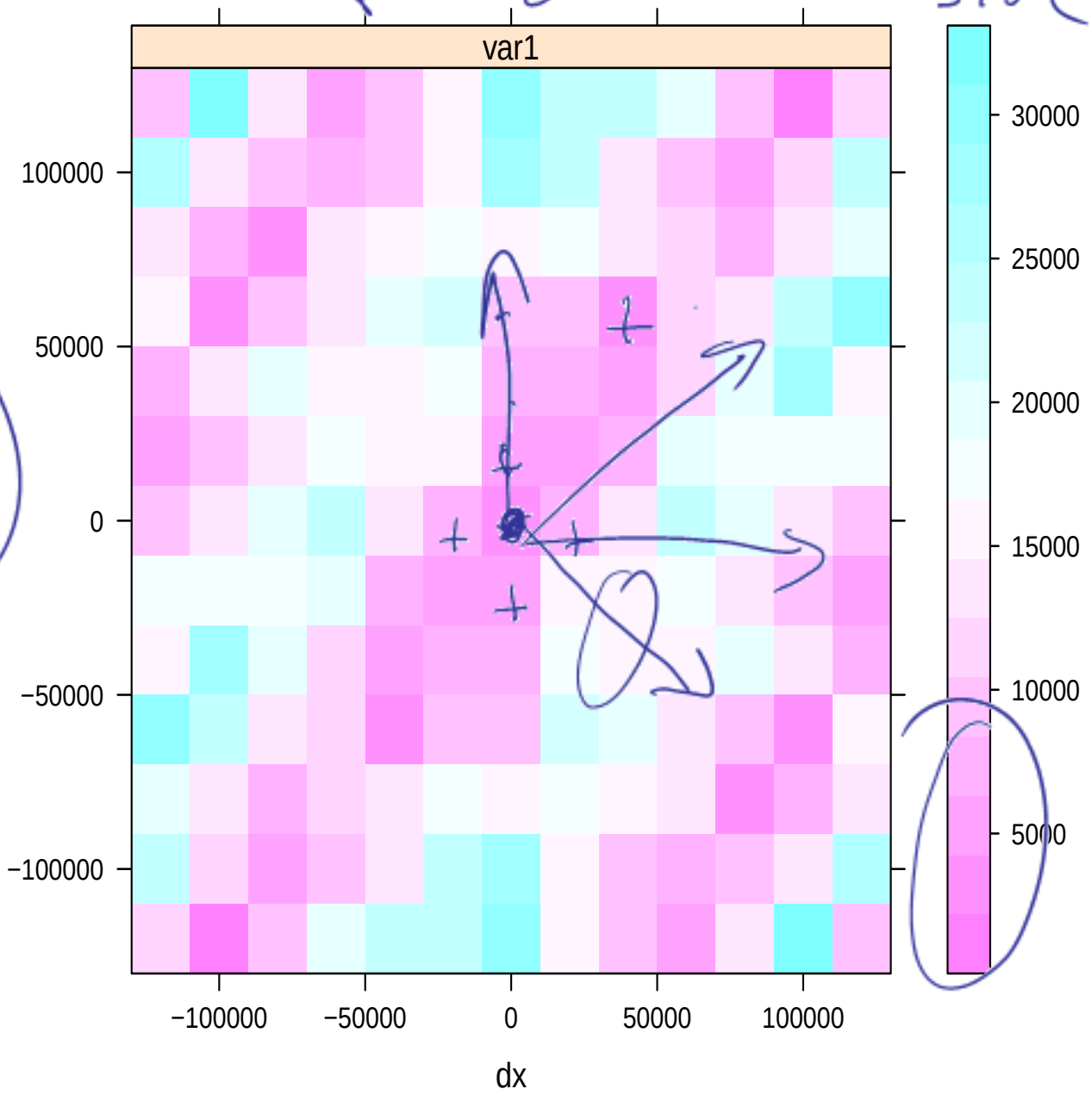

Residual
100
0
−100
−200
100
200
300
400
Predicted


residual
100000
50000
0
−50000
−100000
−150000
−100000
−50000
0
50000
100000
−239.818
−28.85
−5.007
27.622
164.185

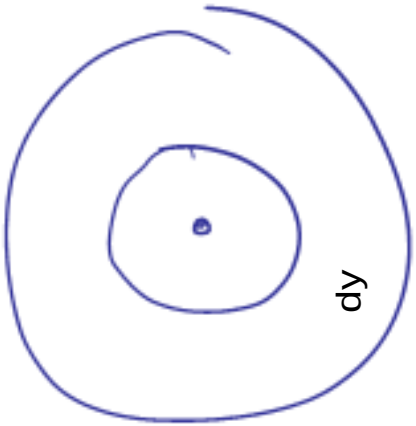
Other ways to fit SV models

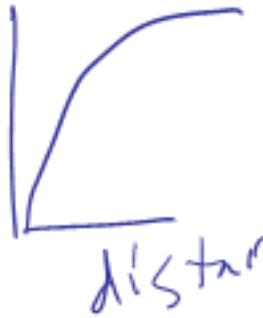
- REML (REstricted/REsidual Maximum Likelihood)
  - likelihood-based, after removing fixed effects
  - avoids binning
  - most commonly used with linear models for trend / treatment effects
- More advanced methods
  - Composite likelihood and generalized estimating equations
  - not commonly used
- Final point about Swiss rainfall data
  - seems not isotropic (plot of 'variogram map' on next page)

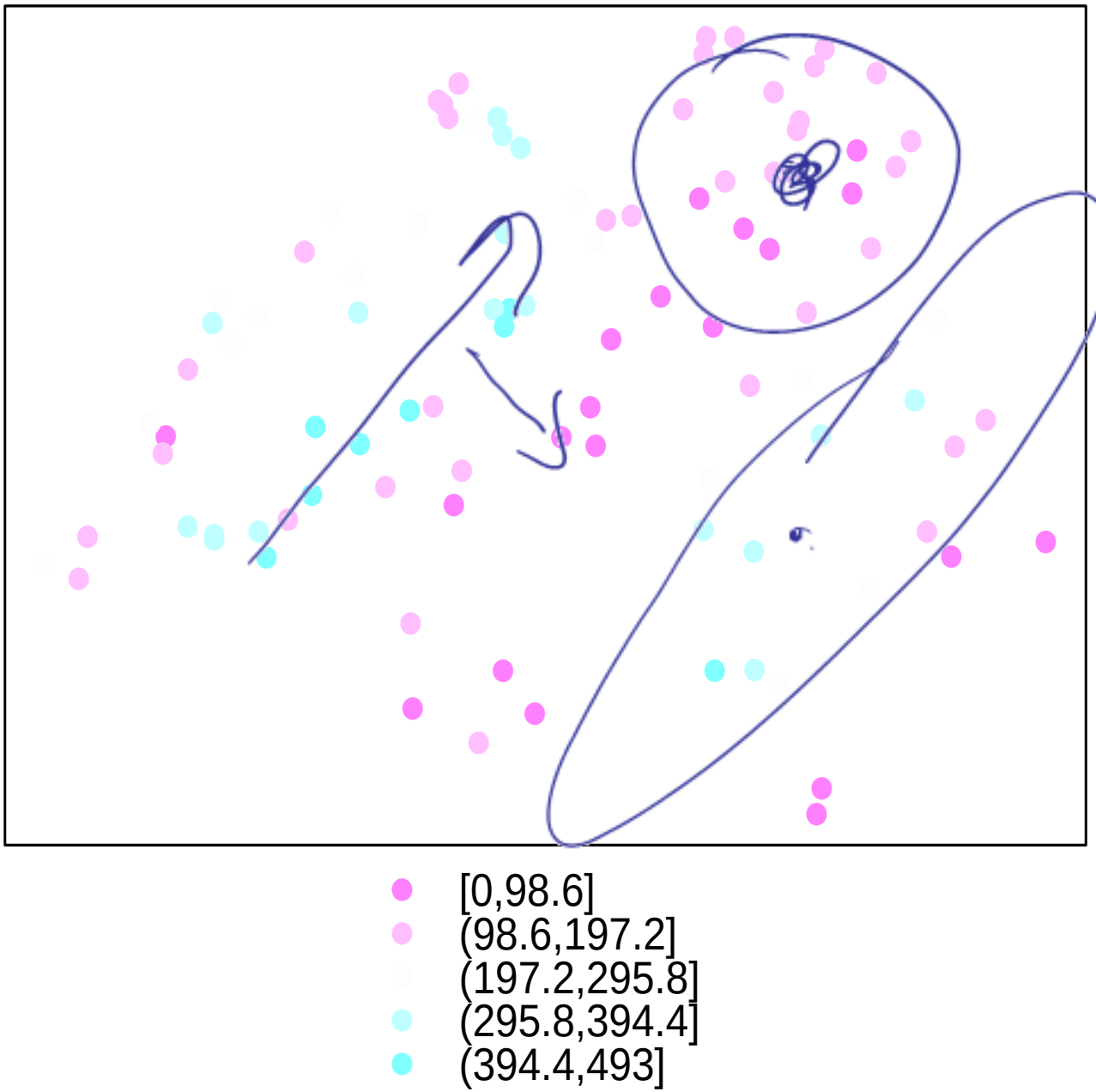
Variogram map: data intensive



isotropic



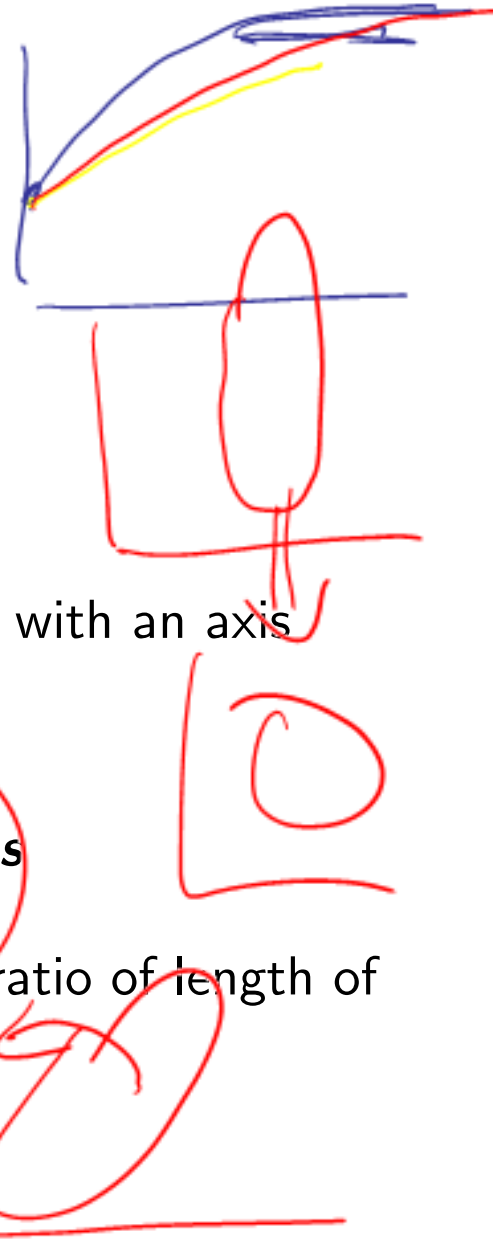



[0,98.6]
(98.6,197.2]
(197.2,295.8]
(295.8,394.4]
(394.4,493]

Local prediction

- Currently, using all observations to make predictions
- Only have to compute $\mathbf{\Sigma}^{-1}$ once
- What if you have a large number of observations (e.g. 1 million) $\mathbf{\Sigma}$ is too large
- Or, believe μ changes over space (and you don't want to model that change)
- use only nearby obs. to predict at a location
- This is called "local prediction"
- Either use some max. # obs., or all obs. within some specified distance of prediction location.

Anisotropy



- We've talked about visual assessment of anisotropy
  - Variogram maps and directional variograms
- How to deal with different types of anisotropy
  - Geometric anisotropy:
  - Range longer in one direction than another
  - Nugget and (partial) sill same in all directions
  - Rotate coordinate system so longer direction aligned with an axis
  - Then rescale that axis
  - Do this by:

$$\boldsymbol{s}^* = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \boldsymbol{s}$$

  - Where θ is the angle of the major axis and λ is the ratio of length of minor to length of major axis

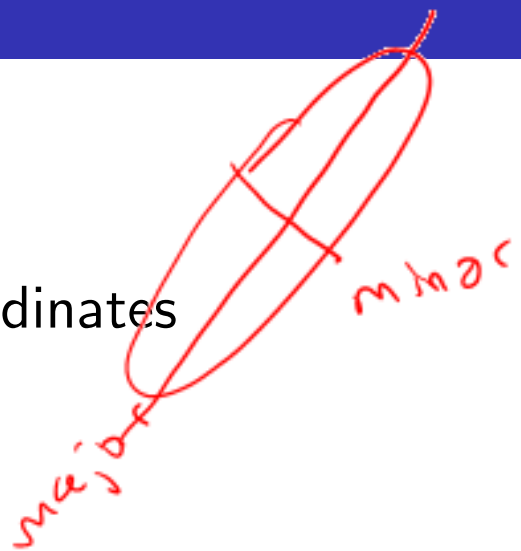
Zonal anisotropy

- variogram sills vary with direction
- The direction with the shorter range also has the shorter sill.
- model by a combining an isotropic model and a model which depends "only on the lag-distance in the direction θ of the greater sill" (Schabenberger and Gotway, 2005, p 152).

$$\gamma(\boldsymbol{h}) = \gamma_1(|| \boldsymbol{h} ||) + \gamma_2(h_\theta)$$

- Chiles and Delfiner (1999, p. 96) warn against axis-specific models, e.g.: $\gamma(\boldsymbol{h}) = \gamma_1(h_x) + \gamma_2(h_y)$
  - Under certain circumstances they can lead to Var $Z(s) = 0$, which not good.

Kriging with anisotropy



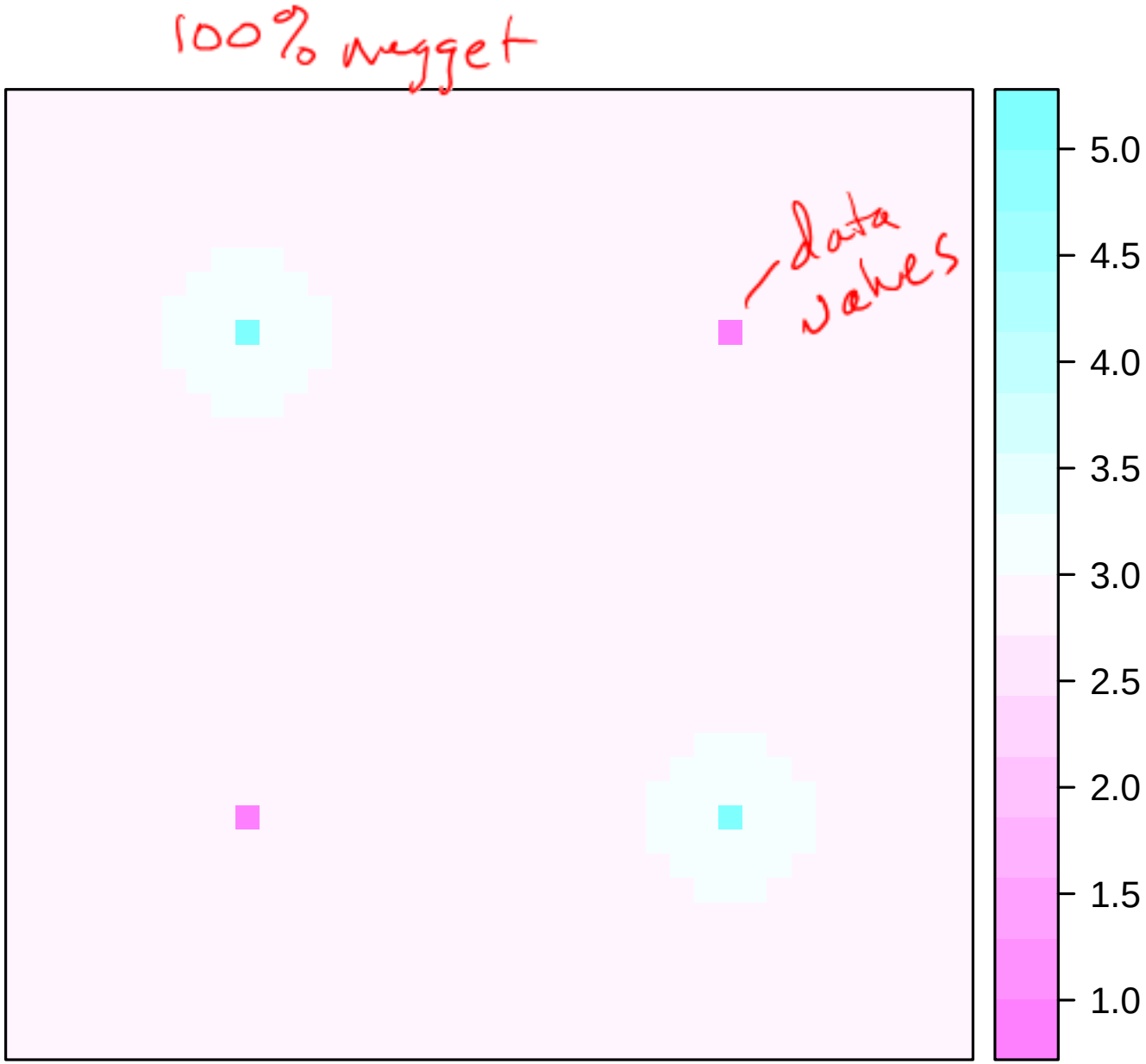
- Geometric anisotropy
- Do not have to transform and back transform coordinates
- krige() can do this for you
- in the vgm() specification of the variogram model,
  - add anis=c(θ, λ)
  - θ is the direction of the longest range (highest correlation) in degrees, measured clockwise from the + vertical axis (N)
  - λ is the ratio of minor range to major range (a value from 0 to 1)
  - so anis=c(60,0.2) specifies a major axis at 2 o'clock with a length 5 times the minor axis.
- Zonal anisotropy
  - "fake it" by setting up a geometrically anisotropic component with a major axis much longer than the minor axis (e.g. $\lambda = 0.00001$.
  - Only distances only along the major axis contribute to $\gamma(\boldsymbol{h})$

Measurement error kriging

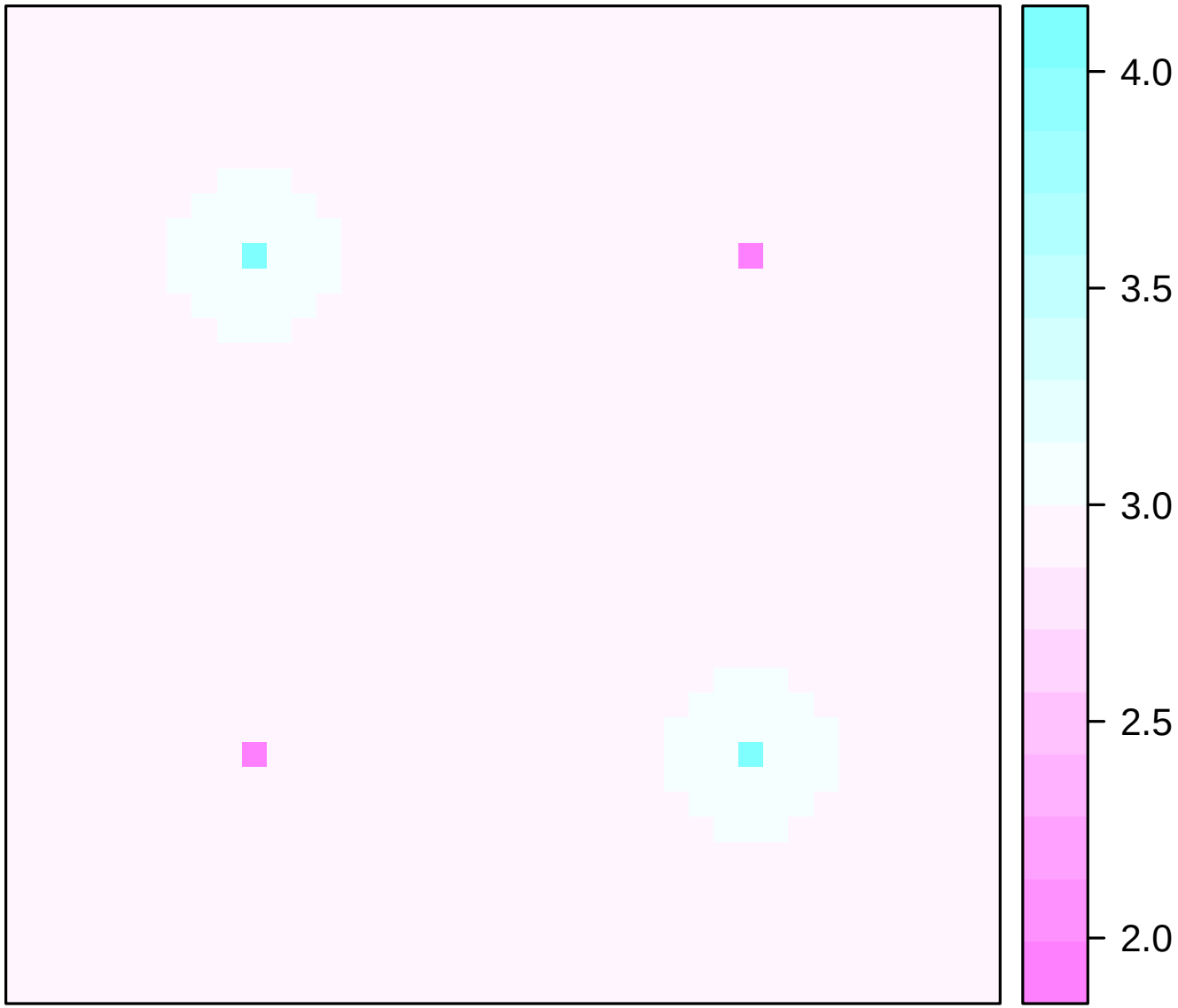
- Remember the effect of the nugget:
  - At the location of an observed value, $\hat{Z}(s) = Z(s)$.
  - and Var $\hat{Z}(s) = 0$ at that location
  - but any small distance away from that s, Var $Z(s + h) = \sigma^2_{nugget}$
- Kriging "honors the data"
- Assumes that a hypothetical repeat observation at s will be exactly the same number
- What if there is measurement error in $Z(s)$?
  - So, a repeat measurement at same location will not be the same value.
- Now, do not want to honor the data (because what we observe includes non-repeatable measurement error)
- Identification problem: have only one obs. per location. Can not separate nugget from measurement error
- Need outside information / guess about the magnitude of the measurement error

Measurement error kriging

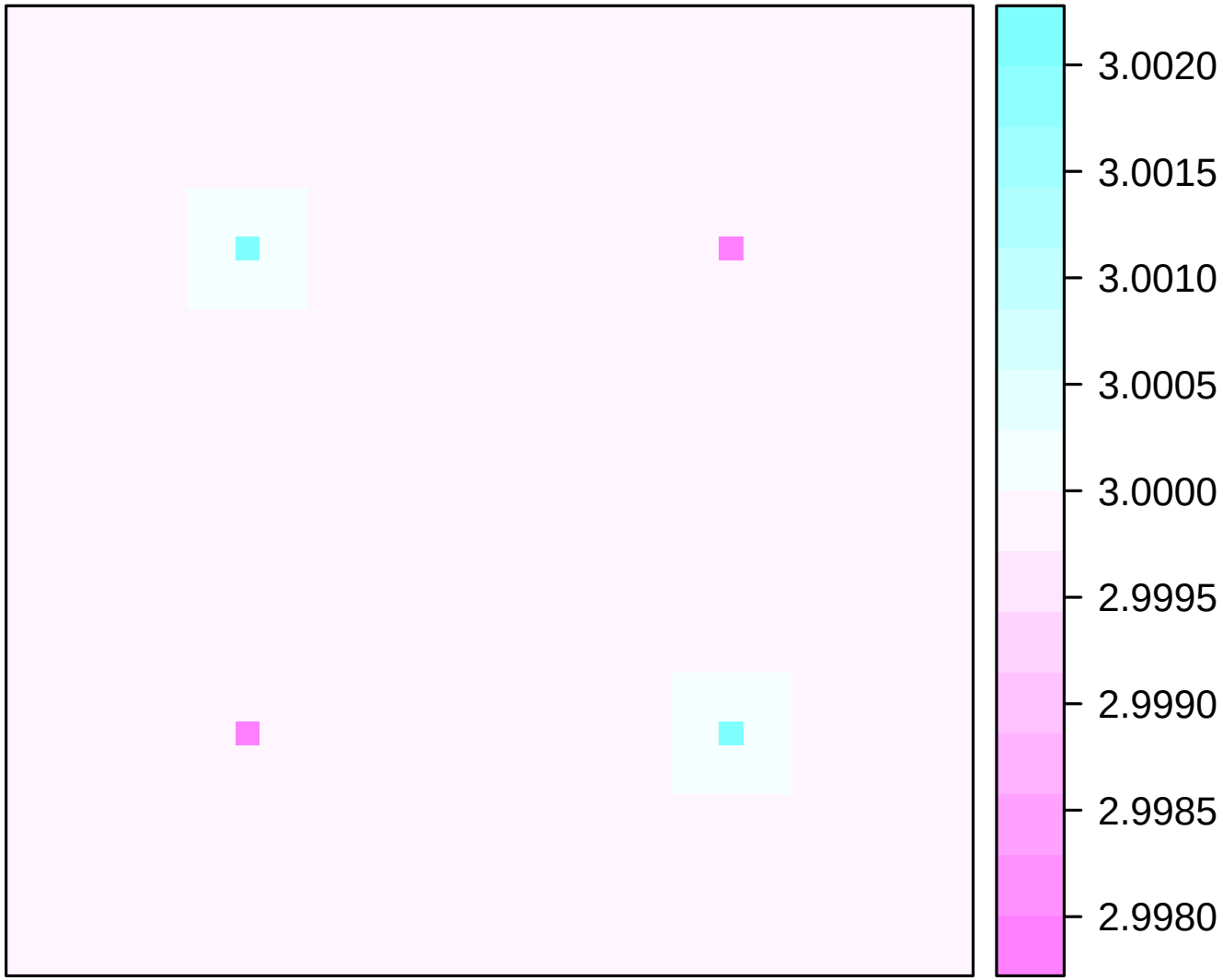
- If you "know" $\sigma^2_{meas.\ error}$, then can account for this when kriging
- Consequence is that prediction at observed locations is a "smoothed" version of the observations.
- Simplest example of the difference is a pure nugget process
  no spatial correlation at the spatial scale of the observations
- If 100% nugget, prediction is the mean value **except** at the observed values
- If 50% nugget, 50% meas. error, prediction at observed values pulled towards the mean
- If 100% meas. error, prediction is almost the mean everywhere
- Pictures on next three slides


100% nugget
data values
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0

50% nugget.



100% me. 0% nugget



Measurement error kriging

variance.

- Specify a known measurement error by adding Err=value to vgm model
- Must specify specific value. Can not (to my knowledge) estimate it
  - Data can not separate nugget (micro-scale variation) and meas. error
  - unless there are repetitions at the same location.
  - 2 obs with same loc. are a problem for most software
  - Correlation = 1, so VC matrix no longer full rank

Block Kriging

Fairfield-Smith variance law

- To now, we've focused on predicting $\boldsymbol{Z(s)}$ individual locations
  - locations assumed to be points w/o area
  - (physical/mathematical simplification, not reality)
- In many applications, want to predict total over some area
  - total and mean interconvertable: total = mean*area
- Areas are not undividable units
  - experiment on people. A person is a clearly defined, undividable unit
  - experiment in a field. You choose the plot size - no clearly defined undividable unit
- Called the Modifiable Areal Unit Problem (MAUP)
  - both size and shape of area matter
  - inferences depend on both
  - e.g. Variance between "replicate" field plots depends on size and shape

Block Kriging

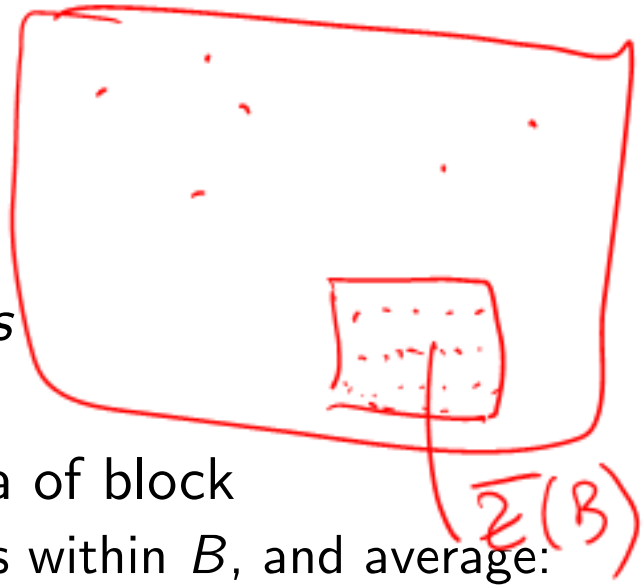
- MAUP is one example of a "change of support" problem
- Support: size, shape, and orientation of a unit associated with a measurement
- Changing support, e.g. by averaging or aggregating,
  - creates new random variables (for the new plots)
  - related to original r.v's, but spatial and statistical properties are different
  - e.g. semivariogram parameters will change

Block Kriging

- Block Kriging is a second ex. of a COSP
- predict $Z(B)$ for block B with area $|B|$:

$$Z(B) = \frac{1}{|B|}\int_B Z(s)\,ds$$

area total

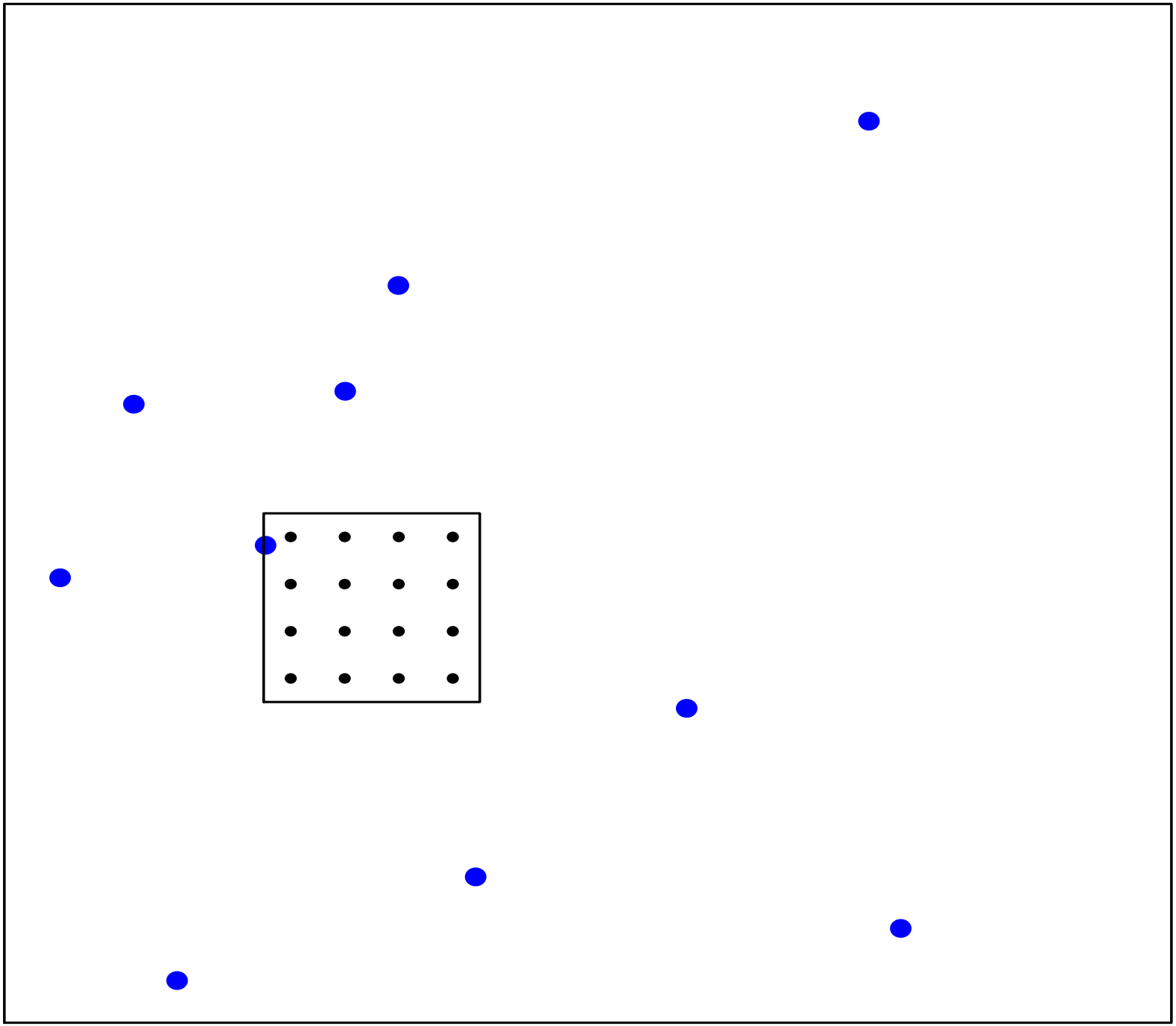


- Z(B) is an average, so total in the block / area of block
  - In practice, predict at a grid of point locations within B, and average: $Z(B) = \Sigma\lambda_i Z(\boldsymbol{s}_i)$
  - choose λ_i to minimize MSEP
  - like OK, but based on "point-to-block" covariances

$$\text{Cov}\,(Z(B), Z(\boldsymbol{s}_i)) = \frac{1}{|B|}\int_B Cov(Z(u), Z(\boldsymbol{s}_i))\,du$$

  - Again, approx. by setting up a grad of pts in B:

$$\text{Cov}\,(Z(B), Z(\boldsymbol{s}_i)) = \frac{1}{N}\Sigma_j Cov(Z(u_j), Z(\boldsymbol{s}_i))\,du$$



Non-linear Kriging

- reminder: Best predictor (smallest MSEP) is E $[Z(\boldsymbol{s}_0) \mid \boldsymbol{Z(s)}]$
  - Gaussian: linear function of observations: $\sum \lambda_i Z(\boldsymbol{s}_i)$
  - Other distributions: still want E $[Z(\boldsymbol{s}_0) \mid \boldsymbol{Z(s)}]$, but that won't be a linear fn of obs.
- Log Normal kriging
  - $\log Z(\boldsymbol{s}_i) \sim$Gaussian, use log-transformed obs., predict $P(\boldsymbol{s}_0) = \widehat{\log Z}(\boldsymbol{s}_0)$
  - then back transform: $\hat{Z}(\boldsymbol{s}_0) = \exp P(\boldsymbol{s}_0)$
  - $P(\boldsymbol{s}_0)$ is an unbiased prediction of $\log Z(\boldsymbol{s}_0)$, but $\exp P(\boldsymbol{s}_0)$ is a biased predictor of $Z(\boldsymbol{s}_0)$.
  - Jensen's inequality, demonstrated by HW 1 question

LogN kriging: Solutions

- 1) Ignore the problem, focus on medians
  - $\exp P(\boldsymbol{s}_0)$ is an asymptotically unbiased estimate of the **median** of $Z(\boldsymbol{s}_0)$.
- 2) Use properties of logN distribution
  - when $\log Z(\boldsymbol{s}_0) \sim N(\mu, \sigma^2)$, E $Z = \exp(\mu + \sigma^2/2$, E $\hat{Z}(\boldsymbol{s}_0) = \exp(\mu +$ Var $\hat{Z}(\boldsymbol{s}_0)/2$
  - so predict $\log Z(\boldsymbol{s}_0)$, calculate $\sigma^2(\boldsymbol{s}_0) =$ Var $\hat{Z}(\boldsymbol{s}_0)$, estimate σ^2 (e.g. by sill)
  - predictor of $Z(\boldsymbol{s}_0)$ is

$$P(\boldsymbol{s}_0) = \exp\left[\widehat{\log Z}(\boldsymbol{s}_0) + \sigma^2/2 - \sigma^2(\boldsymbol{s}_0)/2\right]$$

Trans-Gaussian Kriging

- log is one member of the Box-Cox family of transformations. For these

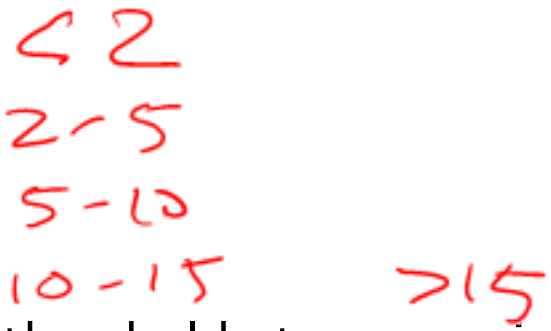
$$Z^*(\boldsymbol{s}_0) = \begin{cases} \frac{Z(\boldsymbol{s}_0)^\lambda - 1}{\lambda} & \lambda > 0 \\ \log Z(\boldsymbol{s}_0) & \lambda = 0 \end{cases}$$

  - $\lambda = 1 \Rightarrow$ no transformation
  - $\lambda = 0.5 \Rightarrow$ proportional to $\sqrt{Z(\boldsymbol{s}_0)}$ transformation
  - $\lambda = -1 \Rightarrow$ proportional to $1/Z(\boldsymbol{s}_0)$ transformation
  - purpose of $\frac{X-1}{\lambda}$ is so that limit as $\lambda \to 0$ is log transformation
- Since kriging minimizes MSEP, want / prefer symmetric distribution of values
- if including covariates (UK), want / prefer symmetric distribution of residuals

Indicator Kriging

- What if you want to predict an exceedance probability, $P[Z(\boldsymbol{s}_0) >$ threshold $]$
- e.g. legal limit on concentration of mercury in fish
- Kriging predicts E $[Z(\boldsymbol{s}_0) \mid \boldsymbol{Z(s)}]$
- Define $Z^*(\boldsymbol{s}_0) = I(Z(\boldsymbol{s}_0) >$ threshold $)$
  - $Z^*(\boldsymbol{s}_0) = 1$ if condition is true $(Z(\boldsymbol{s}_0) >$ threshold $)$
  - $Z^*(\boldsymbol{s}_0) = 0$ if condition is false $(Z(\boldsymbol{s}_0) \leq$ threshold $)$
  - E $Z^*(\boldsymbol{s}_0) = P[Z(\boldsymbol{s}_0) >$ threshold $]$
- Apply indicator transformation to all obs, estimate semivariogram from indicator variables (can be hard) then krige.
- Issues:
  - no guarantee that $0 \geq p \geq 1$
    remember, $\hat{Z}(\boldsymbol{s}_0)$ can exceed range of data
  - variety of ad-hoc fixes
  - there are more complicated methods
    my general sense is they don't work markedly better

Indicator Kriging



- Can use many thresholds to approximate the cdf of $Z(\boldsymbol{s}_0)$
  - Define $Z_1(s)$ to be $I(Z(s) < k_1)$,
  - and $Z_2(s)$ to be $I(Z(s) < k_2)$,
  - for many values of k
  - gives you predictions of $\hat{F}(k_1)$, $\hat{F}(k_2)$, ...
- Note: E $Z^*(\boldsymbol{s}_0) \mid Z^*(s)$ is not the same as E $Z^*(\boldsymbol{s}_0) \mid Z(s)$ because the indicator transformation "throws away" information.

Disjunctive Kriging

- Extend indicator kriging from 2 regions (e.g. $Z(\boldsymbol{s}_0) < 10$ or $Z(\boldsymbol{s}_0) \geq 10$) to many
- e.g. $Z(\boldsymbol{s}_0) < 10$, $10 \geq Z(\boldsymbol{s}_0) < 20$, $20 \geq Z(\boldsymbol{s}_0) < 30$, ...
- Knowing $10 \geq Z(\boldsymbol{s}_0) < 20$ is more informative than knowing only that $Z(\boldsymbol{s}_0) < 20$ (IK)
- very elegant math (which we'll ignore)
- $\rightarrow$ predict any function $g(Z(\boldsymbol{s}_0))$, including the set of indicator functions
- Richard Webster has published a lot of ag-related studies using disjunctive kriging
- Only available in R in the Rgeostats package
- Note: can combine block kriging ideas with any of the non-linear krigers
  - e.g. define 1km x 1km areas and estimate $P[\overline{\text{soil N}} > \textit{threshold}]$