Lecture 13

Gaussian Process Models - Part 2

3/06/2018

EDA and GPs

Variogram

When fitting a Gaussian process model, it is often difficult to fit the covariance parameters (hard to identify). Today we will discuss some EDA approaches for getting a sense of the values for the scale, range and nugget parameters.

σ^2 d σ_w^2

Variogram

When fitting a Gaussian process model, it is often difficult to fit the covariance parameters (hard to identify). Today we will discuss some EDA approaches for getting a sense of the values for the scale, range and nugget parameters.

From the spatial modeling literature the typical approach is to examine an *empirical variogram*, first we will define the *theoretical variogram* and its connection to the covariance.

Variogram

When fitting a Gaussian process model, it is often difficult to fit the covariance parameters (hard to identify). Today we will discuss some EDA approaches for getting a sense of the values for the scale, range and nugget parameters.

From the spatial modeling literature the typical approach is to examine an *empirical variogram*, first we will define the *theoretical variogram* and its connection to the covariance.

Variogram:

$$2\gamma(t_i, t_j) = Var(Y(t_i) - Y(t_j))$$

where $\gamma(t_i, t_j)$ is known as the semivariogram.

Some Properties of the theoretical Variogram / Semivariogram

- are non-negative

$$\gamma(t_i, t_j) \geq 0$$

Some Properties of the theoretical Variogram / Semivariogram

- are non-negative

$$\gamma(t_i, t_j) \geq 0$$

- are equal to 0 at distance 0

$$\gamma(t_i, t_i) = 0$$

Some Properties of the theoretical Variogram / Semivariogram

- are non-negative

$$\gamma(t_i, t_j) \geq 0$$

- are equal to 0 at distance 0

$$\gamma(t_i, t_i) = 0$$

- are symmetric

$$\gamma(t_i, t_j) = \gamma(t_j, t_i)$$

Some Properties of the theoretical Variogram / Semivariogram

- are non-negative

$$\gamma(t_i, t_j) \geq 0$$

- are equal to 0 at distance 0

$$\gamma(t_i, t_i) = 0$$

- are symmetric

$$\gamma(t_i, t_j) = \gamma(t_j, t_i)$$

- if there is no dependence then

$$2\gamma(t_i, t_j) = Var(Y(t_i)) + Var(Y(t_j)) \quad \text{for all } i \neq j$$

Some Properties of the theoretical Variogram / Semivariogram

- are non-negative

$$\gamma(t_i, t_j) \geq 0$$

- are equal to 0 at distance 0

$$\gamma(t_i, t_i) = 0$$

- are symmetric

$$\gamma(t_i, t_j) = \gamma(t_j, t_i)$$

- if there is no dependence then

$$2\gamma(t_i, t_j) = Var(Y(t_i)) + Var(Y(t_j)) \quad \text{for all } i \neq j$$

- if the process *is not* stationary

$$2\gamma(t_i, t_j) = Var(Y(t_i)) + Var(Y(t_j)) - 2\,Cov(Y(t_i), Y(t_j))$$

Some Properties of the theoretical Variogram / Semivariogram

- are non-negative

$$\gamma(t_i, t_j) \geq 0$$

- are equal to 0 at distance 0

$$\gamma(t_i, t_i) = 0$$

- are symmetric

$$\gamma(t_i, t_j) = \gamma(t_j, t_i)$$

- if there is no dependence then

$$2\gamma(t_i, t_j) = Var(Y(t_i)) + Var(Y(t_j)) \quad \text{for all } i \neq j$$

- if the process *is not* stationary

$$2\gamma(t_i, t_j) = Var(Y(t_i)) + Var(Y(t_j)) - 2\,Cov(Y(t_i), Y(t_j))$$

- if the process *is* stationary

$$2\gamma(t_i, t_j) = 2Var(Y(t_i)) - 2\,Cov(Y(t_i), Y(t_j))$$

We will assume that our process of interest is stationary, in which case we will parameterize the semivariagram in terms of $h = |t_i - t_j|$.

Empirical Semivariogram:

$$\hat{\gamma}(h) = \frac{1}{2\,N(h)} \sum_{|t_i - t_j| \in (h-\epsilon, h+\epsilon)} (Y(t_i) - Y(t_j))^2$$

Empirical Semivariogram

We will assume that our process of interest is stationary, in which case we will parameterize the semivariagram in terms of $h = |t_i - t_j|$.

Empirical Semivariogram:

$$\hat{\gamma}(h) = \frac{1}{2\,N(h)} \sum_{|t_i - t_j| \in (h-\epsilon, h+\epsilon)} (Y(t_i) - Y(t_j))^2$$

Practically, for any data set with n observations there are $\binom{n}{2} \cancel{+ n}$ possible data pairs to examine. Each individually is not very informative, so we aggregate into bins and calculate the empirical semivariogram for each bin.

Connection to Covariance

Exp Cov

$$Cov(Y(t_i), Y(t_j)) = \sigma^2 e^{-d|t_i - t_j|}$$

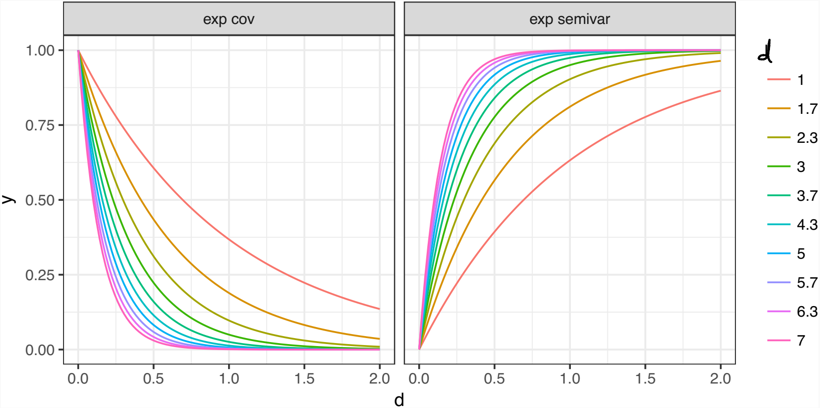
Sq Exp Cov

$$Cov(Y(t_i), Y(t_j)) = \sigma^2 e^{-(d(t_i - t_j))^2}$$

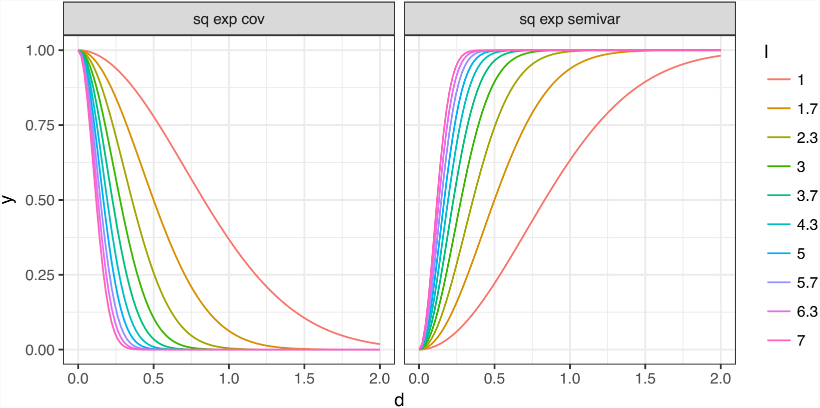
$$\gamma(t_i, t_j) = Var(Y(t_i)) - Cov(Y(t_i), Y(t_j))$$

$$= \sigma^2 - \sigma^2 e^{-(d(t_i - t_j))^2}$$

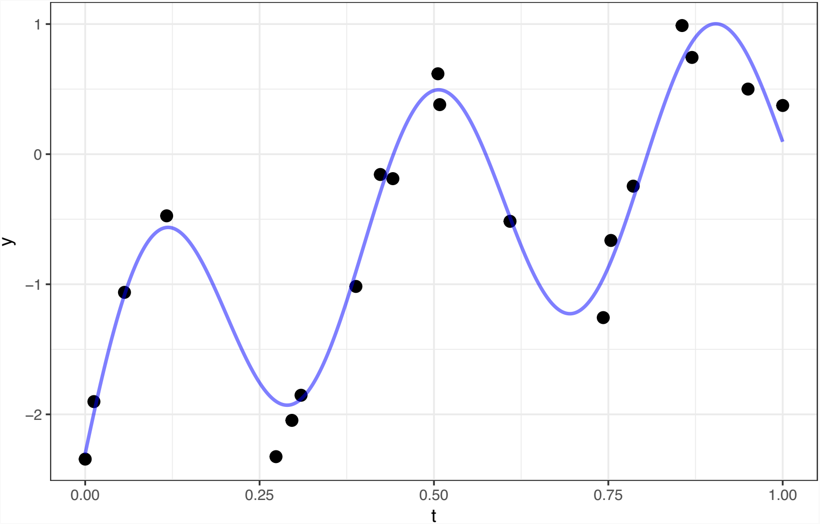
Covariance vs Semivariogram - Exponential



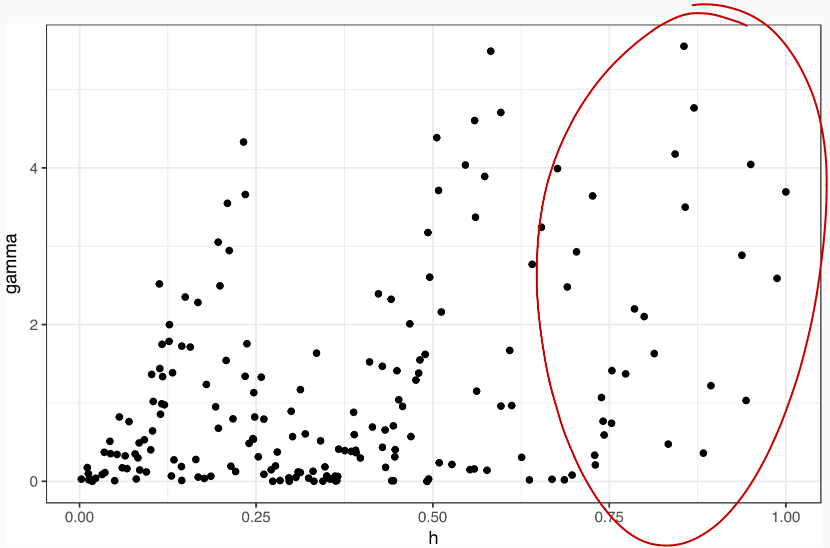
Covariance vs Semivariogram - Square Exponential



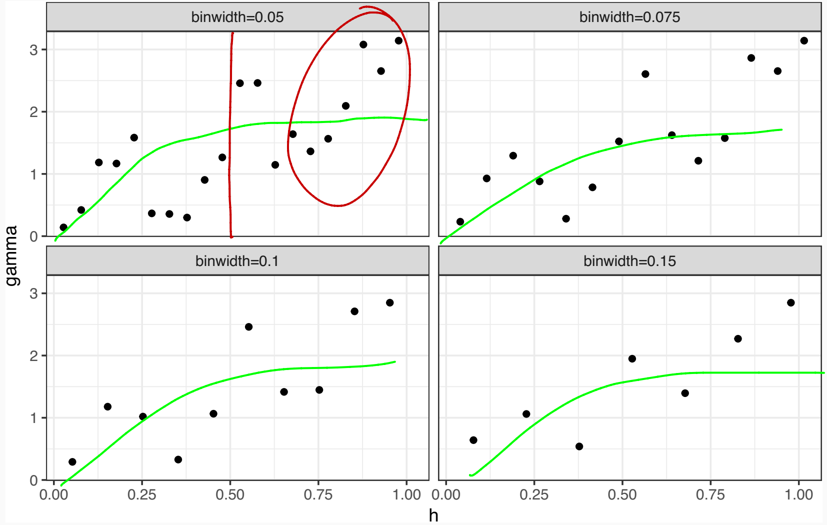
From last time

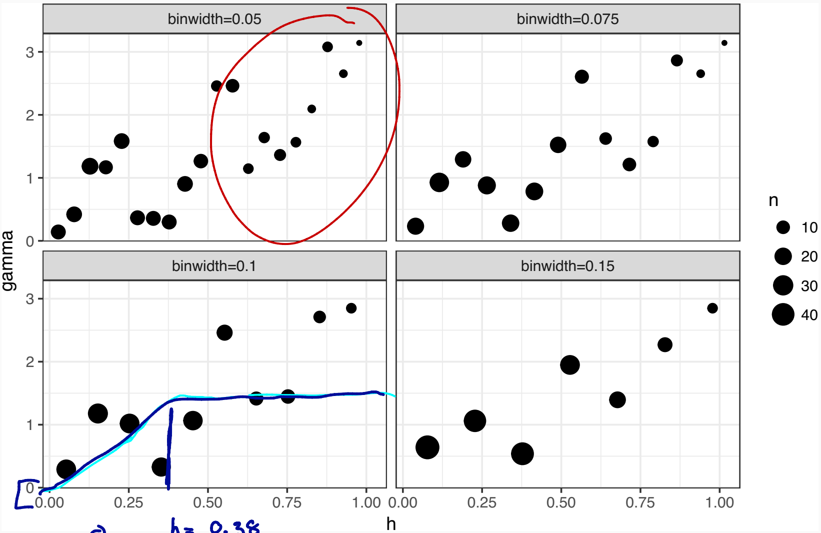


Empirical semivariogram - no bins / cloud



Empirical semivariogram (binned)





nugget = 0

h = 0.38

$\ell = \frac{\sqrt{3}}{0.38}$

$\sigma^2 + \cancel{\sigma^2} = 1.5$

Theoretical vs empirical semivariogram

After fitting the model last time we came up with a posterior median of $\sigma^2 = 1.89$ and $l = 5.86$ for a square exponential covariance.

Theoretical vs empirical semivariogram

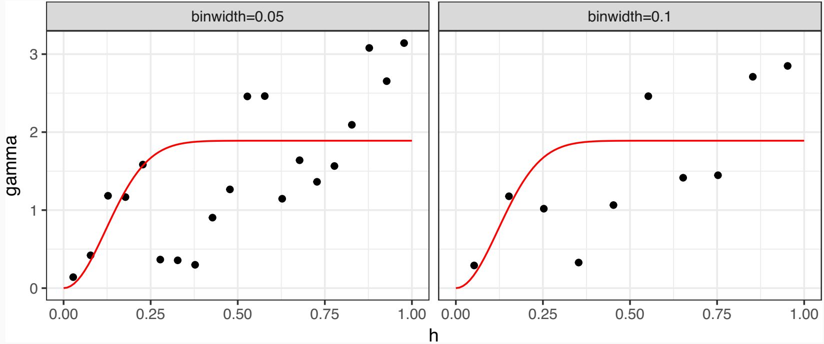
After fitting the model last time we came up with a posterior median of $\sigma^2 = 1.89$ and $l = 5.86$ for a square exponential covariance.

$$
\begin{aligned}
Cov(h) &= \sigma^2 \exp\big(-(h\,l)^2\big) \\
\gamma(h) &= \sigma^2 - \sigma^2 \exp\big(-(h\,l)^2\big) \\
&= 1.89 - 1.89 \exp\big(-(5.86\,h)^2\big)
\end{aligned}
$$

Theoretical vs empirical semivariogram

After fitting the model last time we came up with a posterior median of $\sigma^2 = 1.89$ and $l = 5.86$ for a square exponential covariance.

$$
\begin{aligned}
Cov(h) &= \sigma^2 \exp\left(-(h\,l)^2\right) \\
\gamma(h) &= \sigma^2 - \sigma^2 \exp\left(-(h\,l)^2\right) \\
&= 1.89 - 1.89 \exp\left(-(5.86\,h)^2\right)
\end{aligned}
$$



$$\mathrm{Cov}\left(Y(t_i), Y(t_j)\right) = \sigma^2 e^{-(\ell h)^2} + \sigma_w^2 \mathbb{1}_{h=0}$$

$$h = |t_i - t_j|$$

$$\mathrm{Var}\left(Y(t_i)\right) = \sigma^2 + \sigma_w^2$$

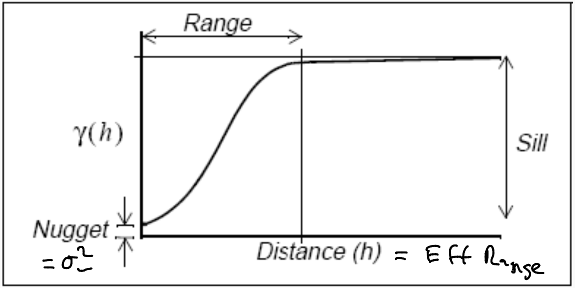
$$\gamma(t_i, t_j) = \sigma^2 + \sigma_w^2 - \left(\sigma^2 e^{-(\ell h)^2} + \sigma_w^2 \mathbb{1}_{h=0}\right)$$

Variogram features

σ^2 - Scale

ℓ - range

σ_w^2 - nugget



Sill $= \sigma^2 + \sigma_w^2$

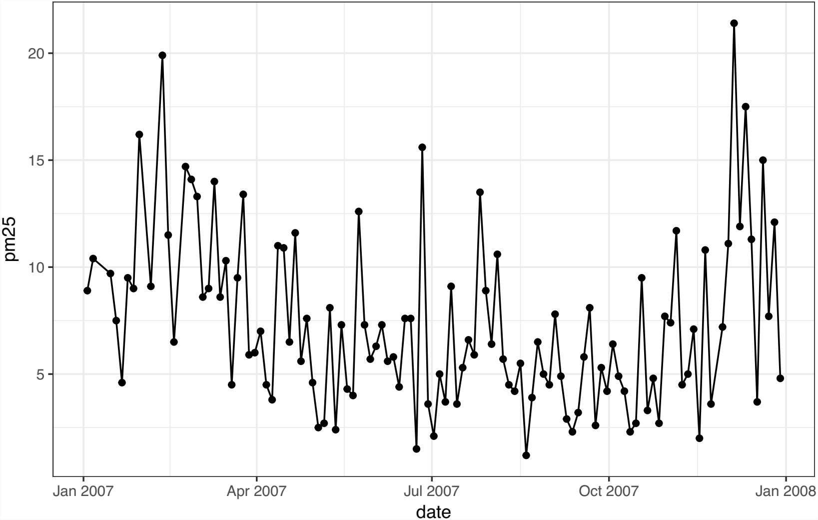
Nugget $= \sigma_w^2$

Distance (h) = Eff Range

$= \frac{\sqrt{3}}{\ell}$

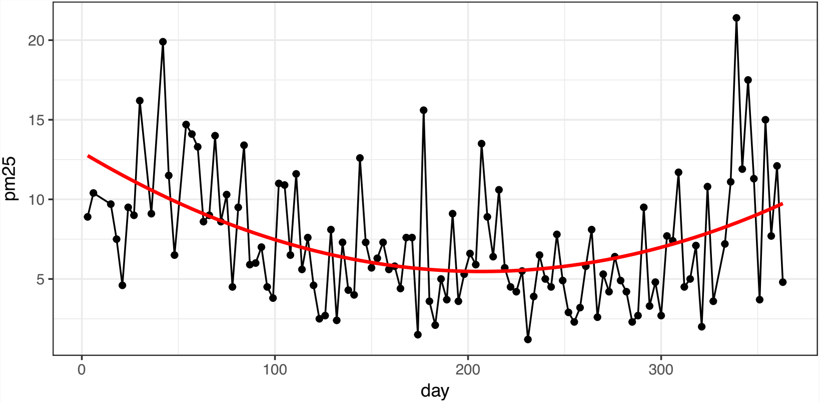
PM2.5 Example

Measured PM2.5 data from an EPA monitoring station in Columbia, NJ.



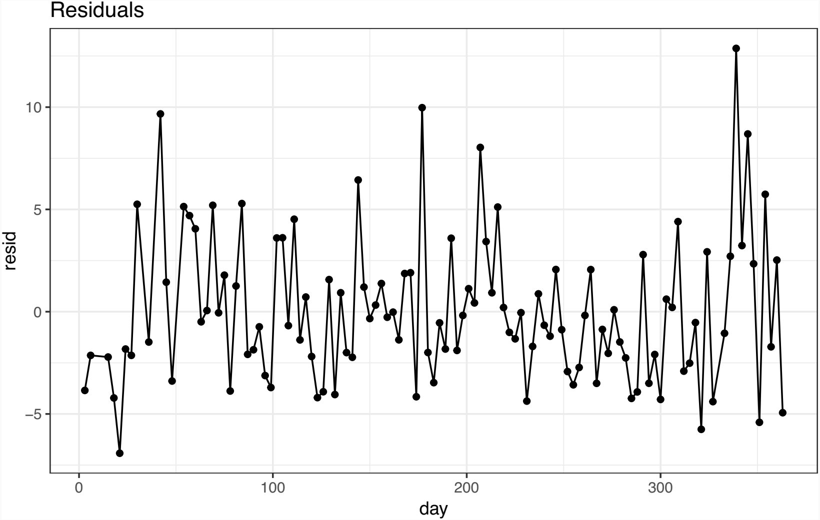
site	latitude	longitude	pm25	date	day
230031011	46.682	-68.016	8.9	2007-01-03	3
230031011	46.682	-68.016	10.4	2007-01-06	6
230031011	46.682	-68.016	9.7	2007-01-15	15
230031011	46.682	-68.016	7.5	2007-01-18	18
230031011	46.682	-68.016	4.6	2007-01-21	21
230031011	46.682	-68.016	9.5	2007-01-24	24
230031011	46.682	-68.016	9.0	2007-01-27	27
230031011	46.682	-68.016	16.2	2007-01-30	30
230031011	46.682	-68.016	9.1	2007-02-05	36
230031011	46.682	-68.016	19.9	2007-02-11	42
230031011	46.682	-68.016	11.5	2007-02-14	45
230031011	46.682	-68.016	6.5	2007-02-17	48
230031011	46.682	-68.016	14.7	2007-02-23	54
230031011	46.682	-68.016	14.1	2007-02-26	57
230031011	46.682	-68.016	13.3	2007-03-01	60
230031011	46.682	-68.016	8.6	2007-03-04	63
230031011	46.682	-68.016	9.0	2007-03-07	66
230031011	46.682	-68.016	14.0	2007-03-10	69
230031011	46.682	-68.016	8.6	2007-03-13	72
230031011	46.682	-68.016	10.3	2007-03-16	75

Mean Model

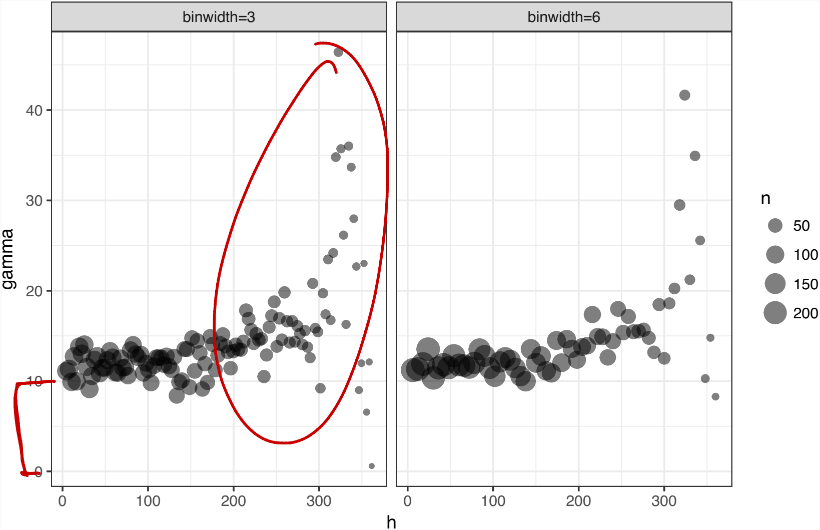



```
##
## Call:
## lm(formula = pm25 ~ day + I(day^2), data = pm25)
##
## Coefficients:
## (Intercept)          day     I(day^2)
##  12.9644351   -0.0724639    0.0001751
##
```

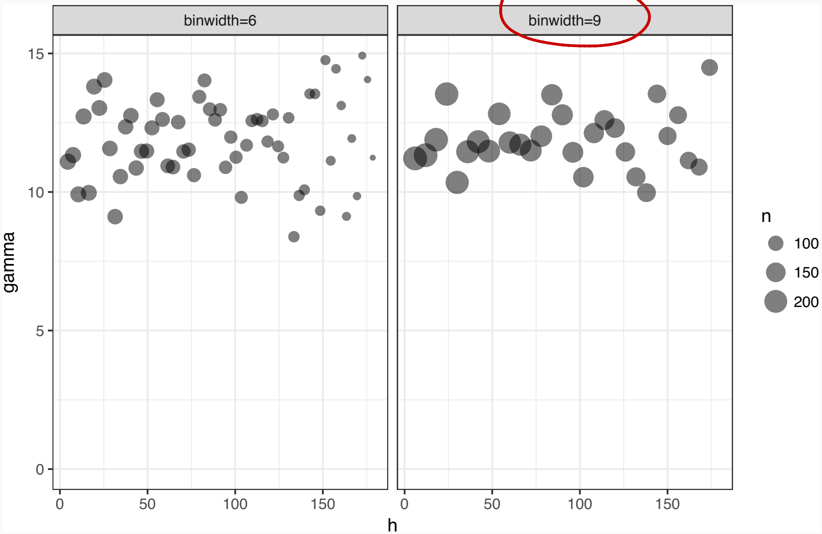
Detrended Residuals



Empirical Variogram



Empirical Variogram



Model

What does the model we are trying to fit actually look like?

What does the model we are trying to fit actually look like?

$$y(t) = \mu(t) + w(t) + \epsilon(t)$$

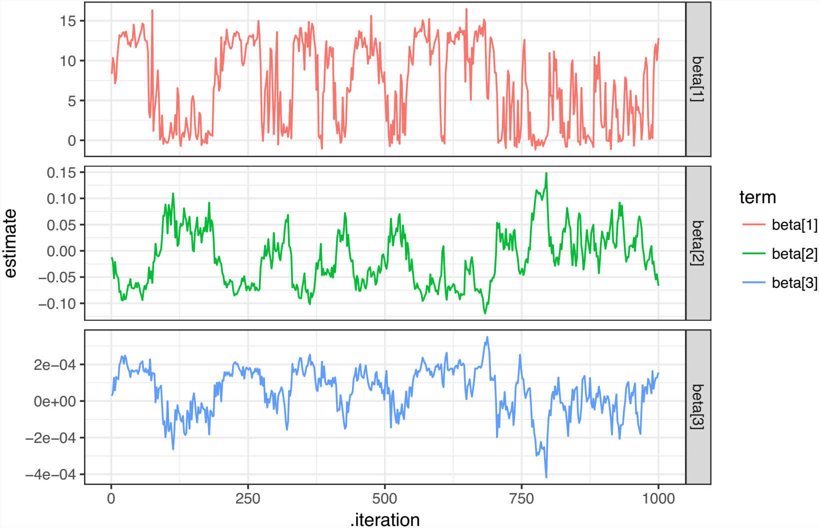
where

$$\begin{aligned}
\boldsymbol{\mu}(\mathbf{t}) &= \beta_0 + \beta_1\,\mathbf{t} + \beta_2\,\mathbf{t}^2 \\
\mathbf{w}(\mathbf{t}) &\sim \mathcal{GP}(0, \boldsymbol{\Sigma}) \\
\epsilon(t) &\sim \mathcal{N}(0, \sigma^2_w)
\end{aligned}$$

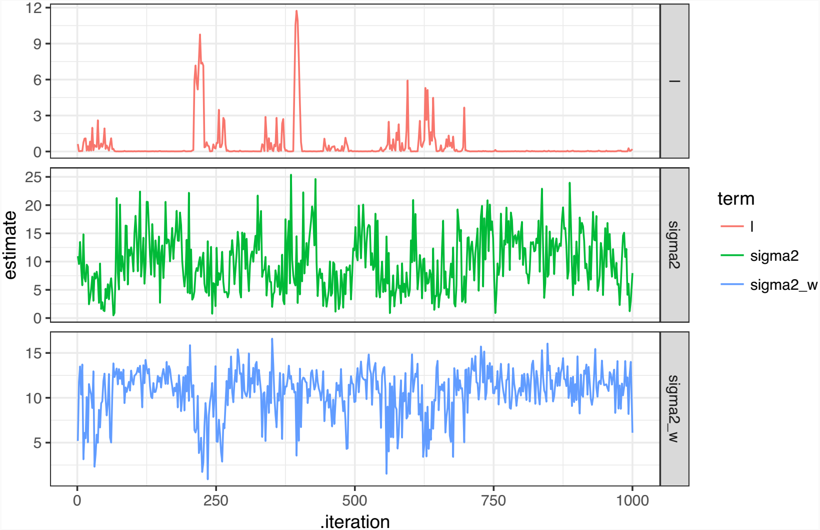
JAGS Model

```
gp_exp_model = "model{
  y ~ dmnorm(mu, inverse(Sigma))

  for (i in 1:N) {
    mu[i] <- beta[1]+ beta[2] * x[i] + beta[3] * x[i]^2
  }

  for (i in 1:(N-1)) {
    for (j in (i+1):N) {
      Sigma[i,j] <- sigma2 * exp(- pow(l*d[i,j],2))
      Sigma[j,i] <- Sigma[i,j]
    }
  }

  for (k in 1:N) {
    Sigma[k,k] <- sigma2 + sigma2_w
  }

  for (i in 1:3) {
    beta[i] ~ dt(0, 2.5, 1)
  }
  sigma2_w ~ dnorm(10, 1/25) T(0,)
  sigma2   ~ dnorm(10, 1/25) T(0,)
  l        ~ dt(0, 2.5, 1) T(0,)
}"
```

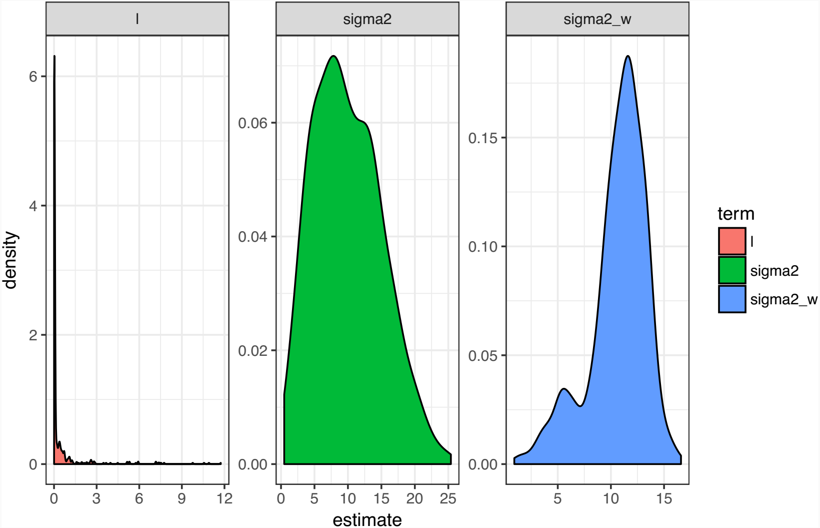
Posterior - Betas



Posterior - Covariance Parameters



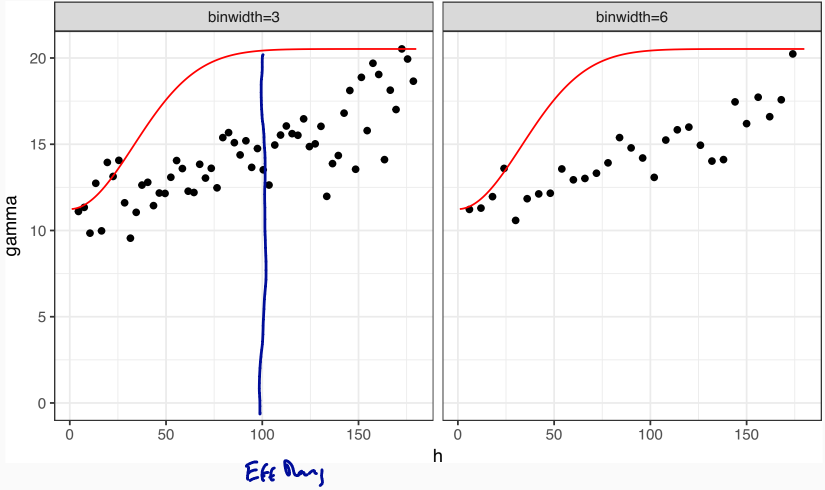
Posterior - Covariance Parameters



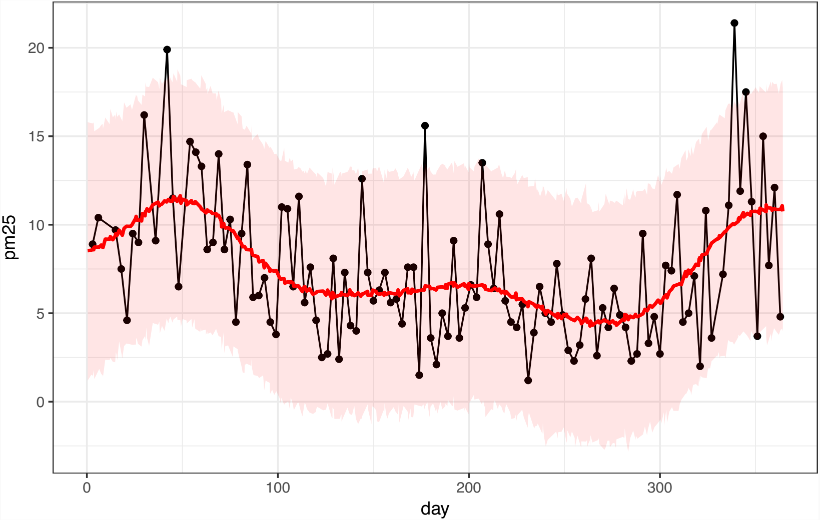
Posterior

term	post_mean	post_med	post_lower	post_upper
beta[1]	6.922	7.624	-0.720	14.369
beta[2]	-0.015	-0.018	-0.091	0.092
beta[3]	0.000	0.000	0.000	0.000
l	0.462	0.021	0.007	5.303
sigma2	9.862	9.284	1.610	20.555
sigma2_w	10.748	11.239	3.832	14.693

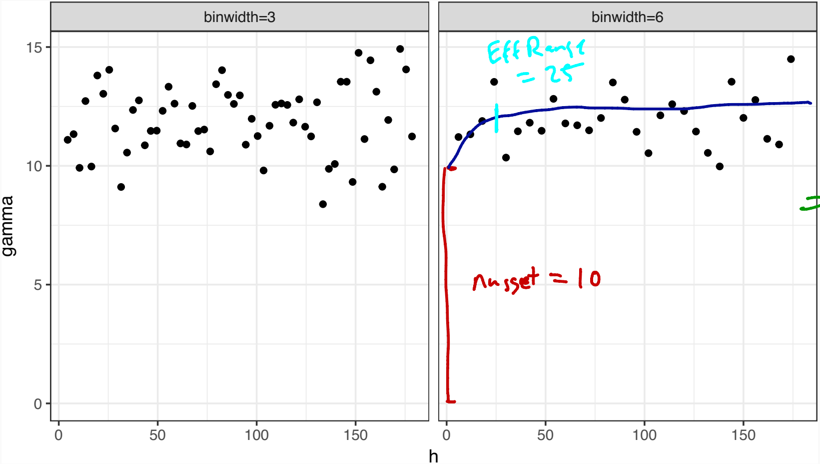
Empirical + Fitted Variogram



Fitted Model + Predictions

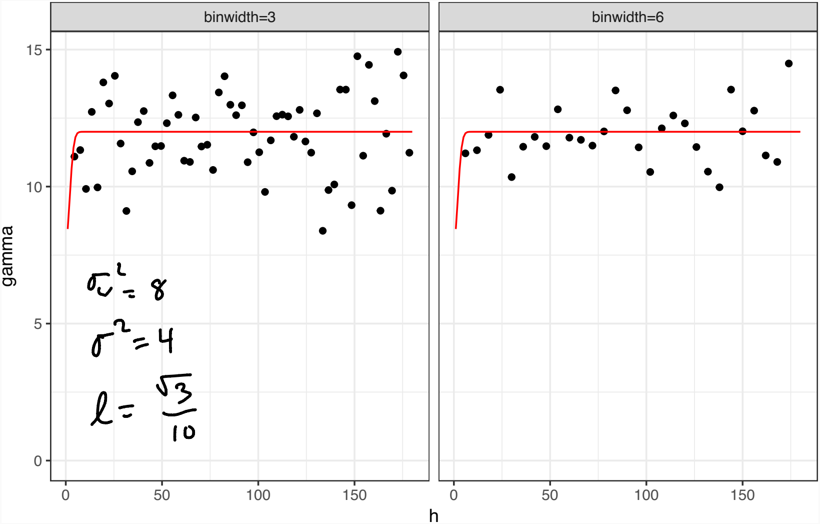


Empirical Variogram Model

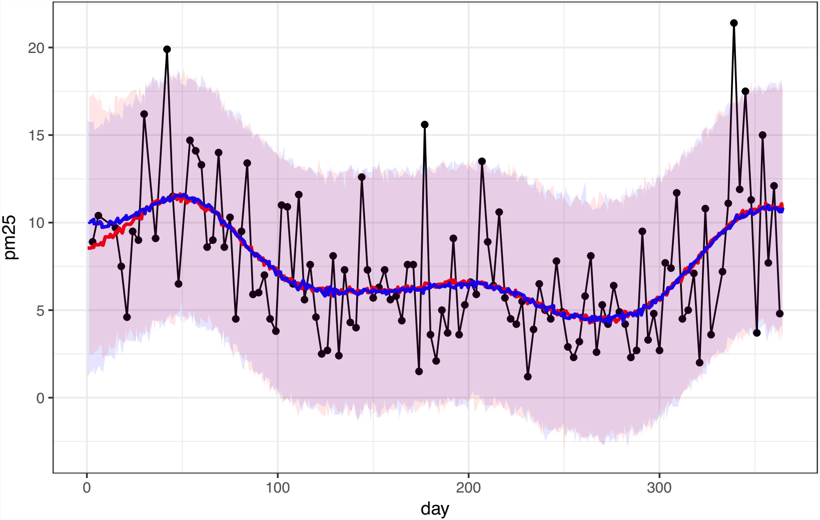


Sill = 12.5 $\Rightarrow \sigma_v^2 = 2.5$

Fit


binwidth=3
binwidth=6
gamma
h
0
5
10
15
0
50
100
150
$\sigma_w^2 = 8$
$\sigma^2 = 4$
$l = \frac{\sqrt{3}}{10}$

Predictions



Full Posterior Predictive Distribution

Plug in Prediction

```
l          = filter(post, term == 'l')         %>% pull(post_med)
sigma2     = filter(post, term == 'sigma2')    %>% pull(post_med)
sigma2_w = filter(post, term == 'sigma2_w') %>% pull(post_med)
beta0      = filter(post, term == 'beta[1]') %>% pull(post_med)
beta1      = filter(post, term == 'beta[2]') %>% pull(post_med)
beta2      = filter(post, term == 'beta[3]') %>% pull(post_med)

reps=1000

x = pm25$day
y = pm25$pm25
x_pred = 1:365 + rnorm(365, 0.01)

mu = beta0 + beta1*x + beta2*x^2
mu_pred = beta0 + beta1*x_pred + beta2*x_pred^2

dist_o = fields::rdist(x)
dist_p = fields::rdist(x_pred)
dist_op = fields::rdist(x, x_pred)
dist_po = t(dist_op)

cov_o  = sq_exp_cov(dist_o,  sigma2 = sigma2, l = l) + nugget_cov(dist_o, sigma2 = sig
cov_p  = sq_exp_cov(dist_p,  sigma2 = sigma2, l = l) + nugget_cov(dist_p, sigma2 = sig
cov_op = sq_exp_cov(dist_op, sigma2 = sigma2, l = l) + nugget_cov(dist_op, sigma2 = si
cov_po = sq_exp_cov(dist_po, sigma2 = sigma2, l = l) + nugget_cov(dist_po, sigma2 = si

inv = solve(cov_o, cov_op)
cond_cov = cov_p - cov_po %*% inv
```

Full Posterior Predictive Distribution

Our posterior consists of samples from

$$l, \sigma^2, \sigma_w^2, \beta_0, \beta_1, \beta_2 \mid \mathbf{y}$$

and for the purposes of generating the posterior predictions we sampled

$$\mathbf{y}_{pred} \mid l^{(m)}, {\sigma^2}^{(m)}, {\sigma_w^2}^{(m)}, {\beta_0}^{(m)}, {\beta_1}^{(m)}, {\beta_2}^{(m)}, \mathbf{y}$$

where $l^{(m)}, \ldots, {\beta_2}^{(m)}$, etc. are the posterior median of that parameter.

Our posterior consists of samples from

$$l, \sigma^2, \sigma_w^2, \beta_0, \beta_1, \beta_2 \mid \mathbf{y}$$

and for the purposes of generating the posterior predictions we sampled

$$\mathbf{y}_{pred} \mid l^{(m)}, {\sigma^2}^{(m)}, {\sigma_w^2}^{(m)}, {\beta_0}^{(m)}, {\beta_1}^{(m)}, {\beta_2}^{(m)}, \mathbf{y}$$

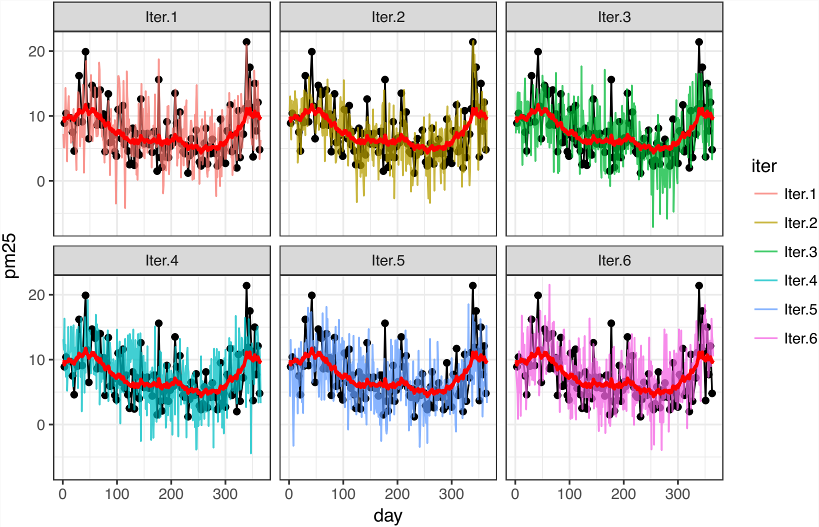
where $l^{(m)}, \ldots, {\beta_2}^{(m)}$, etc. are the posterior median of that parameter.

In practice we should instead be sampling

$$\mathbf{y}_{pred}^{(i)} \mid l^{(i)}, {\sigma^2}^{(i)}, {\sigma_w^2}^{(i)}, {\beta_0}^{(i)}, {\beta_1}^{(i)}, {\beta_2}^{(i)}, \mathbf{y}$$

since this takes into account the additional uncertainty in the model parameters.

Full Posterior Predictive Distribution - Plots



Full Posterior Predictive Distribution - Median + CI