

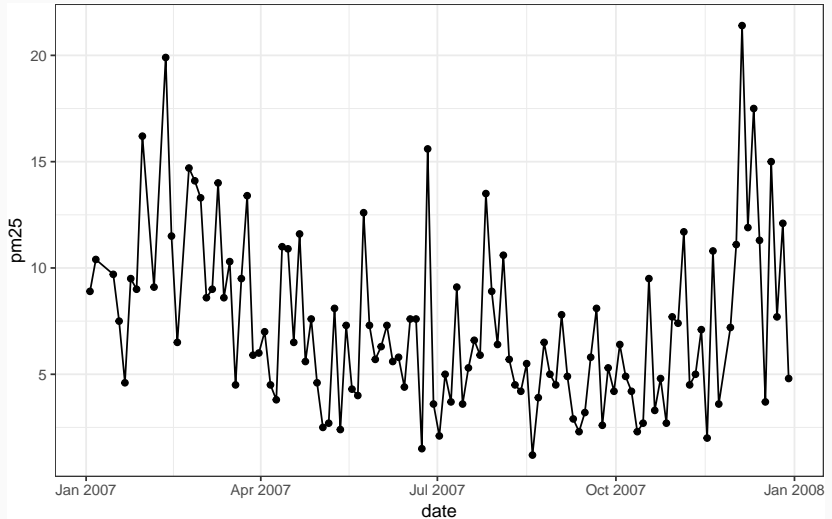
Lecture 14

Covariance Functions

10/24/2018

PM2.5 Example - Revisited

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



Previous JAGS Model

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

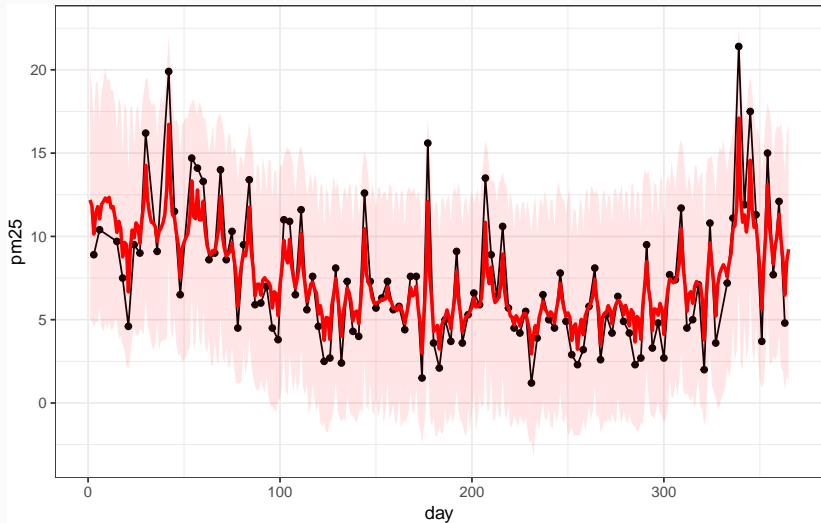
  for (i in 1:N) {
    mu[i] <- beta[1]+ beta[2] * x[i] + beta[3] * x[i]^2
  }

  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(coef[i], 2.5, 1)
  }
  sigma2_w ~ dnorm(5, 1/25) T(0,)
  sigma2    ~ dnorm(12.5, 1/25) T(0,)
  l         ~ dt(0, 2.5, 1) T(0,)
}"
```

Posterior Predictive Distribution



Revised JAGS Model

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

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

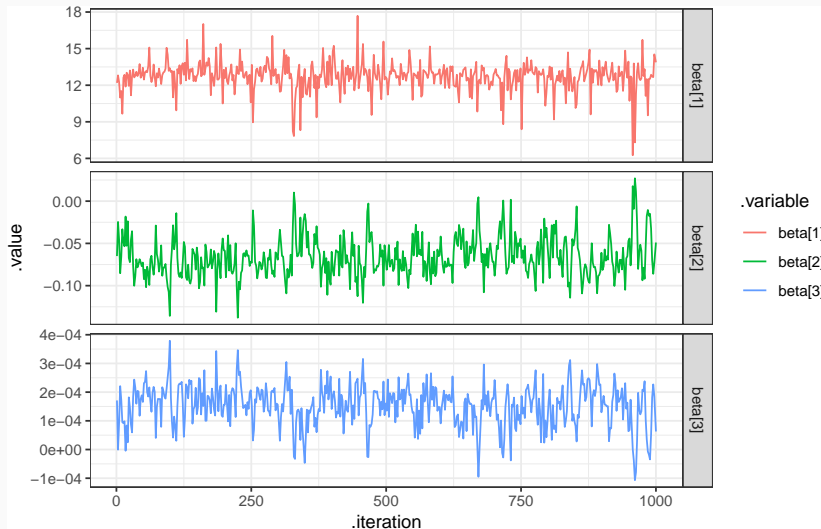
  for (i in 1:(N-1)) {
    for (j in (i+1):N) {
      Sigma[i,j] <- sigma2 * exp(- pow(l*d[i,j],2))
      Sigma[j,i] <- Sigma[i,j]
    }
  }

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

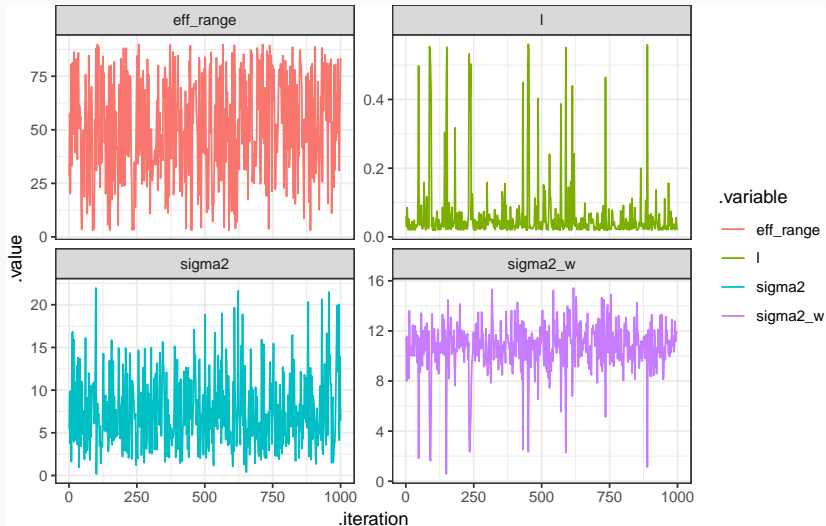
  for (i in 1:3) {
    beta[i] ~ dt(coef[i], 2.5, 1)
  }
  sigma2_w ~ dnorm(5, 1/25) T(0,)
  sigma2 ~ dnorm(12.5, 1/25) T(0,)

  eff_range ~ dunif(3,90)
  l = sqrt(3) / eff_range
}"
```

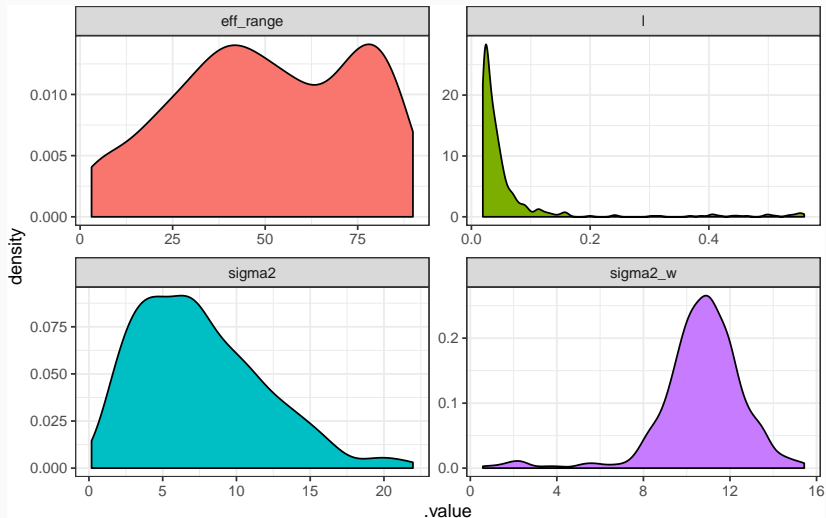
Posterior - Betas



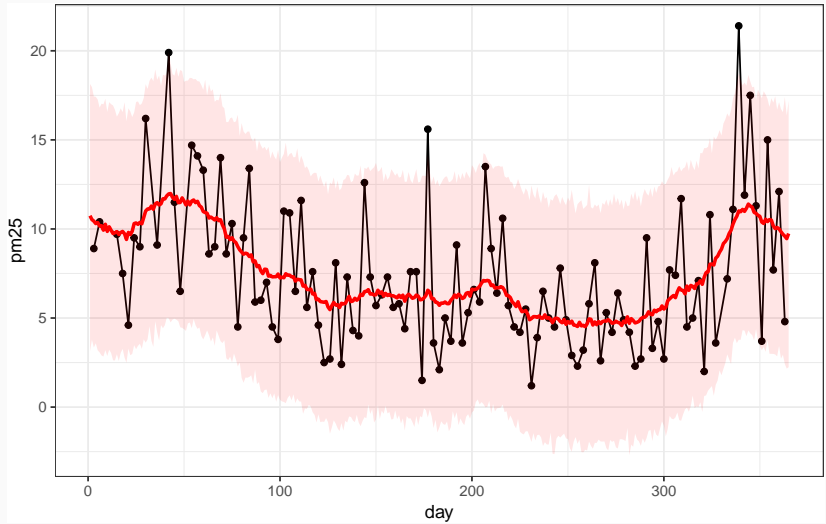
Posterior - Covariance Parameters



Posterior - Covariance Parameters



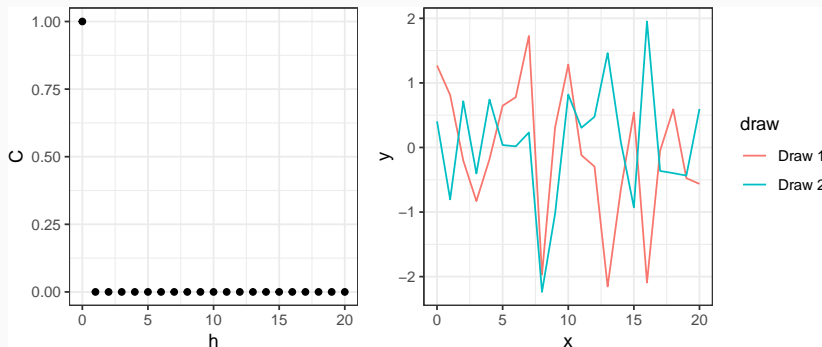
Posterior Predictive Distribution



More Covariance Functions

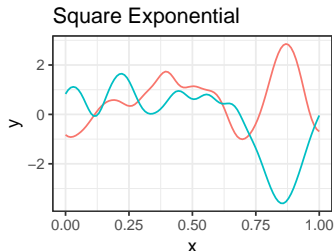
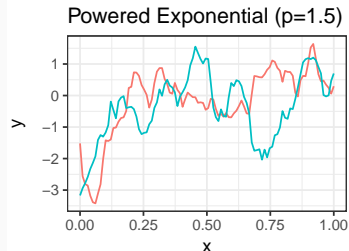
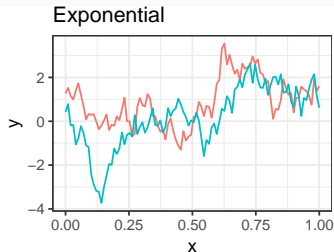
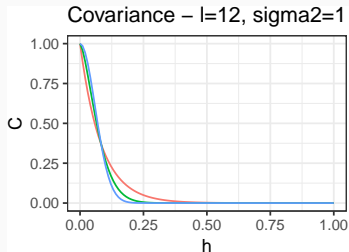
Nugget Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \mathbf{1}_{\{h=0\}} \text{ where } h = |t_i - t_j|$$



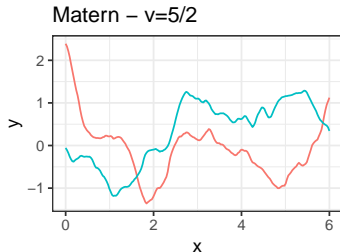
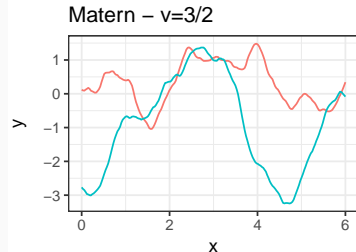
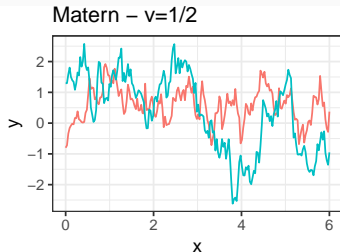
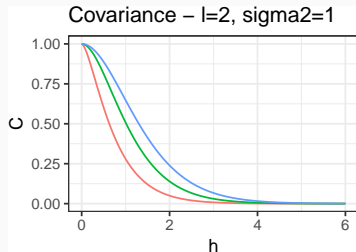
(- / Power / Square) Exponential Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \exp(-(hl)^p) \text{ where } h = |t_i - t_j|$$



Matern Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} h \cdot l\right)^\nu K_\nu \left(\sqrt{2\nu} h \cdot l\right) \text{ where } h = |t_i - t_j|$$



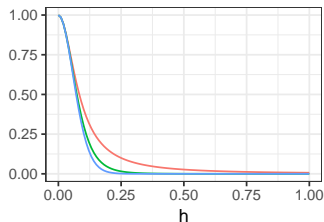
Matern Covariance

- K_ν is the modified Bessel function of the second kind.
- A Gaussian process with Matérn covariance has sample functions that are $\lceil \nu - 1 \rceil$ times differentiable.
- When $\nu = 1/2 + p$ for $p \in \mathbb{N}^+$ then the Matern has a simplified form (product of an exponential and a polynomial of order p).
- When $\nu = 1/2$ the Matern is equivalent to the exponential covariance.
- As $\nu \rightarrow \infty$ the Matern converges to the square exponential covariance.

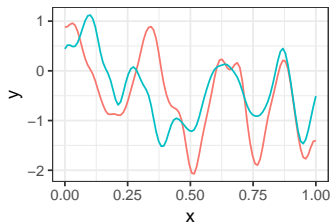
Rational Quadratic Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \left(1 + \frac{h^2 l^2}{\alpha} \right)^{-\alpha} \quad \text{where } h = |t_i - t_j|$$

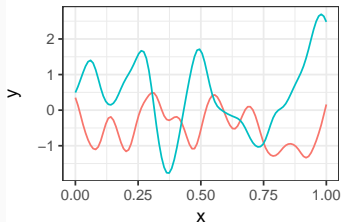
Covariance - l=12, sigma2=1



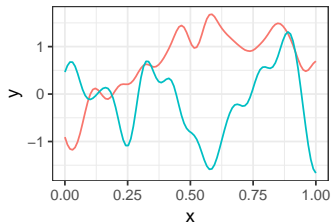
Rational Quadratic - alpha=1



Rational Quadratic - alpha=3



Rational Quadratic - alpha=10



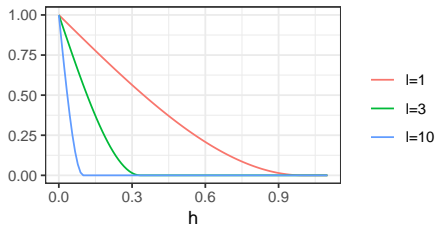
Rational Quadratic Covariance

- is a scaled mixture of squared exponential covariance functions with different characteristic length-scales (l).
- As $\alpha \rightarrow \infty$ the rational quadratic converges to the square exponential covariance.
- Has sample functions that are infinitely differentiable for any value of α

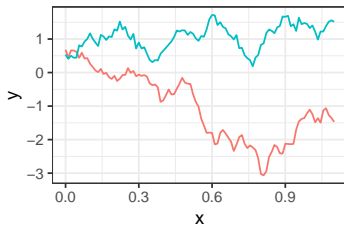
Spherical Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \begin{cases} \sigma^2 \left(1 - \frac{3}{2}h \cdot l + \frac{1}{2}(h \cdot l)^3\right) & \text{if } 0 < h < 1/l \\ 0 & \text{otherwise} \end{cases} \quad \text{where } h = |t_i - t_j|$$

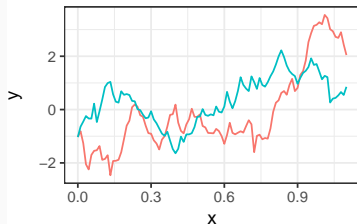
Covariance – sigma2=1



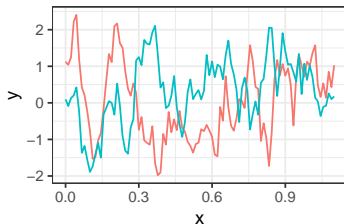
Spherical – l=1



Spherical – l=3



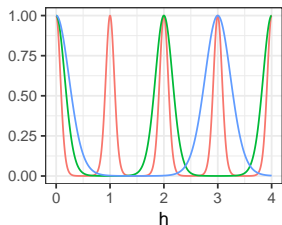
Spherical – l=10



Periodic Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \exp\left(-2l^2 \sin^2\left(\pi \frac{h}{p}\right)\right) \text{ where } h = |t_i - t_j|$$

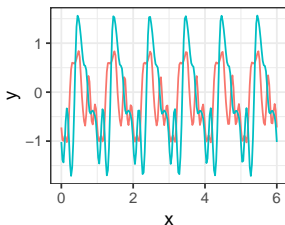
Covariance - l=2, sigma2=1



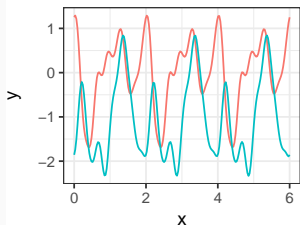
forcats::as_factor(Cov)

- p=1
- p=2
- p=3

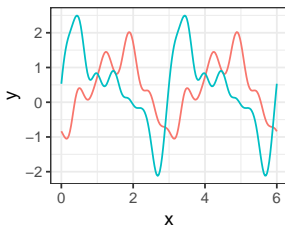
Periodic - p=1



Periodic - p=2

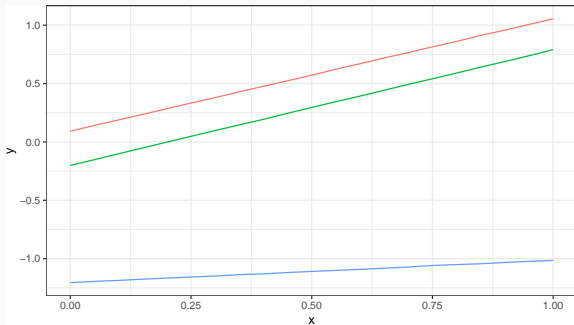


Periodic - p=3



Linear Covariance

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma_b^2 + \sigma_v^2 (t_i - c)(t_j - c)$$



Combining Covariances

If we define two valid covariance functions, $Cov_a(y_{t_i}, y_{t_j})$ and $Cov_b(y_{t_i}, y_{t_j})$ then the following are also valid covariance functions,

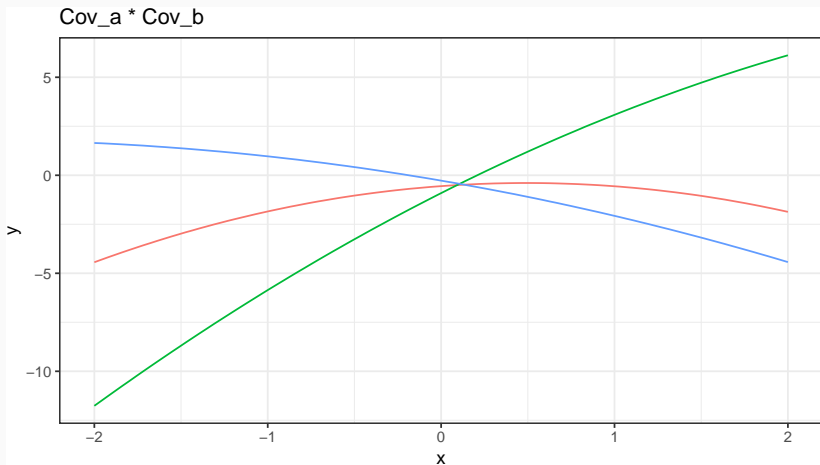
$$Cov_a(y_{t_i}, y_{t_j}) + Cov_b(y_{t_i}, y_{t_j})$$

$$Cov_a(y_{t_i}, y_{t_j}) \times Cov_b(y_{t_i}, y_{t_j})$$

Linear \times Linear \rightarrow Quadratic

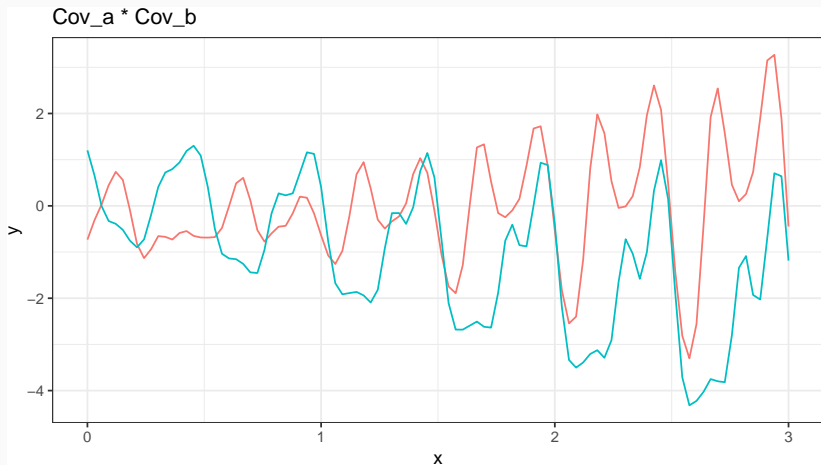
$$\text{Cov}_a(y_{t_i}, y_{t_j}) = 1 + 2(t_i \times t_j)$$

$$\text{Cov}_b(y_{t_i}, y_{t_j}) = 2 + 1(t_i \times t_j)$$



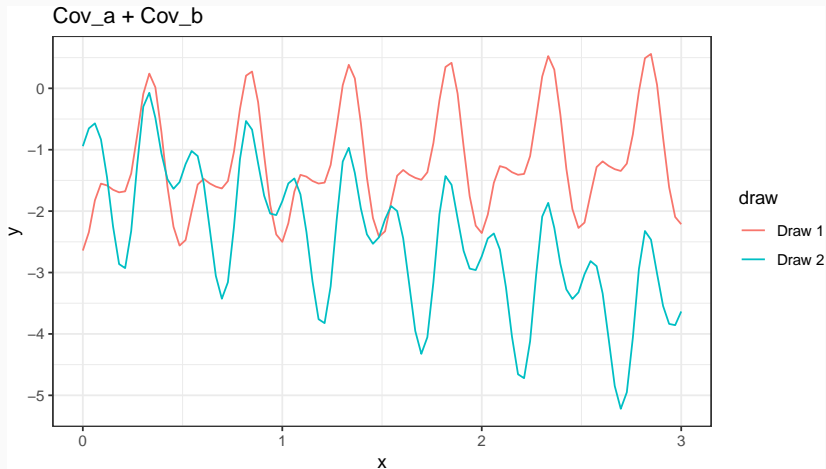
$$\text{Cov}_a(y_{t_i}, y_{t_j}) = 1 + 1 (t_i \times t_j)$$

$$\text{Cov}_b(y_{t_i}, y_{t_j}) = \exp(-2 \sin^2(2\pi h))$$



$$\text{Cov}_a(y_{t_i}, y_{t_j}) = 1 + 1 (t_i \times t_j)$$

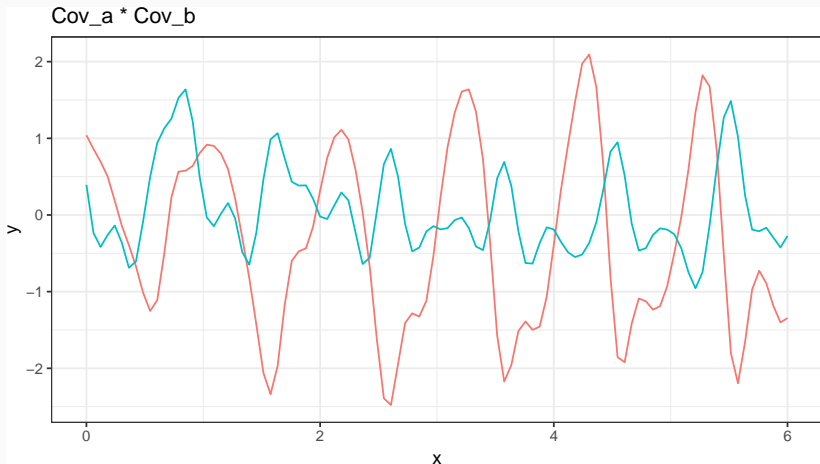
$$\text{Cov}_b(h = |t_i - t_j|) = \exp(-2 \sin^2(2\pi h))$$



Sq Exp \times Periodic \rightarrow Locally Periodic

$$\text{Cov}_a(h = |t_i - t_j|) = \exp(-(1/3)h^2)$$

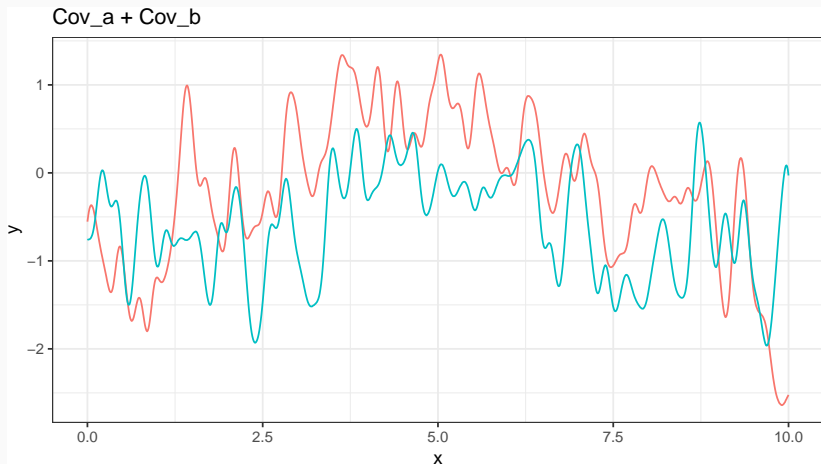
$$\text{Cov}_b(h = |t_i - t_j|) = \exp(-2 \sin^2(\pi h))$$



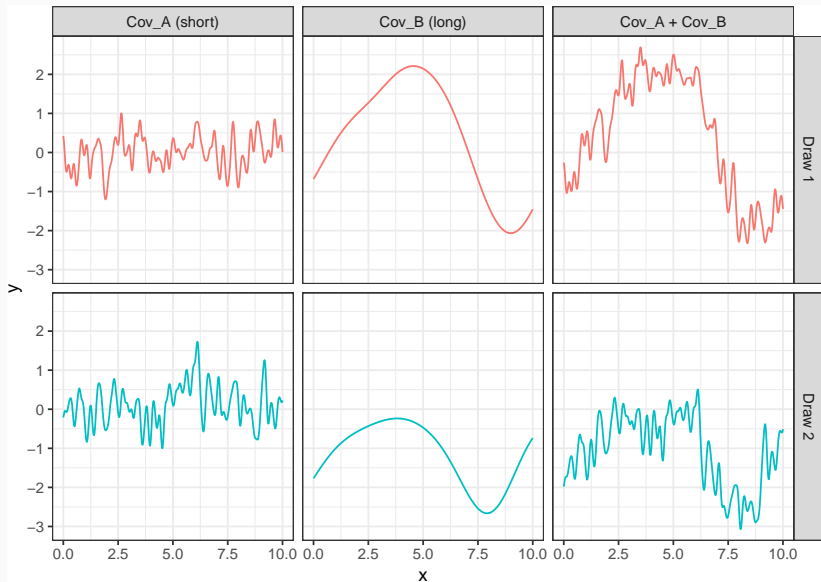
Sq Exp (short) + Sq Exp (long)

$$\text{Cov}_a(h = |t_i - t_j|) = (1/4) \exp(-4\sqrt{3}h^2)$$

$$\text{Cov}_b(h = |t_i - t_j|) = \exp(-(\sqrt{3}/2)h^2)$$



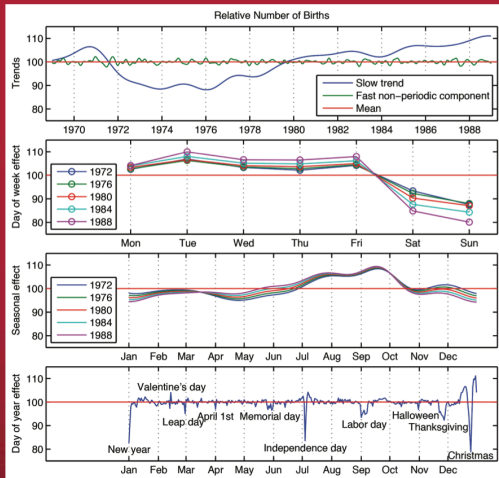
Sq Exp (short) + Sq Exp (long) (Seen another way)



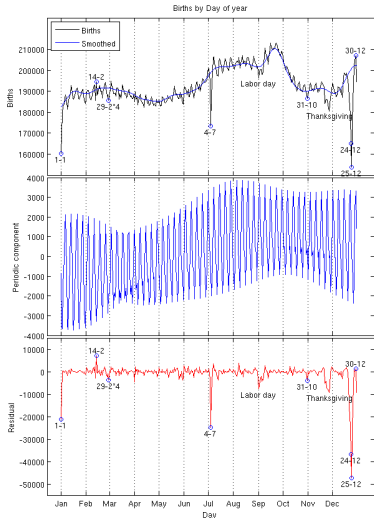
BDA3 example

Bayesian Data Analysis

Third Edition



Births (one year)



1. Smooth long term trend
($sq \exp cov$)
2. Seven day periodic trend with decay
($periodic \times sq \exp cov$)
3. Constant mean

Component Contributions

We can view our GP in the following ways,

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \sigma^2 \mathbf{I})$$

but with appropriate conditioning we can also think of \mathbf{y} as being the sum of multiple independent GPs

$$\mathbf{y} = \boldsymbol{\mu} + w_1(\mathbf{t}) + w_2(\mathbf{t}) + w_3(\mathbf{t})$$

where

$$w_1(\mathbf{t}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_1)$$

$$w_2(\mathbf{t}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_2)$$

$$w_3(\mathbf{t}) \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$$

Decomposition of Covariance Components

$$\begin{bmatrix} y \\ w_1 \\ w_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_1 + \Sigma_2 + \sigma^2 \mathbf{I} & \Sigma_1 & \Sigma_2 \\ & \Sigma_1 & 0 \\ & & \Sigma_2 \end{bmatrix} \right)$$

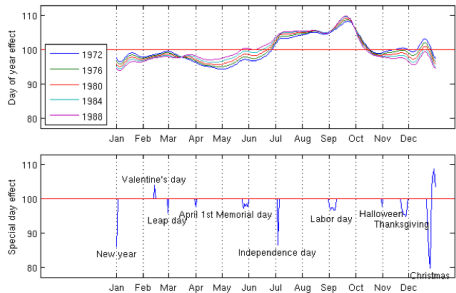
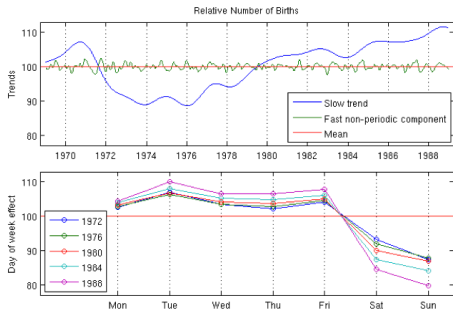
therefore

$$w_1 \mid \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{cond}, \boldsymbol{\Sigma}_{cond})$$

$$\boldsymbol{\mu}_{cond} = 0 + \Sigma_1 (\Sigma_1 + \Sigma_2 + \sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

$$\boldsymbol{\Sigma}_{cond} = \Sigma_1 - \Sigma_1 (\Sigma_1 + \Sigma_2 + \sigma^2 \mathbf{I})^{-1} \Sigma_1^t$$

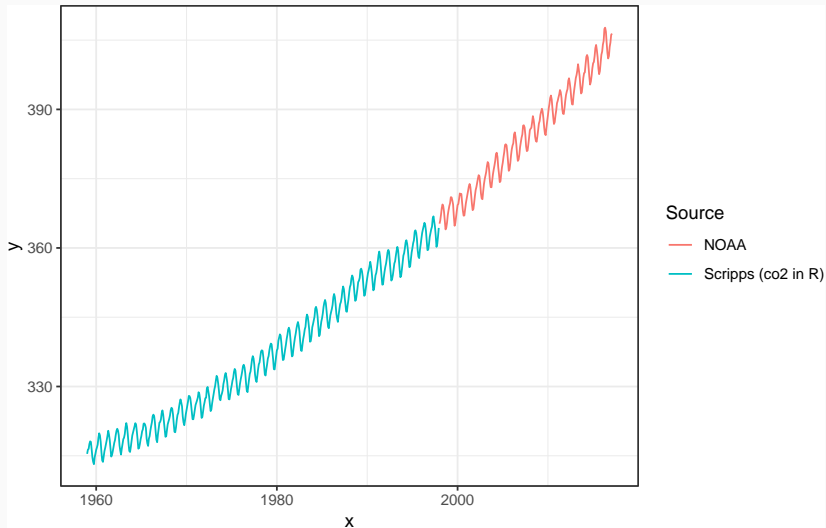
Births (multiple years)



1. slowly changing trend ($sq \exp cov$)
2. small time scale correlating noise ($sq \exp cov$)
3. 7 day periodical component capturing day of week effect ($periodic \times sq \exp cov$)
4. 365.25 day periodical component capturing day of year effect ($periodic \times sq \exp cov$)
5. component to take into account the special days and interaction with weekends ($linear cov$)
6. independent Gaussian noise ($nugget cov$)
7. constant mean

Mauna Loa Example

Atmospheric CO₂



Based on Rasmussen 5.4.3 (we are using slightly different data and parameterization)

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_3 + \boldsymbol{\Sigma}_4 + \sigma^2 I)$$

$$\{\boldsymbol{\mu}\}_i = \bar{y}$$

$$\{\boldsymbol{\Sigma}_1\}_{ij} = \sigma_1^2 \exp(-(l_1 \cdot d_{ij})^2)$$

$$\{\boldsymbol{\Sigma}_2\}_{ij} = \sigma_2^2 \exp(-(l_2 \cdot d_{ij})^2) \exp(-2(l_3)^2 \sin^2(\pi d_{ij}/p))$$

$$\{\boldsymbol{\Sigma}_3\}_{ij} = \sigma_3^2 \left(1 + \frac{(l_4 \cdot d_{ij})^2}{\alpha}\right)^{-\alpha}$$

$$\{\boldsymbol{\Sigma}_4\}_{ij} = \sigma_4^2 \exp(-(l_5 \cdot d_{ij})^2)$$

```

ml_model = "model{
  y ~ dnorm(mu, inverse(Sigma))

  for (i in 1:(length(y)-1)) {
    for (j in (i+1):length(y)) {
      k1[i,j] <- sigma2[1] * exp(- pow(l[1] * d[i,j],2))
      k2[i,j] <- sigma2[2] * exp(- pow(l[2] * d[i,j],2) - 2 * pow(l[3] * sin(pi*d[i,j]
      k3[i,j] <- sigma2[3] * pow(1+pow(l[4] * d[i,j],2)/alpha, -alpha)
      k4[i,j] <- sigma2[4] * exp(- pow(l[5] * d[i,j],2))

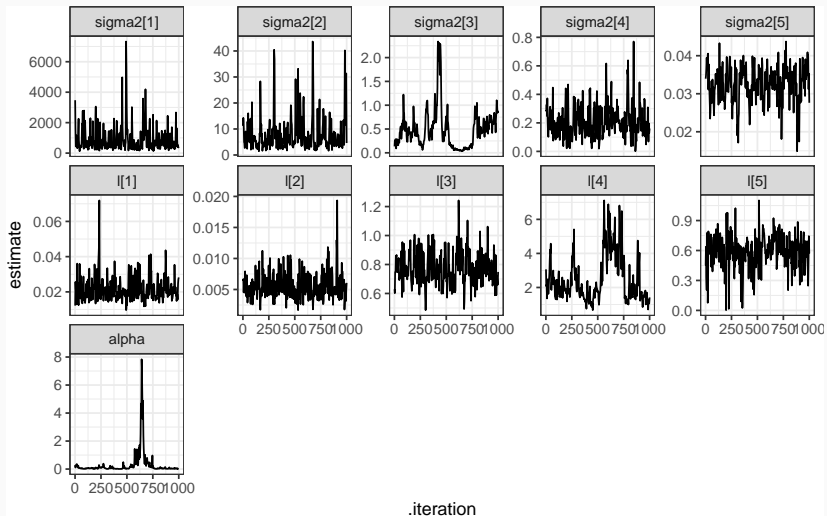
      Sigma[i,j] <- k1[i,j] + k2[i,j] + k3[i,j] + k4[i,j]
      Sigma[j,i] <- Sigma[i,j]
    }
  }

  for (i in 1:length(y)) {
    Sigma[i,i] <- sigma2[1] + sigma2[2] + sigma2[3] + sigma2[4] + sigma2[5]
  }

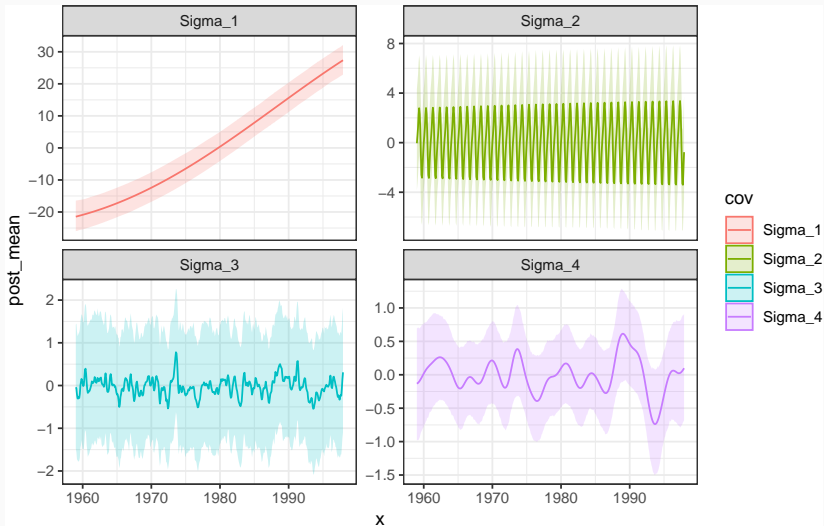
  for(i in 1:5){
    sigma2[i] ~ dt(0, 2.5, 1) T(0,)
    l[i] ~ dt(0, 2.5, 1) T(0,)
  }
  alpha ~ dt(0, 2.5, 1) T(0,)
}"

```

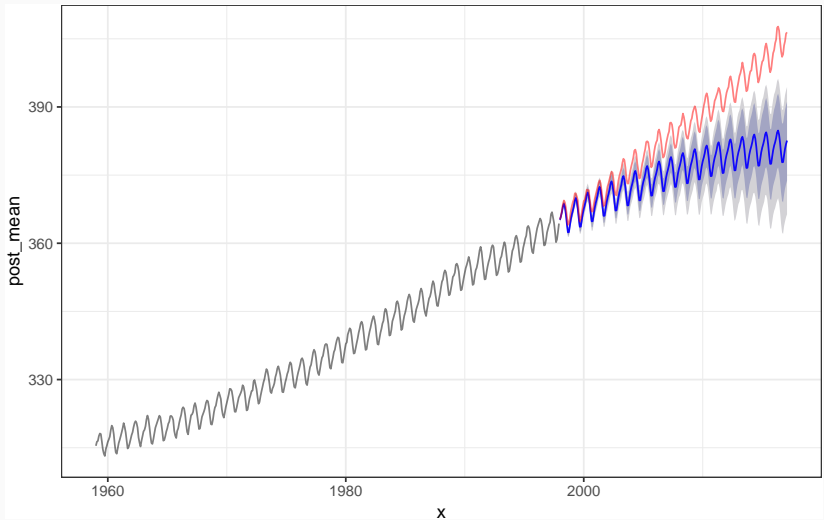
Diagnostics



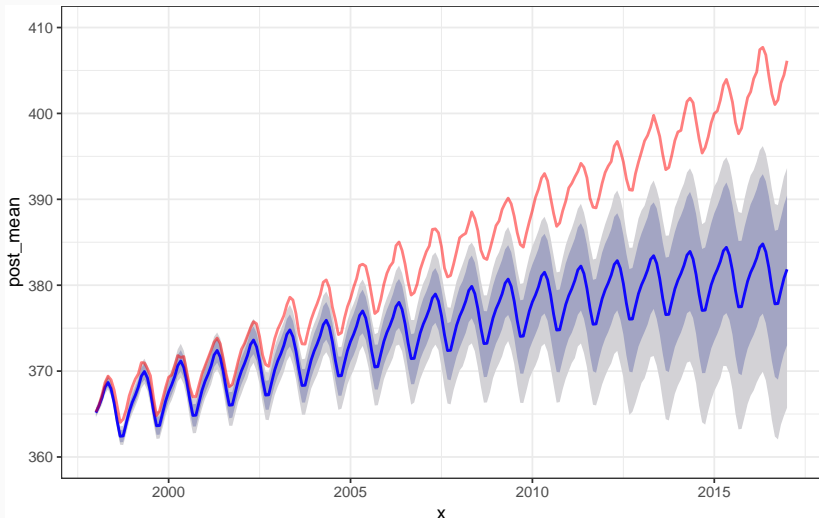
Fit Components



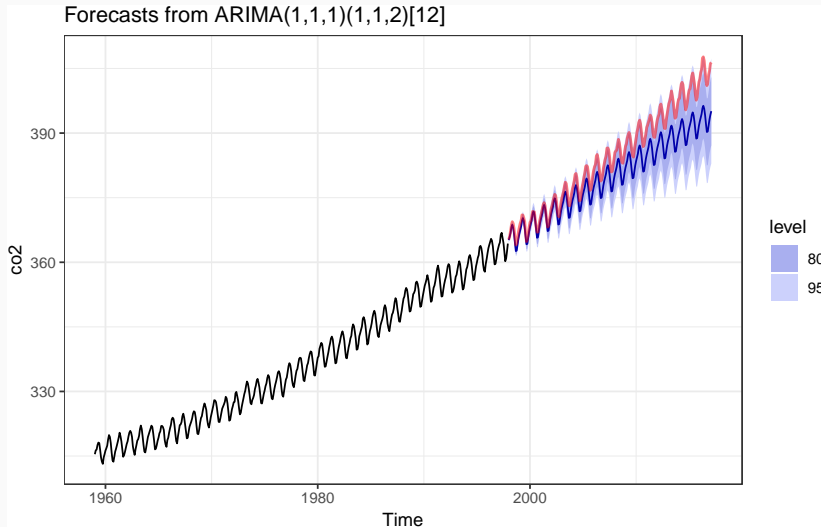
Forecasting



Forecasting (zoom)



Forecasting ARIMA (auto)



dates	RMSE (arima)	RMSE (gp)
Jan 1998 - Jan 2003	1.103	1.911
Jan 1998 - Jan 2008	2.506	4.575
Jan 1998 - Jan 2013	3.824	7.706
Jan 1998 - Mar 2017	5.461	11.395

Forecasting Components

