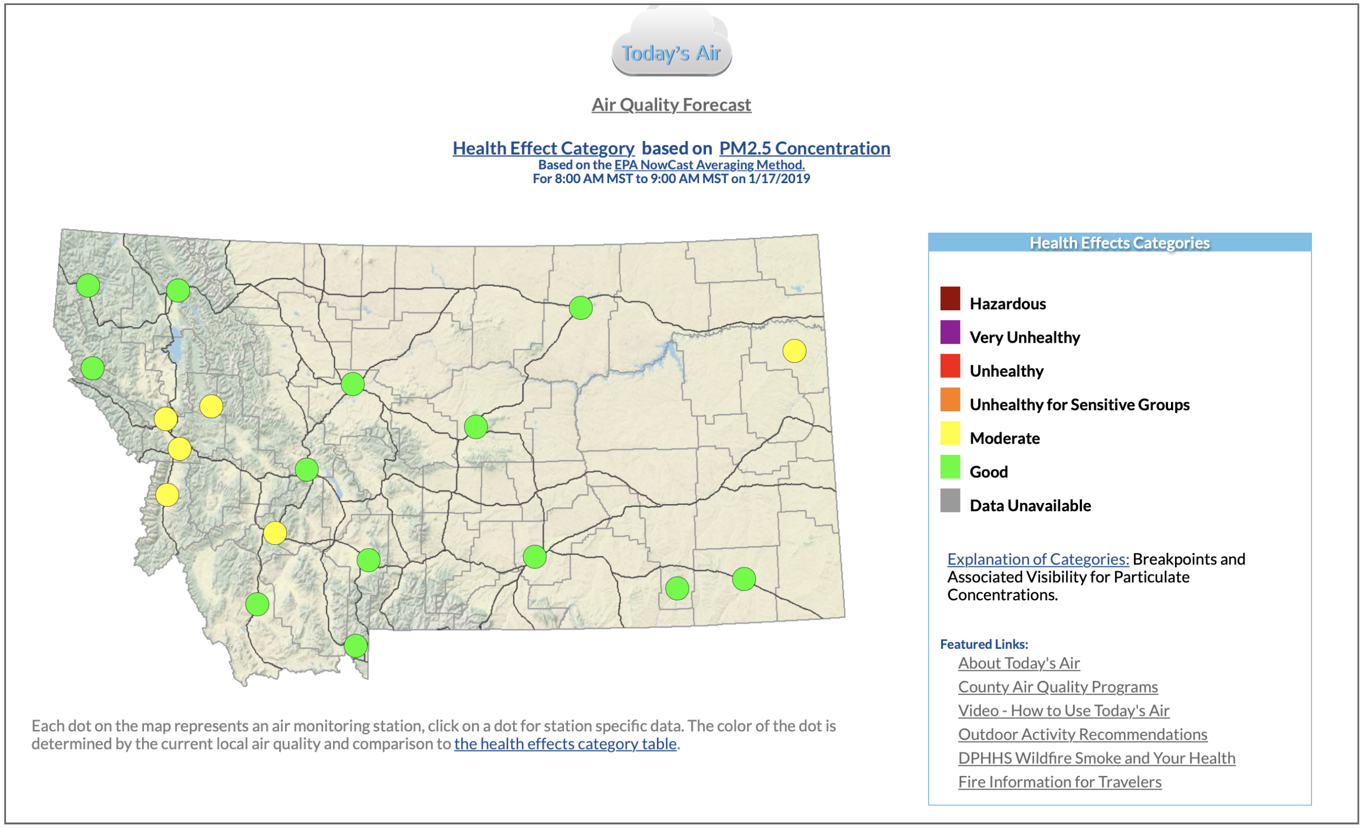
Lecture 4: Point Level Models

Class Intro

- What did we learn last time?
- For Today
  - What is point referenced data?
  - How do we predict the response at unobserved locations?

Overview of Point Level Models or Geostatistical Models

Visual Motivation



source: airnow.gov

Basic Notation

- The response, $Y(\boldsymbol{s})$, is a random variable at location(s) $\boldsymbol{s}$
- $\boldsymbol{s}$ varies continuously over the space
- Typically, $\boldsymbol{s} \in \mathcal{R}^2$ or $\boldsymbol{s} \in \mathcal{D}$, where $\mathcal{D} \subset \mathcal{R}^2$, such as Montana.
- $Y(\boldsymbol{s})$ is a stochastic process, or a collection of random variables.
- In spatial settings, $Y(\boldsymbol{s})$ is referred to as a *spatial process*.

Conceptual Challenge: Questions

Goals

- Inference about spatial process $Y(\boldsymbol{s})$
- Create a continous map of air quality, or more generally predict the response at new locations given the partial realization of the process

Discussion Questions

- What is the dimension of the spatial process, $Y(\boldsymbol{s})$?
- How are predictions made at unsampled locations, and a related question, what does "inference about a spatial process" mean?

Conceptual Challenge: Solutions

Goals

- Inference about spatial process $Y(\boldsymbol{s})$
- Create a continous map of air quality, or more generally predict the response at new locations given the partial realization of the process

Discussion Questions

- What is the dimension of the spatial process, $Y(\boldsymbol{s})$?
  It is defined across the entire space, so infinite dimensional.
- How are predictions made at unsampled locations, and a related question, what does "inference about a spatial process" mean?
  the spatial process is assumed to have mean (or covariate) properties as well as a specified covariance structure as a function of distance. Similar to a standard regression model, predictions are made using the covariates and covariance.

Spatial Covariance

- Spatial covariance is characterized by the distance between locations
- By specifying a structured form for the covariance function, few parameters are necessary
- A covariance function must be symmetric: $Cov(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)) = Cov(Y(\boldsymbol{s}_2), Y(\boldsymbol{s}_1))$
- A common covariance is the exponential covariance, where $Cov(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)) = \sigma^2 \exp(-\phi d_{12})$, where d_{12} is the distance between $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Exponential Covariance: Questions

The exponential covariance function has two parameters:

$Cov(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)) = \sigma^2 \exp(-\phi d_{12}),$

Describe the how the parameters contribute to the covariance:

- ϕ
- σ^2

Exponential Covariance: Solution

The exponential covariance function has two parameters:

$$Cov(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)) = \sigma^2 \exp(-\phi d_{12}),$$

Describe the how the parameters contribute to the covariance:

- ϕ:
  controls the correlation. In particular, $\exp(-\phi d_{12})$ will be between 0 and 1 (with $\phi > 0$ and $d_{12} \geq 0$). When $d_{12} \gg \frac{1}{\phi}$, the correlation goes to zero and When $d_{12} \ll \frac{1}{\phi}$, the correlation goes to one. $\frac{1}{\phi}$ is known as the *range parameter*.
- σ^2: is a scaling parameter associated with the correlation. σ^2 is called the partial sill

Exponential Covariance: Nugget

- Spatial models include a *nugget* effect to account point level variability. This can be thought of as measurement error, or randomness inherent in the process.
- Hence for a single point, $Cov(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_i)) = \sigma^2 \exp(-\phi d_{12}) + \tau^2$,
- In this setting, $\sigma^2 + \tau^2$ is called the *sill*.

Joint Distribution

A joint distribution is specified to model the spatial covariance using standard likelihood-based techniques.

- Let $\boldsymbol{Y} := \{Y(\boldsymbol{s}_i)\}$ for locations $\boldsymbol{s_i}, i = 1, \ldots, n$
- Then assume
  $\boldsymbol{Y}|\mu, \boldsymbol{\theta} \sim MVN_n(\mu \mathbf{1}_n, \Sigma(\boldsymbol{\theta}))$,
  where $\boldsymbol{\theta}$ is a vector of parameters that determine the variance.

Parameter Estimation: Questions

- Write out the likelihood for $\boldsymbol{Y}$
- Describe the process for estimating the parameters in the model

Parameter Estimation: Solutions

- Write out the likelihood for $\boldsymbol{Y}$:
  $$\boldsymbol{Y}|\mu, \boldsymbol{\theta} = (2\pi)^{n/2}|\Sigma(\boldsymbol{\theta})|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{Y} - \mathbf{1}\mu)^T\Sigma(\boldsymbol{\theta})^{-1}(\boldsymbol{Y} - \mathbf{1}\mu)\right]$$
- Describe the process for estimating the parameters in the model
  maximum likelihood for μ, and $\boldsymbol{\theta}$ or with Bayesian techniques using prior distributions and Markov Chain Monte Carlo

Gaussian Process

Given that we assume
$\boldsymbol{Y}|\mu, \boldsymbol{\theta} \sim MVN_n(\mu \mathbf{1}_n, \Sigma(\boldsymbol{\theta}))$,

- it is also reasonable to assume that the response at a set of unobserved locations, $\boldsymbol{Y}(\boldsymbol{s}^*)$ also follows a multivariate normal distribution.
- this results in an infinite dimensional normal distribution, which is also know as a Gaussian Process (GP)
- prediction, say $\boldsymbol{Y}(\boldsymbol{s}^*)|\boldsymbol{Y}(\boldsymbol{s})$ follows a conditional multivariate normal distribution

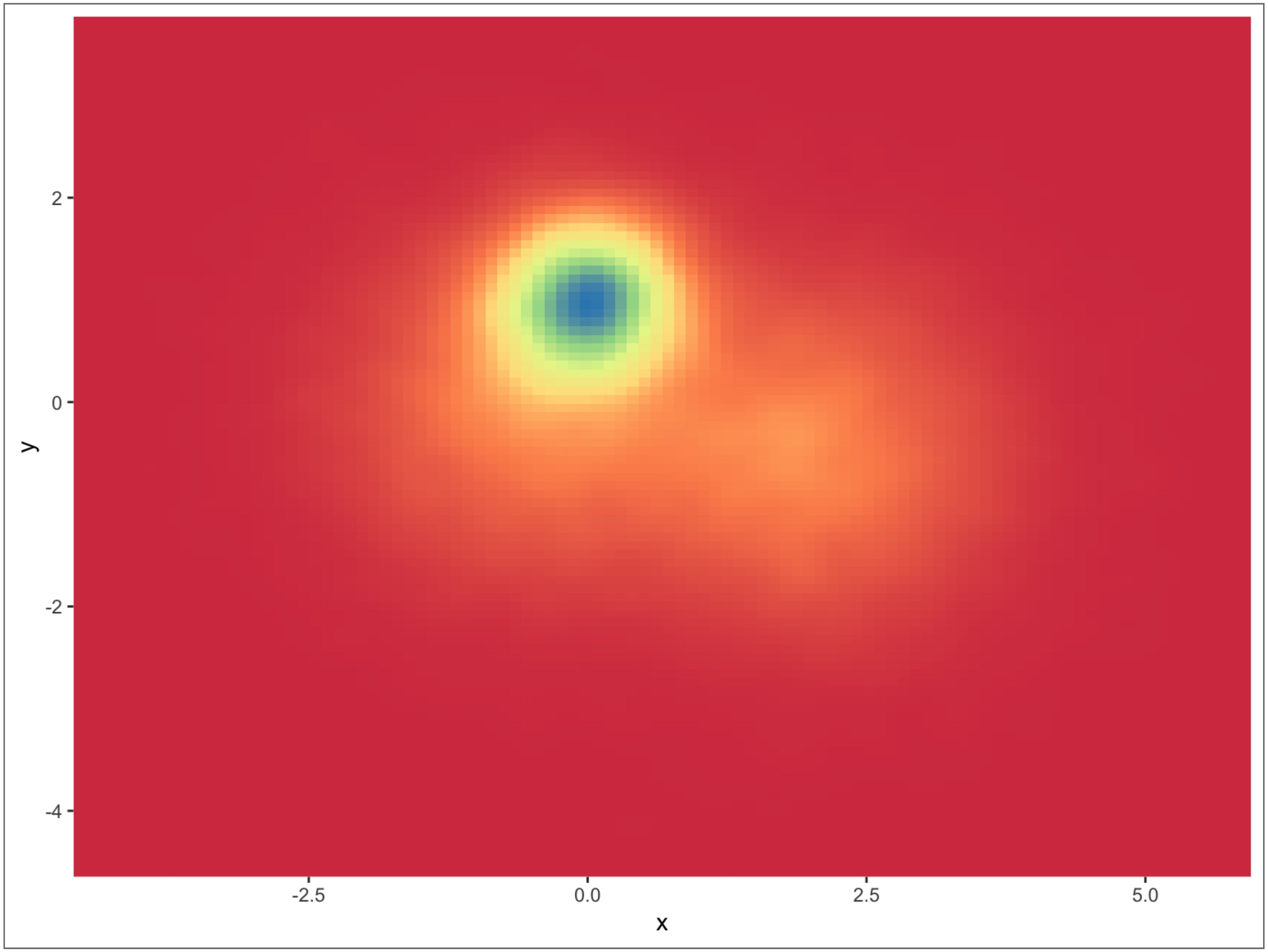
Elements of Geostatistical models

Stationarity

- Assume the spatial process, $\boldsymbol{Y}(\boldsymbol{s})$ has a mean, $\boldsymbol{\mu}(\boldsymbol{s})$, and that the variance of $\boldsymbol{Y}(\boldsymbol{s})$ exists everywhere.
- The process is **strictly stationary** if: for any $n \geq 1$, any set of n sites $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ and any $\boldsymbol{h} \in \mathcal{R}^r$ (typically r =2), the distribution of $(\boldsymbol{Y}(\boldsymbol{s}_1). \ldots, \boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}_n))$ is the same as $\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}_1)$

Strict Stationarity

Is this strictly stationary?

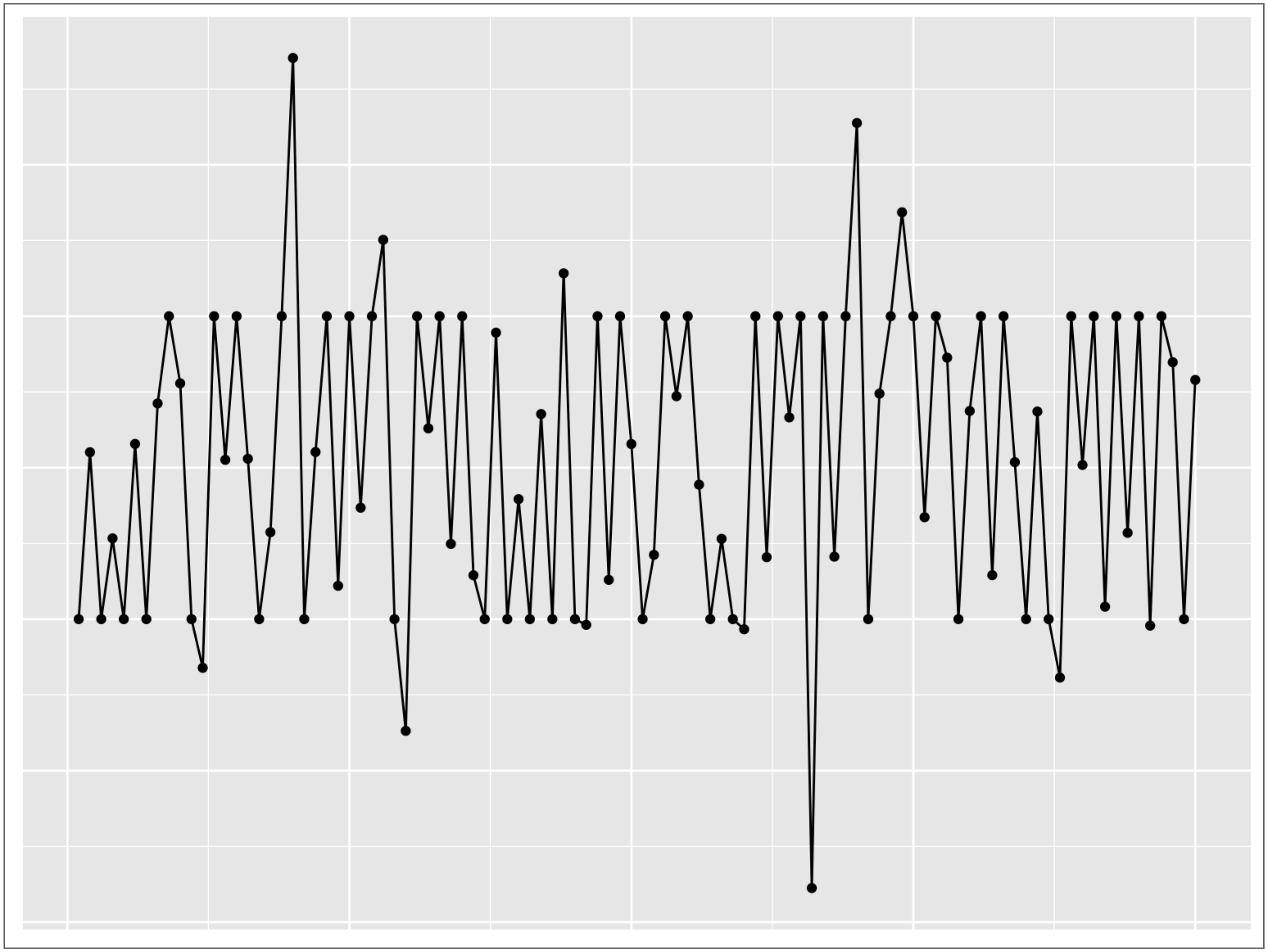

2
0
-2
-4
y
-2.5
0.0
2.5
5.0
x

Weak Stationarity

- Weak stationarity, or second-order stationarity of a spatial process, requires a constant mean and covariance that is a function of $\boldsymbol{h}$, $Cov(\boldsymbol{Y}(\boldsymbol{s}), \boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h})) = C(\boldsymbol{h})$, where $C(\boldsymbol{h})$ is a covariance function that only requires the distance between the points.
- Typically with spatial models, the process is assumed to be mean zero as covariates explain the mean structure. So second-order stationarity is primarily focused on the covariance structure.

Weak Stationarity

Is this weakly stationary?



Intrinsic Stationarity

- A third kind of stationarity, known as intrinsic stationarity, describes the behavior of differences in the spatial process, rather than the data.
- Intrinsic stationarity assumes $E[\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})] = 0$ then define $E[\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})]^2 = Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})) = 2\gamma(\boldsymbol{h}),$
  which only works, and satisfies intrinsic stationarity, if the equation only depends on $\boldsymbol{h}$.

Variograms

- Intrinsic stationary justifies the use of variograms $2\gamma(\boldsymbol{h})$
- Variograms are often used to visualize spatial patterns:
  - If the distance between points is small, the variogram is expected to be small
  - As the distance between points increases, the variogram increases
- There is a mathematical link between the covariance function $C(\boldsymbol{h})$ and the variogram $2\gamma(\boldsymbol{h})$.

Isotropy

- If the variogram $2\gamma(\boldsymbol{h})$ depends only on the length of $\boldsymbol{h}$, $||\boldsymbol{h}||$, and not the direction, then the variogram is isotropic.
- If the direction of $\boldsymbol{h}$ impacts the variogram, then the variogram is anisotropic.

More about Variograms

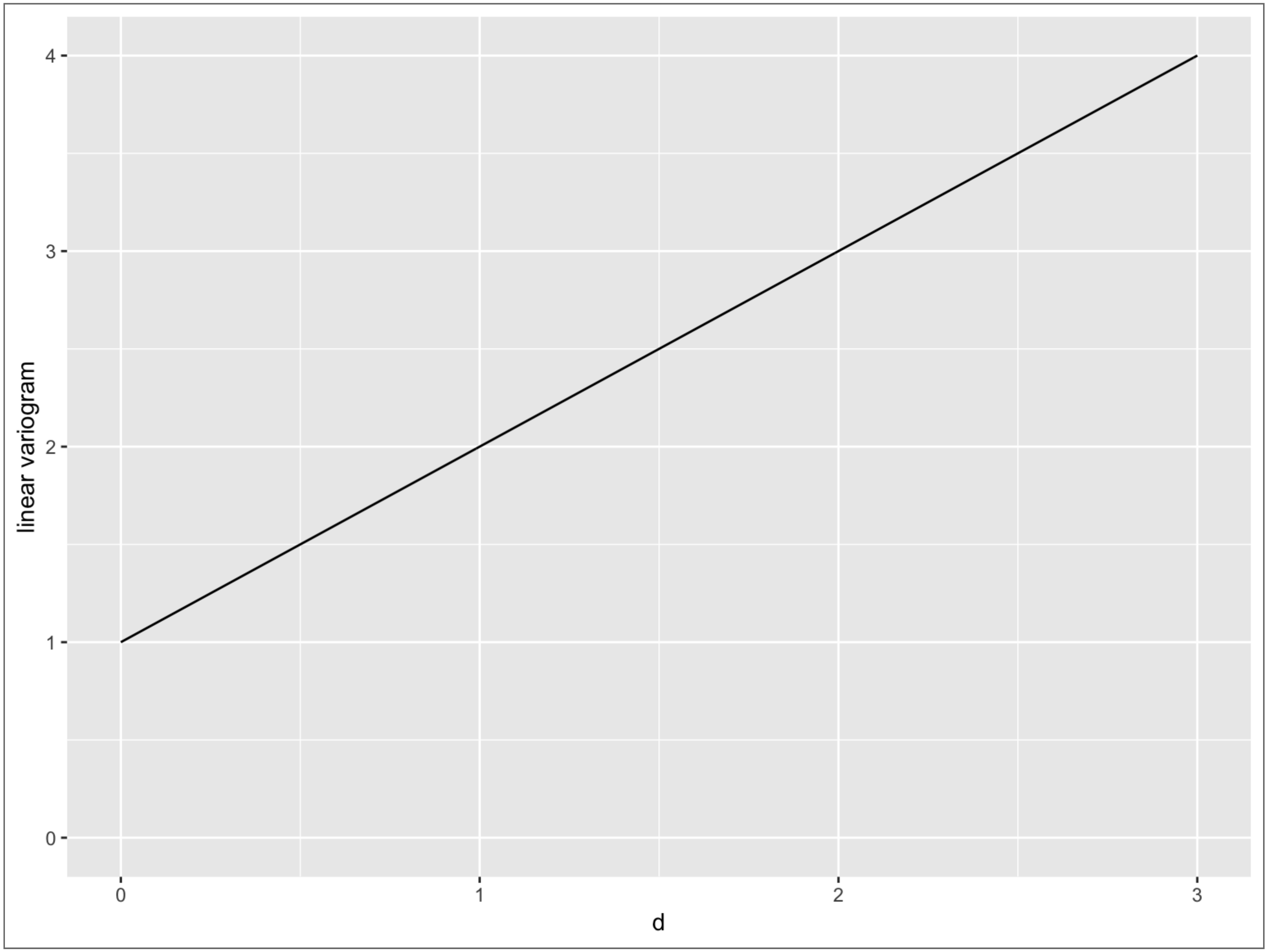
Variogram to the Covariance Function

$$
\begin{aligned}
2\gamma(\boldsymbol{h}) &= Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})) \\
&= Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h})) + Var(\boldsymbol{Y}(\boldsymbol{s})) - 2Cov(Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}), \boldsymbol{Y}(\boldsymbol{s}))) \\
&= C(\boldsymbol{0}) + C(\boldsymbol{0}) - 2C(\boldsymbol{h}) \\
&= 2\left[C(\boldsymbol{0}) + C(\boldsymbol{h})\right]
\end{aligned}
$$

- Given $C()$, the variogram can easily be recovered
- Going the other way, $2\gamma() \rightarrow C()$, requires additional assumptions

Linear semivariogram

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases}$$


linear variogram
4
3
2
1
0
0
1
2
3
d

Spherical semivariogram: exercise

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d \geq 1/\phi \\ \tau^2 + \sigma^2 \left[\frac{3\phi d}{2} - \frac{1}{2}(\phi d)^3 \right] \\ 0 & \text{otherwise} \end{cases}$$

- Sketch, or generate in R, a spherical semivariogram
- On this figure label the nugget, sill, partial sill, and range.