Lecture 14

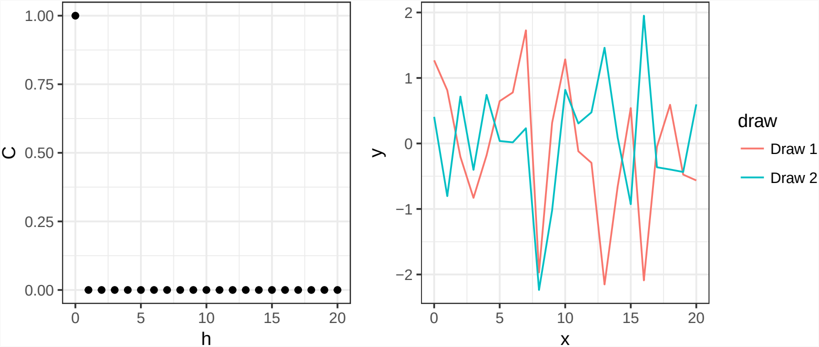
Covariance Functions

3/08/2018

More on Covariance Functions

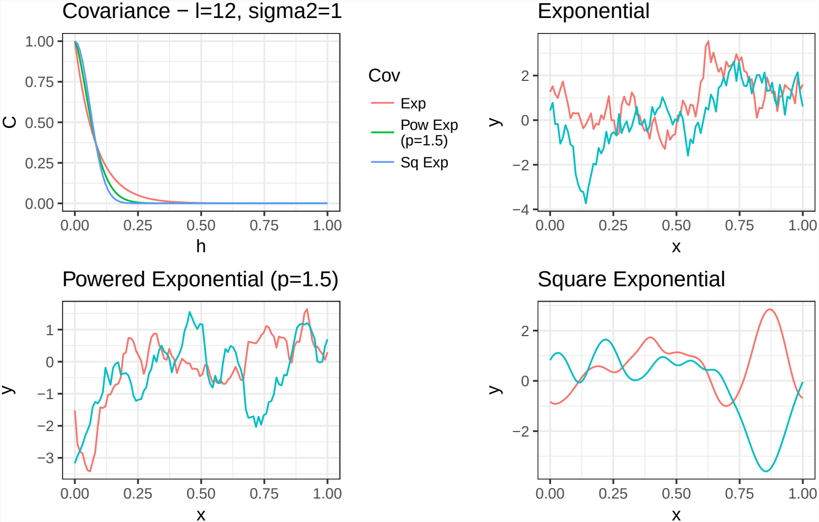
Nugget Covariance

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



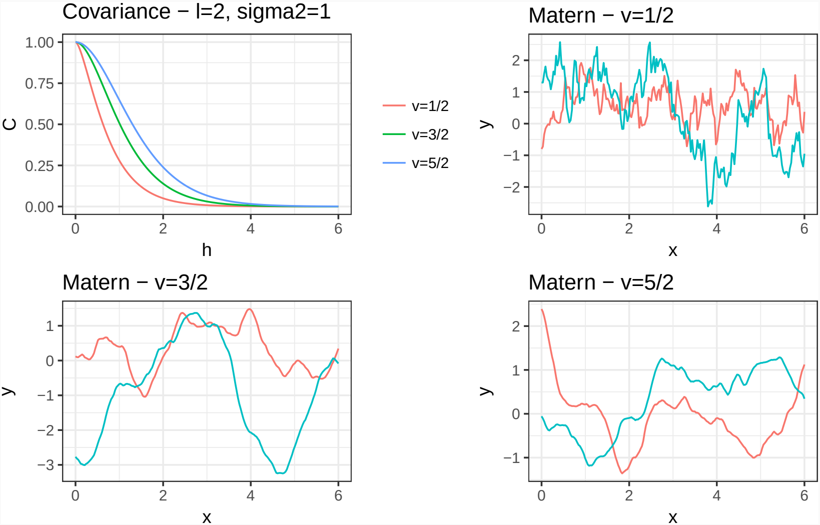
(- / Power / Square) Exponential Covariance

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



Matern Covariance

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

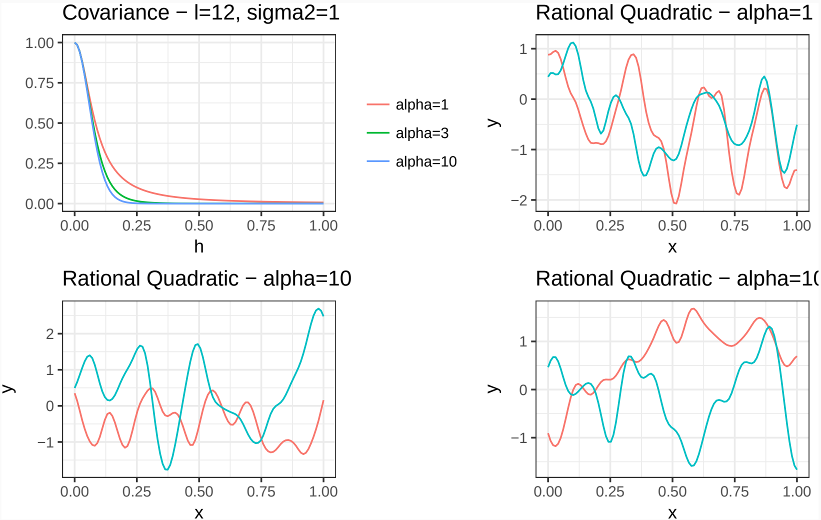


Matern Covariance

- 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 $\nu \to \infty$ the Matern converges to the square exponential covariance.
- A Gaussian process with Matérn covariance has paths that are $\lceil \nu \rceil - 1$ times differentiable.

Rational Quadratic Covariance

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

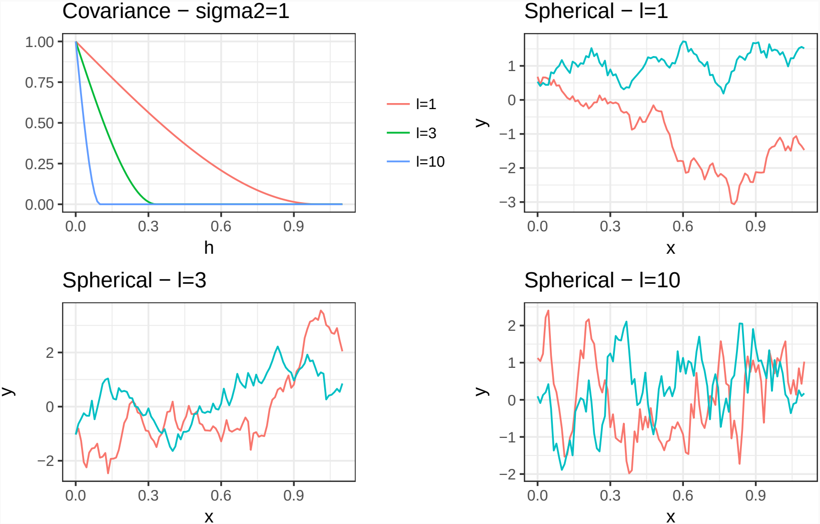


Rational Quadratic Covariance

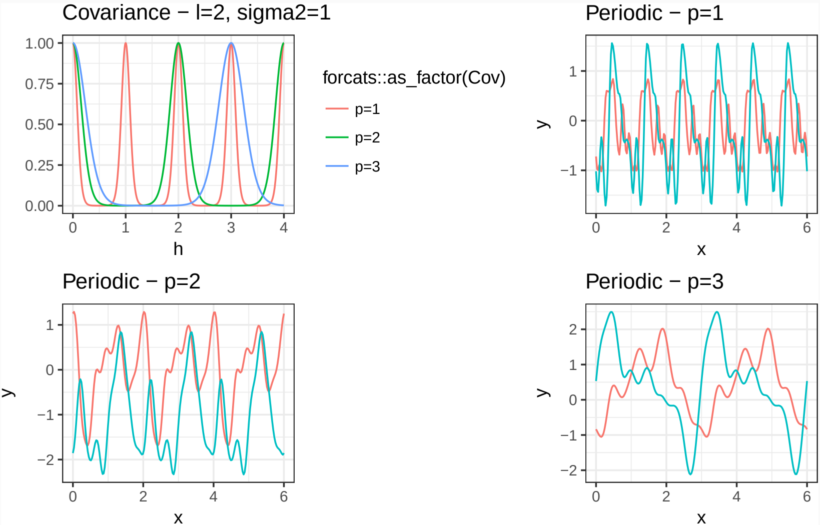
- is a scaled mixture of squared exponential covariance functions with different characteristic length-scales (l).
- As $\alpha \to \infty$ the rational quadratic converges to the square exponential covariance.
- Has sample functions that are infinitely differentiable for any value of α

Spherical Covariance

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

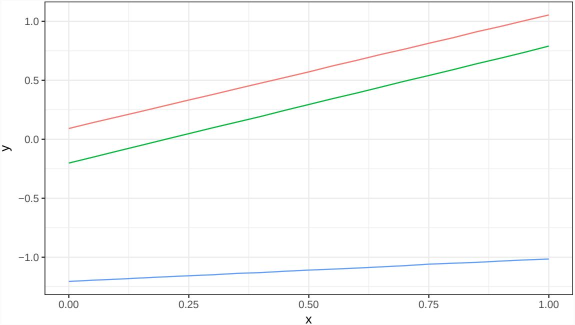


$$Cov(y_{t_i}, y_{t_j}) = \sigma^2 \exp\left(-2\, l^2 \sin^2\left(\pi \frac{h}{p}\right)\right) \text{ where } h = |t_i - t_j|$$



Linear Covariance

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



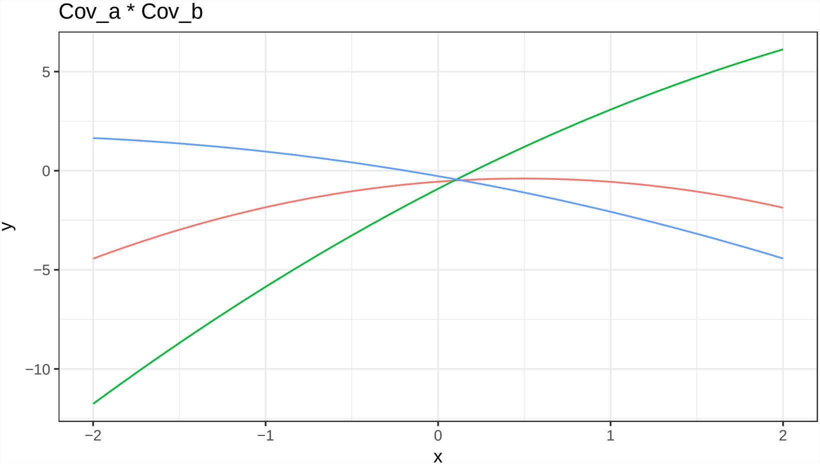
Combining Covariances

If we definite 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,

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

$$Cov_a(y_{t_i}, y_{t_j}) = 1 + 2\,(t_i \times t_j)$$

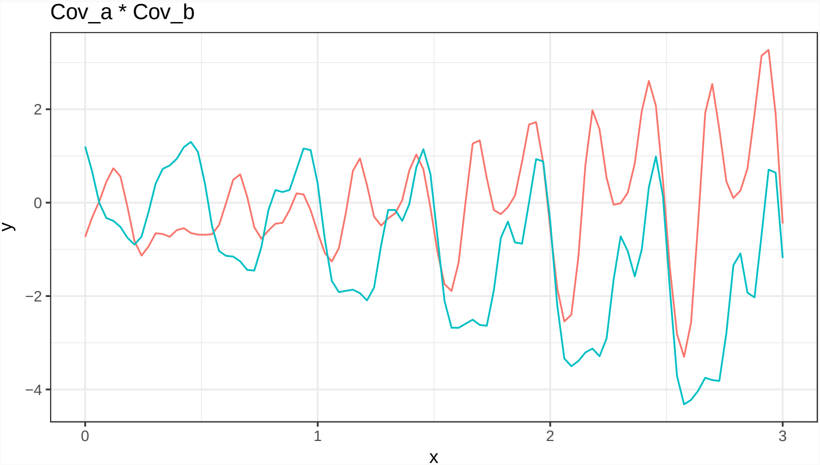
$$Cov_b(y_{t_i}, y_{t_j}) = 2 + 1\,(t_i \times t_j)$$



Linear × Periodic

$$Cov_a(y_{t_i}, y_{t_j}) = 1 + 1\,(t_i \times t_j)$$

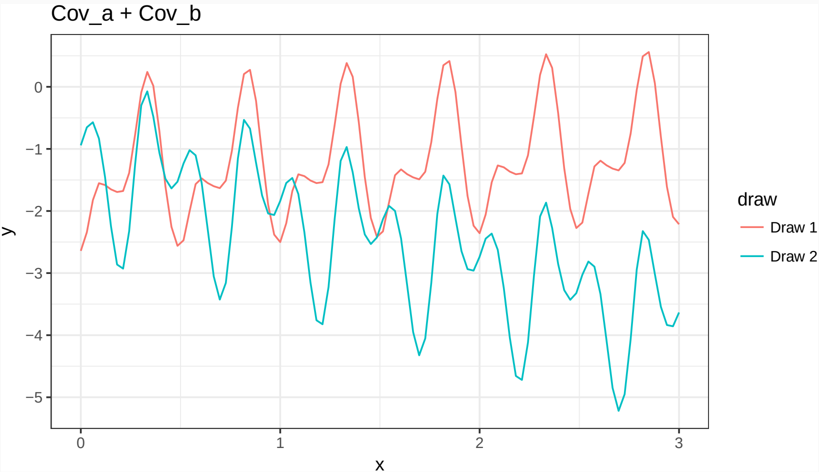
$$Cov_b(y_{t_i}, y_{t_j}) = \exp\left(-2\,\sin^2\,(2\pi\,h)\right)$$



Linear + Periodic

$$Cov_a(y_{t_i}, y_{t_j}) = 1 + 1\,(t_i \times t_j)$$

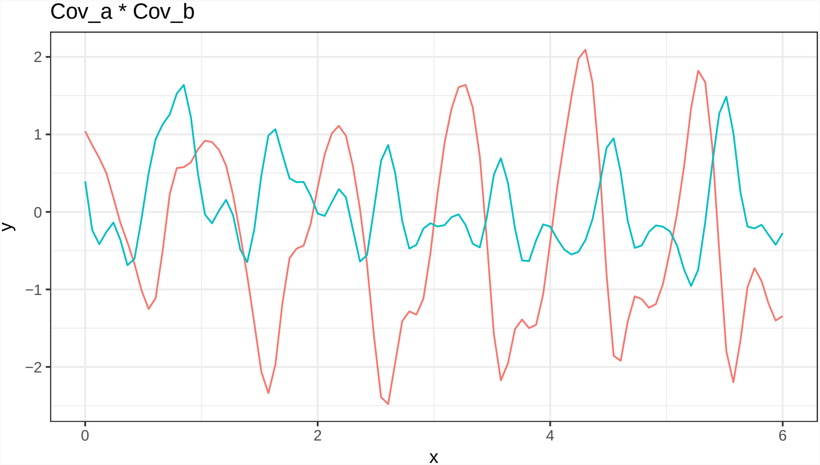
$$Cov_b(h = |t_i - t_j|) = \exp\left(-2\,\sin^2\left(2\pi\,h\right)\right)$$



Sq Exp $\times$ Periodic $\longrightarrow$ Locally Periodic

$$Cov_a(h = |t_i - t_j|) = \exp(-(1/3)h^2)$$

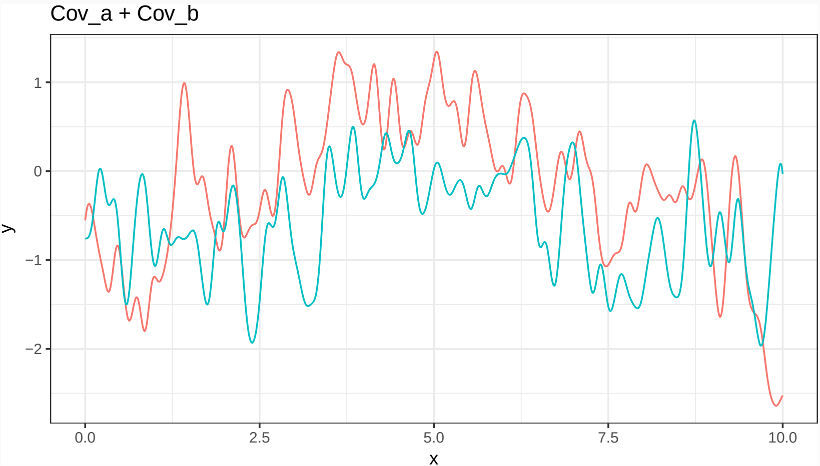
$$Cov_b(h = |t_i - t_j|) = \exp\left(-2\,\sin^2\left(\pi\, h\right)\right)$$



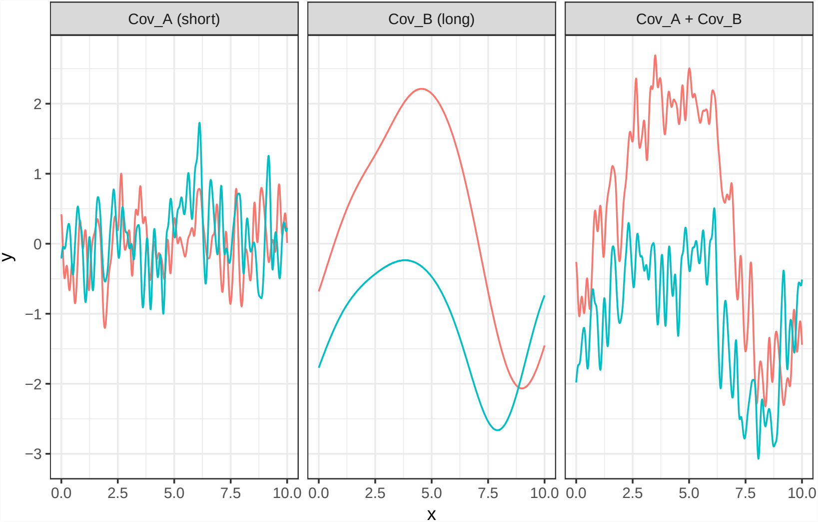
Sq Exp (short) + Sq Exp (long)

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

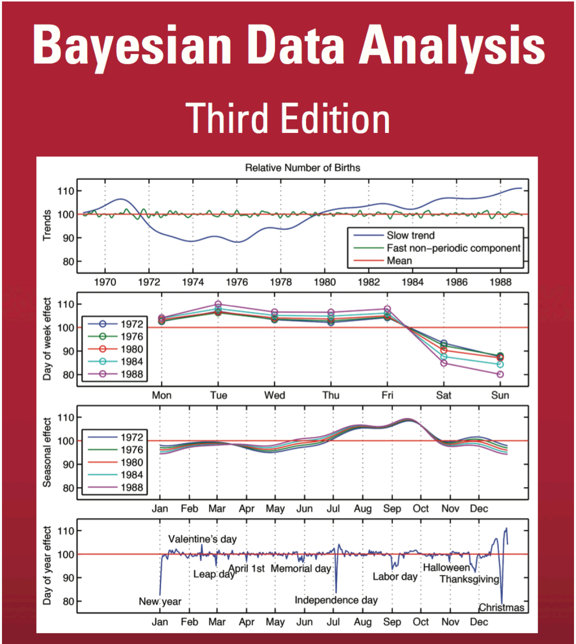
$$Cov_b(h = |t_i - t_j|) = \exp(-(\sqrt{3}/2)h^2)$$



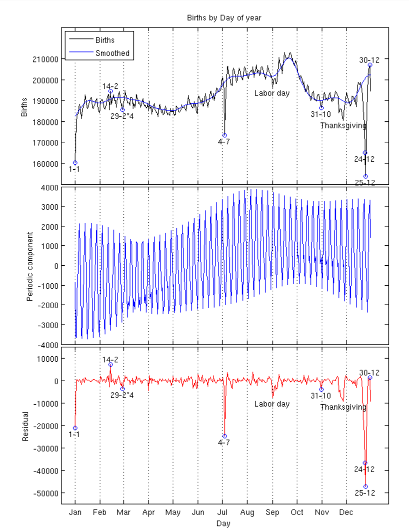
Sq Exp (short) + Sq Exp (long) (Seen another way)



BDA3 example


Bayesian Data Analysis
Third Edition
Relative Number of Births
Trends
Slow trend
Fast non−periodic component
Mean
1970 1972 1974 1976 1978 1980 1982 1984 1986 1988
Day of week effect
1972 1976 1980 1984 1988
Mon Tue Wed Thu Fri Sat Sun
Seasonal effect
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Day of year effect
Valentine's day
Leap day
April 1st
Memorial day
Labor day
Halloween
Thanksgiving
New year
Independence day
Christmas

Births (one year)



1. Smooth long term trend (*sq exp cov*)
2. Seven day periodic trend with decay (*periodic* × *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 multipe independent GPs

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

where

$$\begin{aligned} w_1(\mathbf{t}) &\sim \mathcal{N}(0, \boldsymbol{\Sigma}_1\,) \\ w_2(\mathbf{t}) &\sim \mathcal{N}(0, \boldsymbol{\Sigma}_2) \\ w_3(\mathbf{t}) &\sim \mathcal{N}(0, \sigma^2 \mathbf{I}\,) \end{aligned}$$

Decomposition of Covariance Components

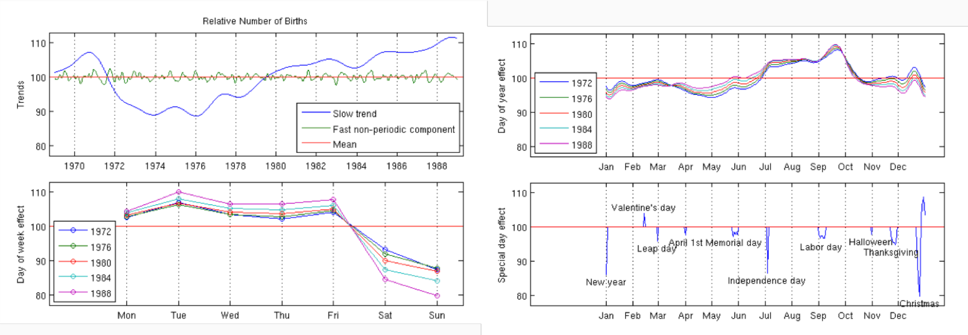
$$\begin{bmatrix} y \\ w_1 \\ w_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_1 + \Sigma_2 + \sigma^2 \mathbf{I} & \Sigma_1 & \Sigma_2 \\ \Sigma_1 & \Sigma_1 & 0 \\ \Sigma_2 & 0 & \Sigma_2 \end{bmatrix} \right)$$

therefore

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

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

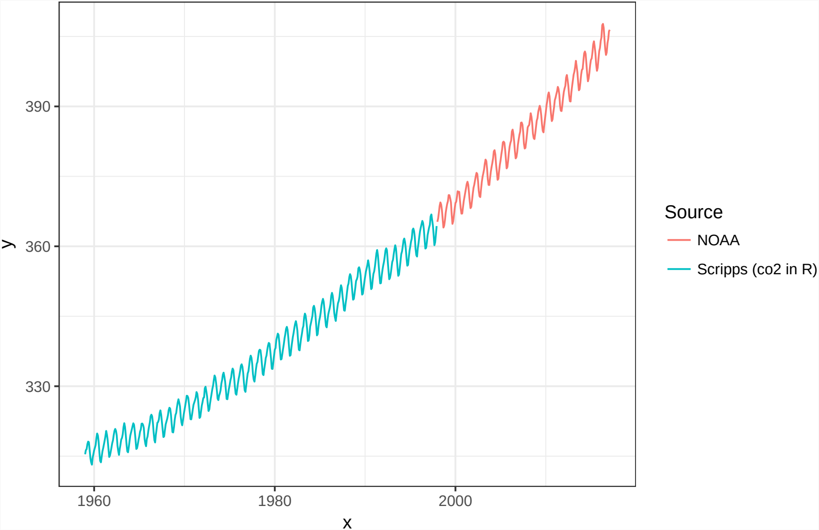
Births (multiple years)



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* × *sq exp cov*)
4. 365.25 day periodical component capturing day of year effect (*periodic* × *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 Exampel

Atmospheric CO_2



GP Model

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

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

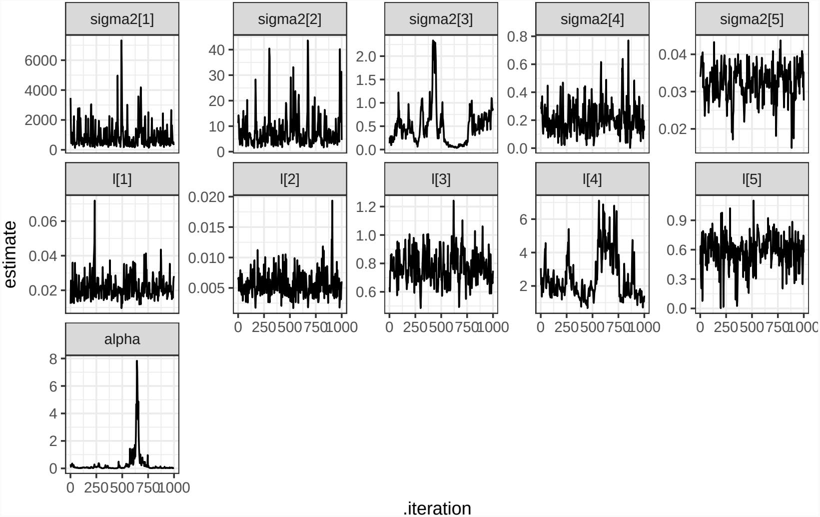
$$\{\boldsymbol{\mu}\}_i = \bar{y}$$

$$\begin{aligned}
\{\boldsymbol{\Sigma}_1\}_{ij} &= \sigma_1^2 \exp\left(-(l_1 \cdot d_{ij})^2\right) \\
\{\boldsymbol{\Sigma}_2\}_{ij} &= \sigma_2^2 \exp\left(-(l_2 \cdot d_{ij})^2\right) \exp\left(-2\,(l_3)^2 \sin^2(\pi\, d_{ij}/p)\right) \\
\{\boldsymbol{\Sigma}_3\}_{ij} &= \sigma_3^2 \left(1 + \frac{(l_4 \cdot d_{ij})^2}{\alpha}\right)^{-\alpha} \\
\{\boldsymbol{\Sigma}_4\}_{ij} &= \sigma_4^2 \exp\left(-(l_5 \cdot d_{ij})^2\right)
\end{aligned}$$

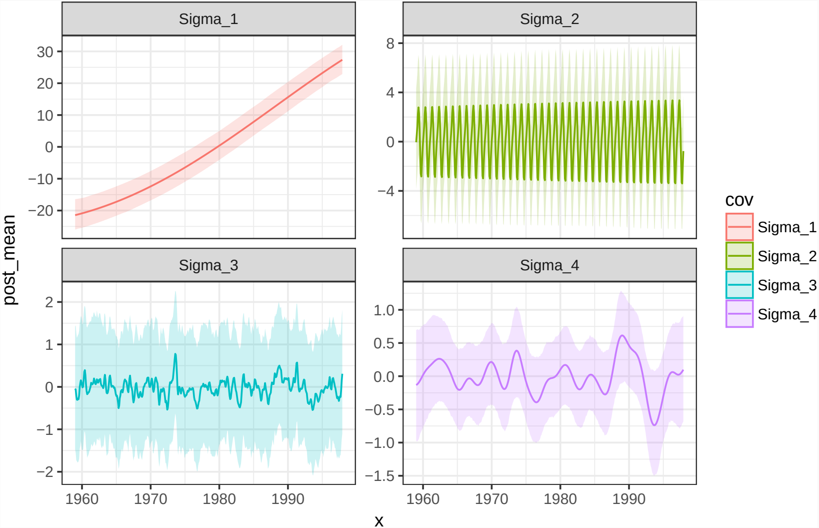
JAGS Model

```
ml_model = "model{
  y ~ dmnorm(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,)
}"
```

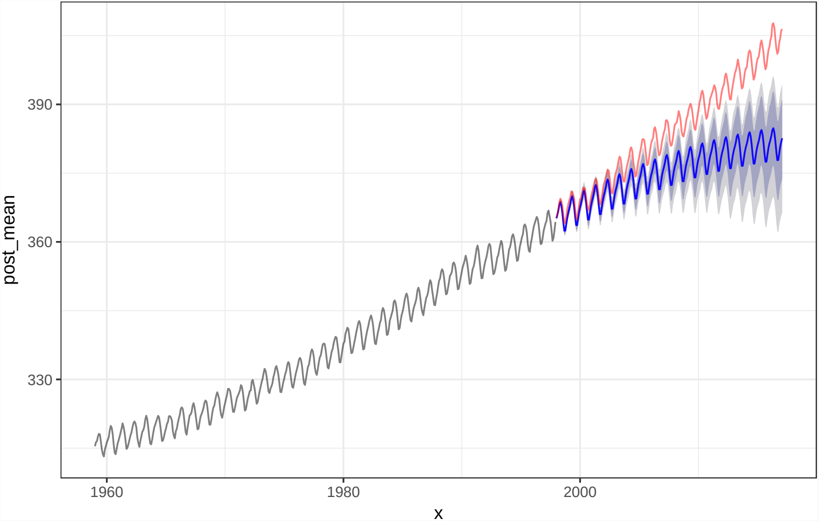
Diagnostics



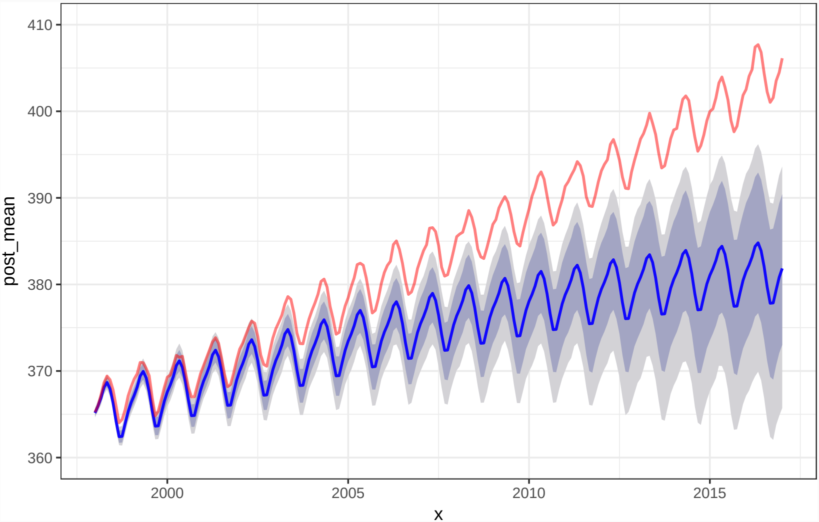
Fit Components



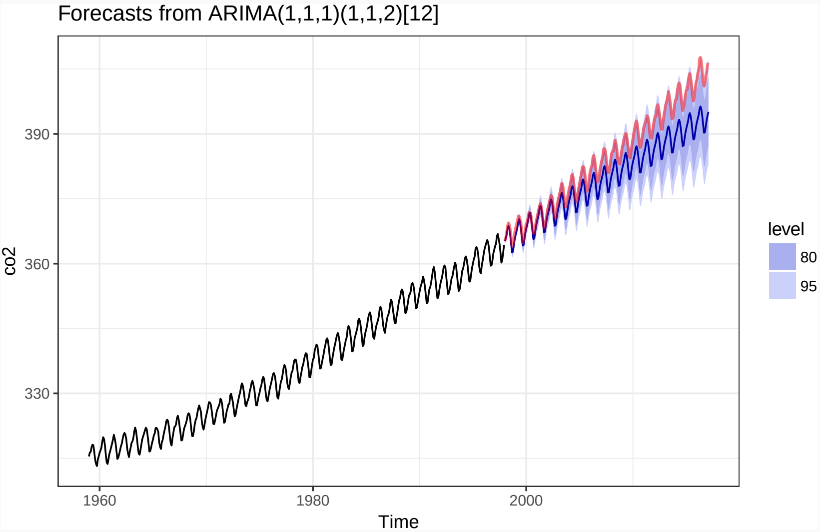
Forecasting



Forecasting (zoom)



Forecasting ARIMA (auto)



Forecasting RMSE

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