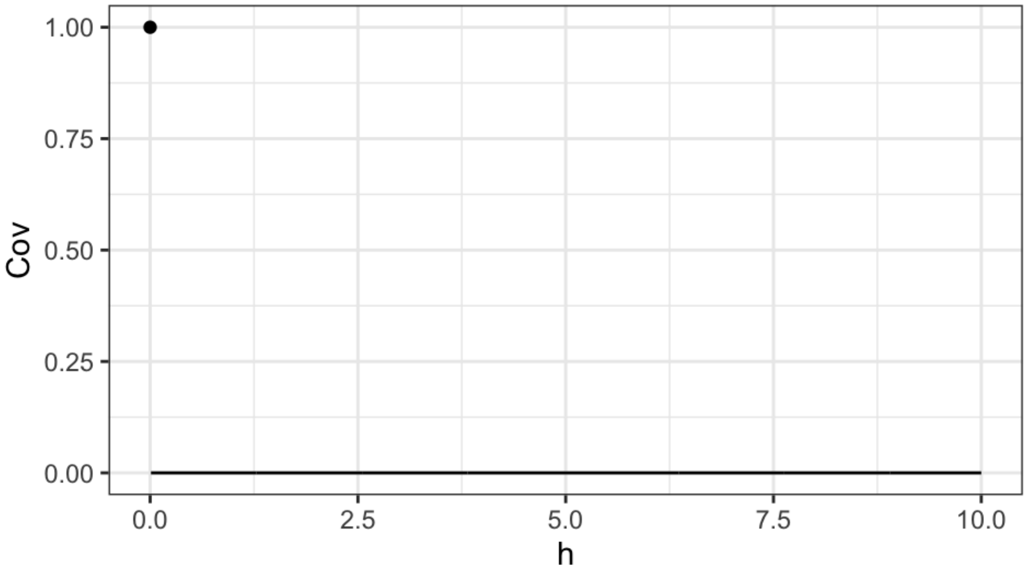
Covariance Functions

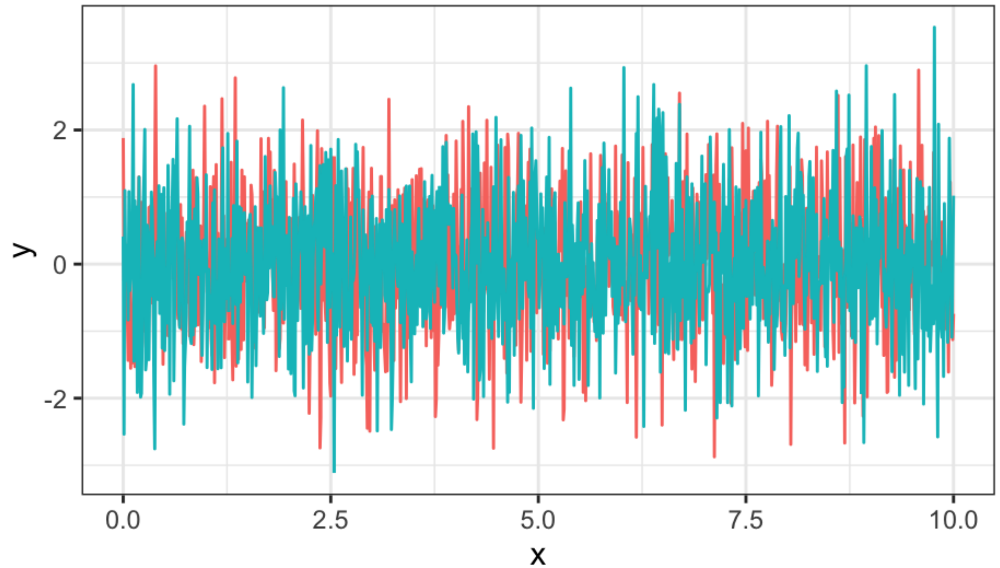
Lecture 15

Dr. Colin Rundel

Nugget Covariance

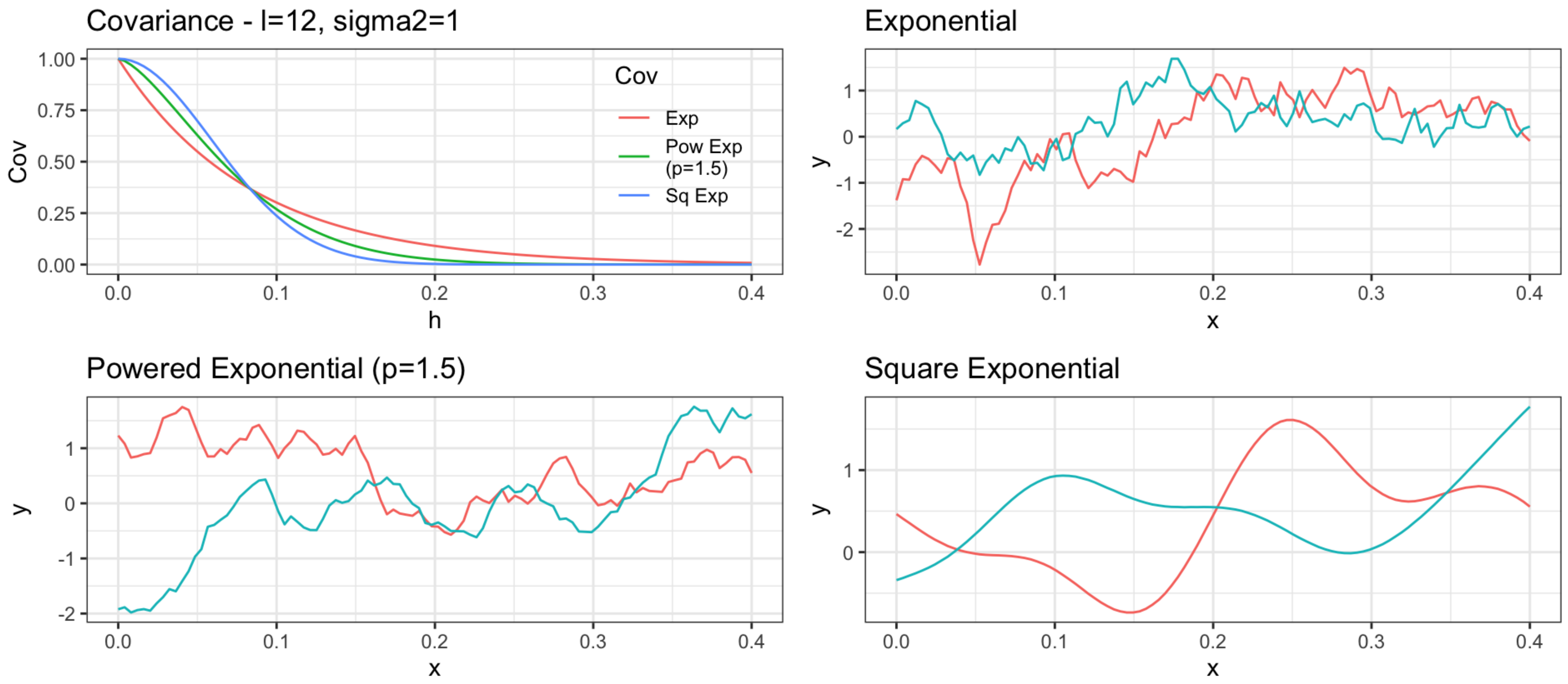
$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 1_{\{h=0\}} \;\; \text{where } h = |t_i - t_j|$$





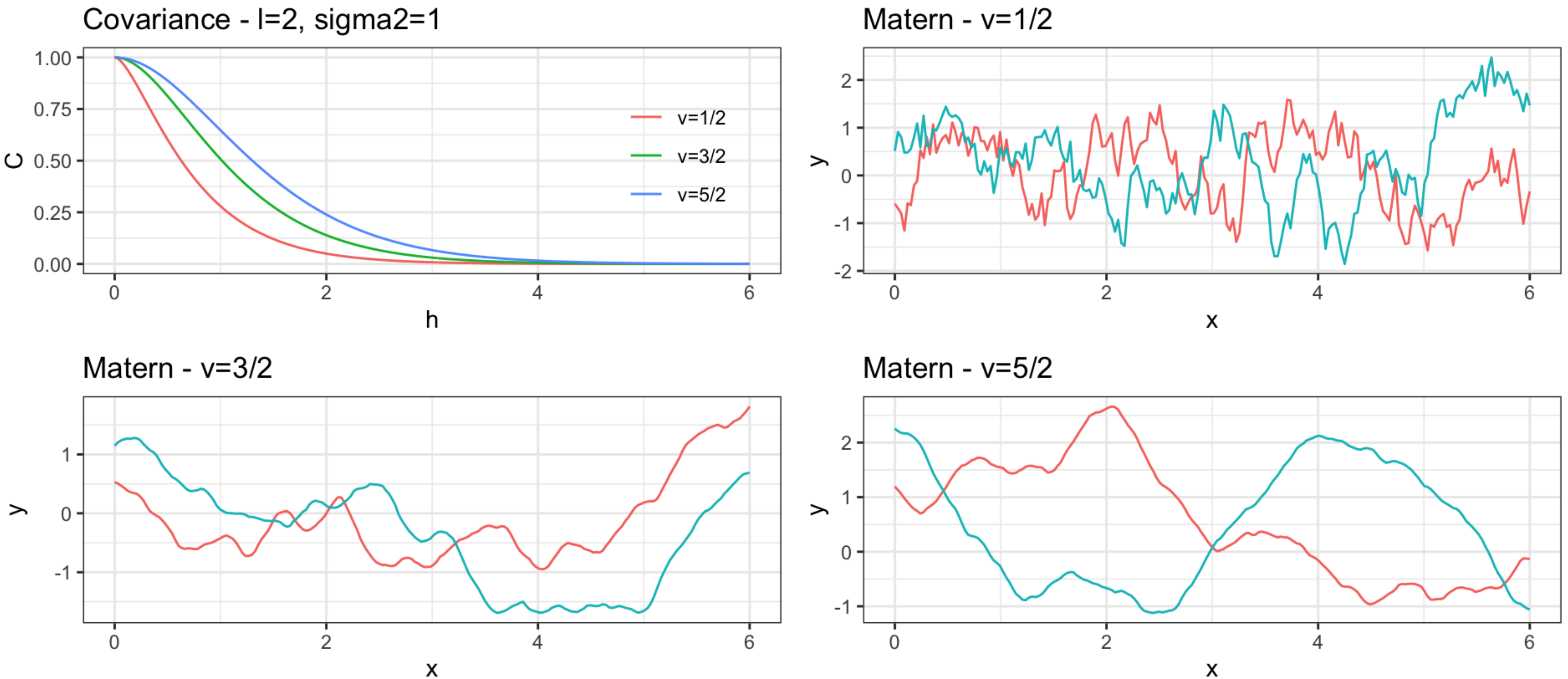
(- / Powered / Square) Exponential Covariance

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



Matern Covariance

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



Matern Covariance

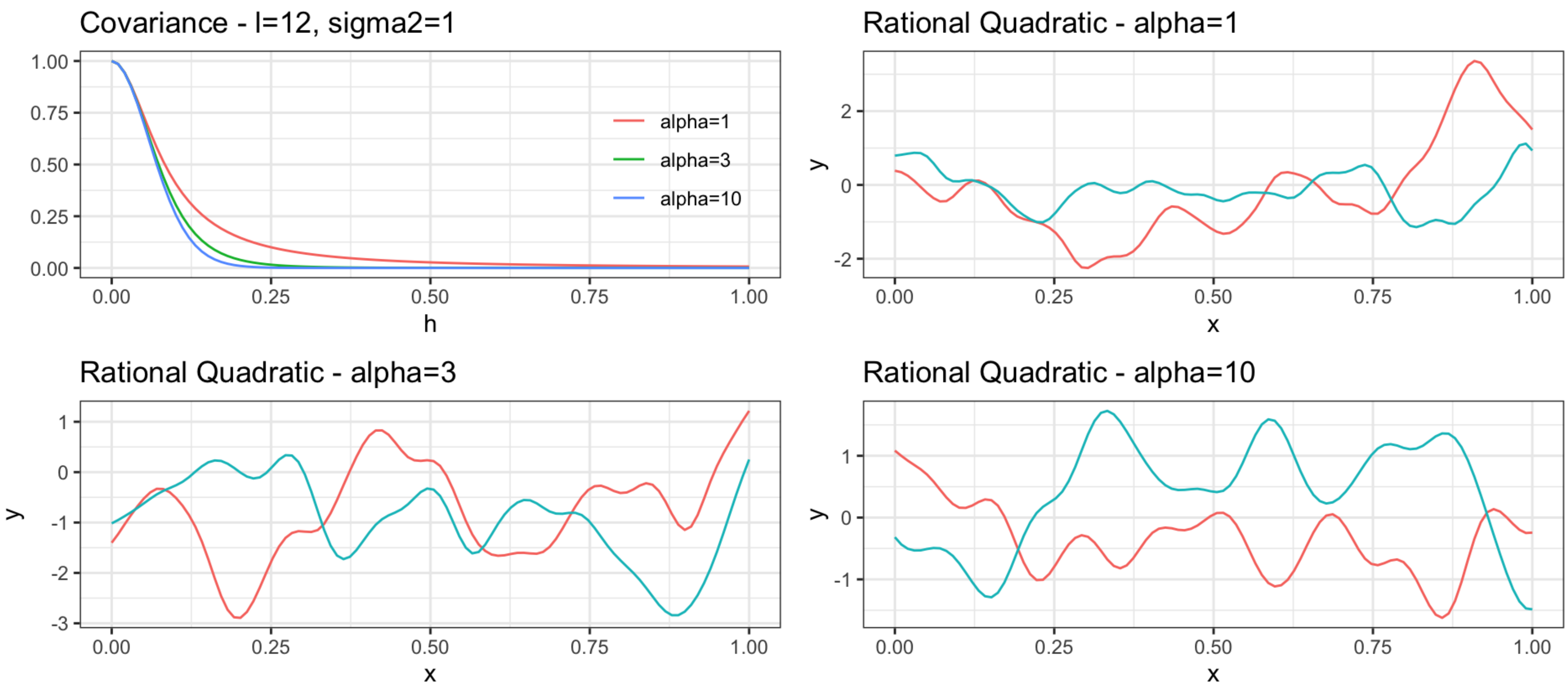
- $K_\nu()$ is the modified Bessel function of the second kind.
- A Gaussian process with Matérn covariance has sample functions that are $\lceil \nu - 1 \rceil$ times differentiable.
- When $\nu = 1/2 + p$ for $p \in \mathbb{N}^+$ then the Matern has a simplified form

$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \exp\left(-\sqrt{2p+1}\; h \cdot l\right) \frac{p!}{(2p)!} \sum_{i=0}^{p} \frac{(p+i)!}{i!(p-i)!} \left(2\sqrt{2p+1}\; h \cdot l\right)^{p-i}$$

- When $\nu = 1/2$ the Matern is equivalent to the exponential covariance.
- As $\nu \to \infty$ the Matern converges to the squared exponential covariance.

Rational Quadratic Covariance

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

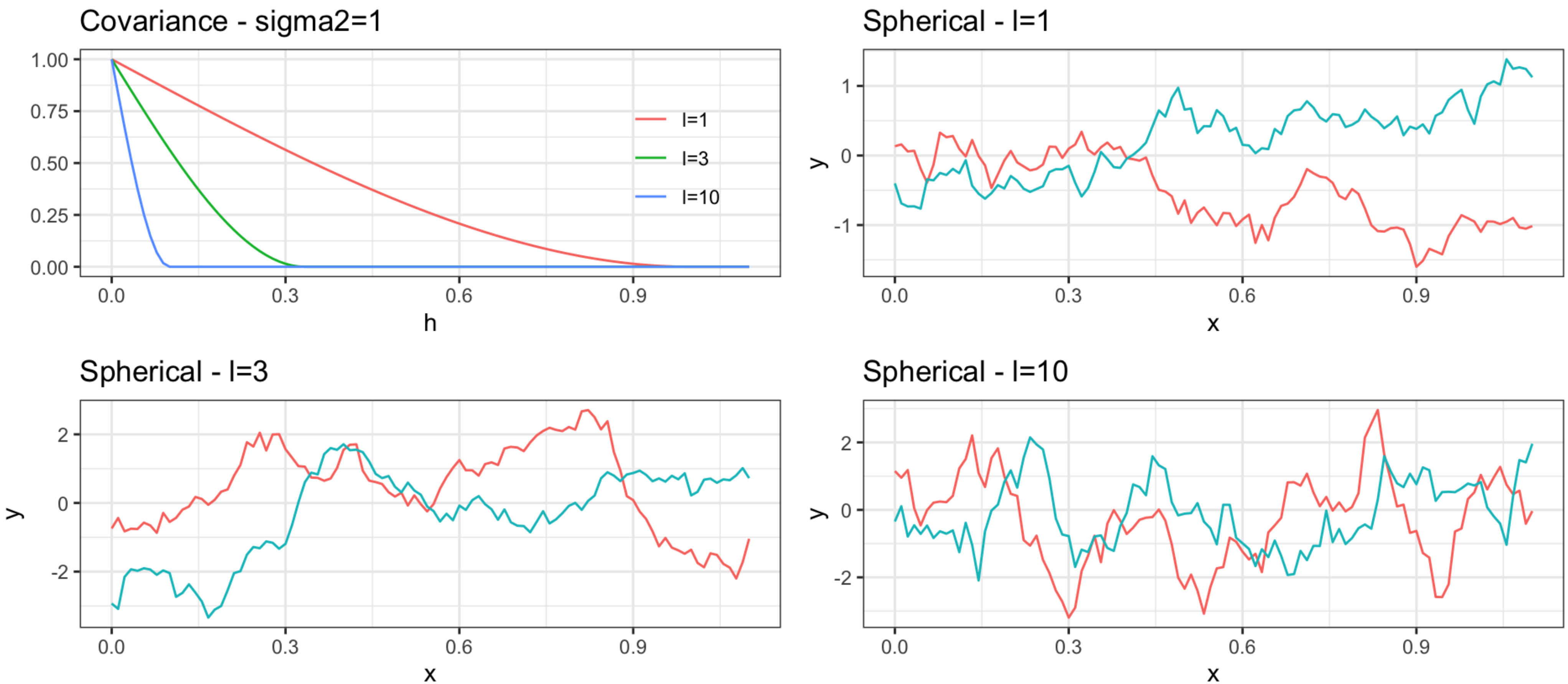


Rational Quadratic Covariance

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

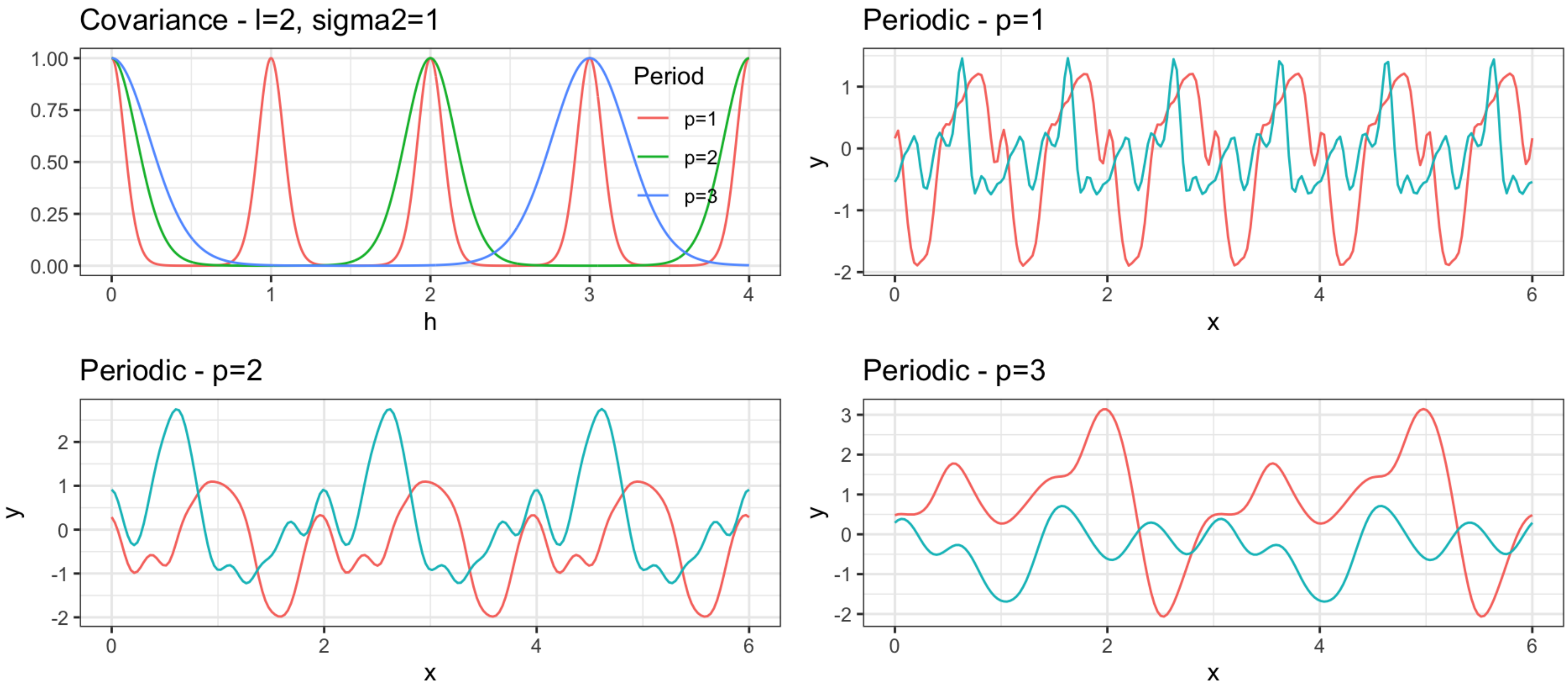
Spherical Covariance

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



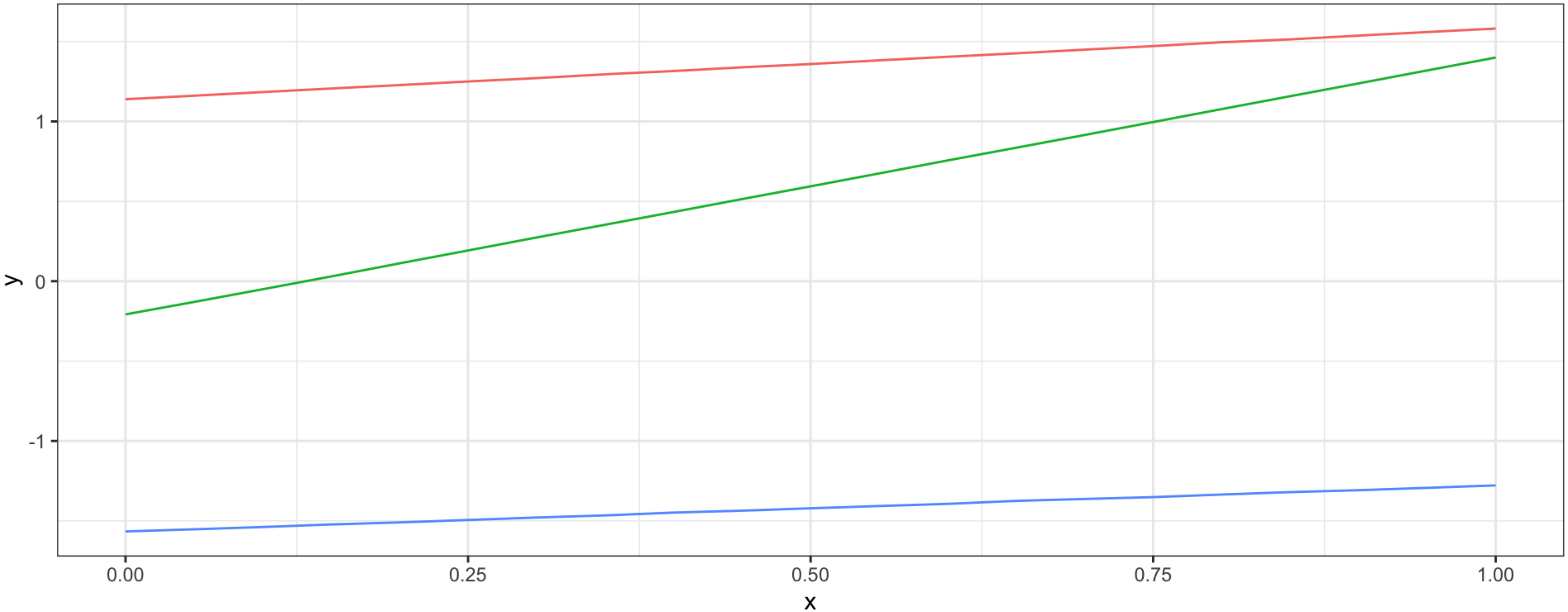
Periodic Covariance

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



Linear Covariance

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



Combining covariance functions

If we definite two valid covariance functions, $\text{Cov}_a(y_{t_i}, y_{t_j})$ and $\text{Cov}_b(y_{t_i}, y_{t_j})$ then the following are also valid covariance functions,

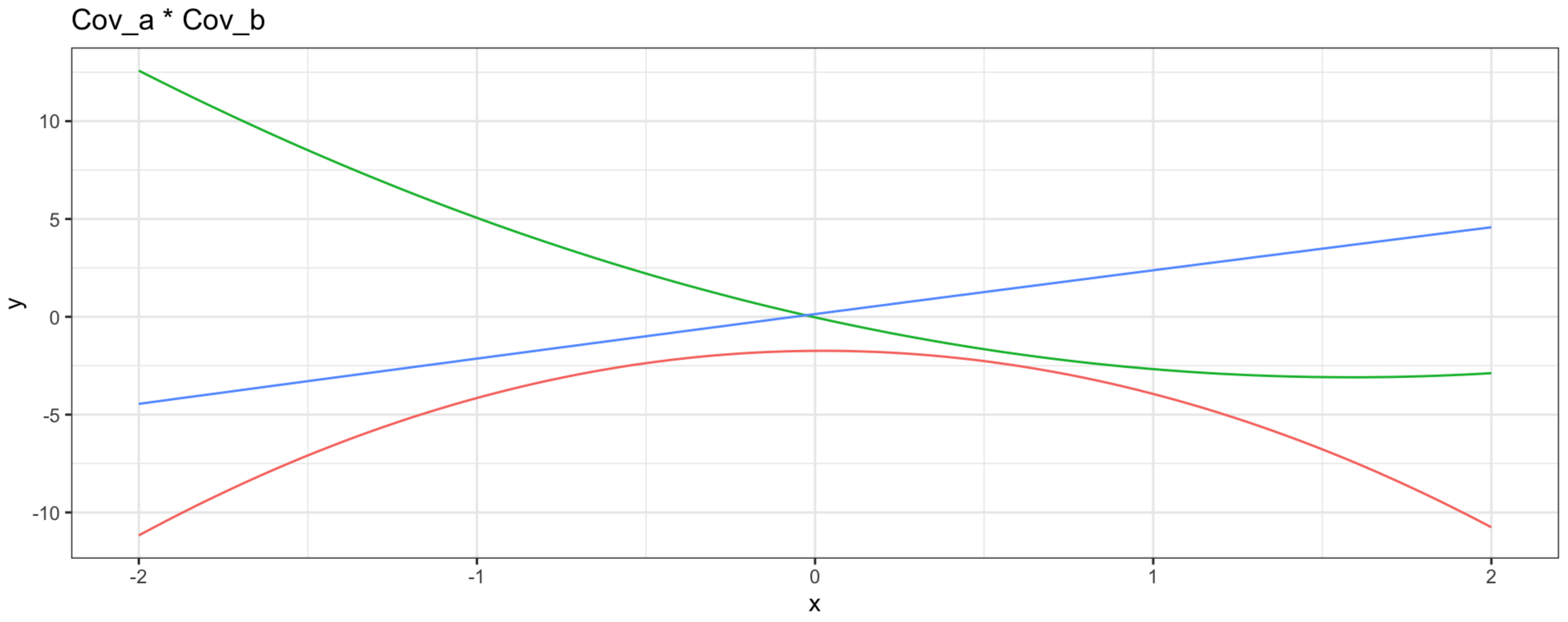
$$\text{Cov}_a(y_{t_i}, y_{t_j}) + \text{Cov}_b(y_{t_i}, y_{t_j})$$

$$\text{Cov}_a(y_{t_i}, y_{t_j}) \times \text{Cov}_b(y_{t_i}, y_{t_j})$$

Linear × Linear → Quadratic

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

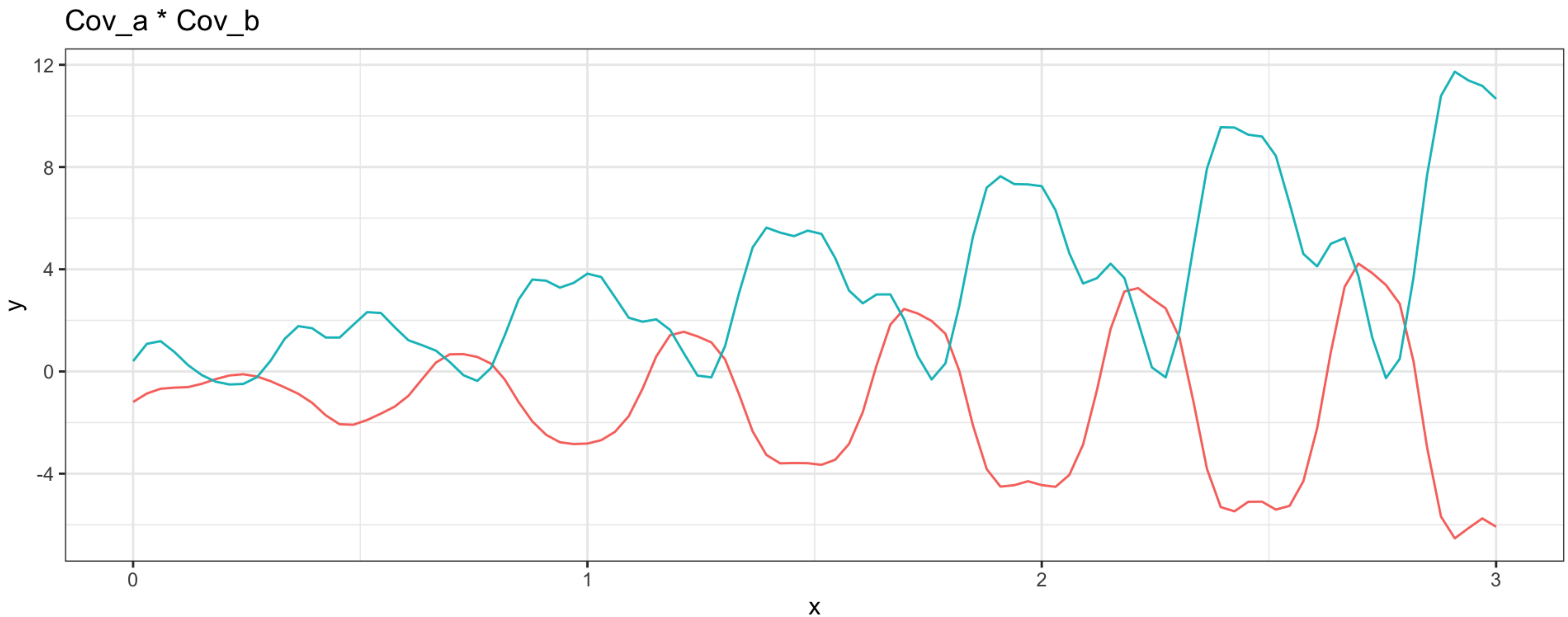
$$\text{Cov}_b(y_{t_i}, y_{t_j}) = 2 + 1\,(t_i \times t_j)$$



Linear × Periodic

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

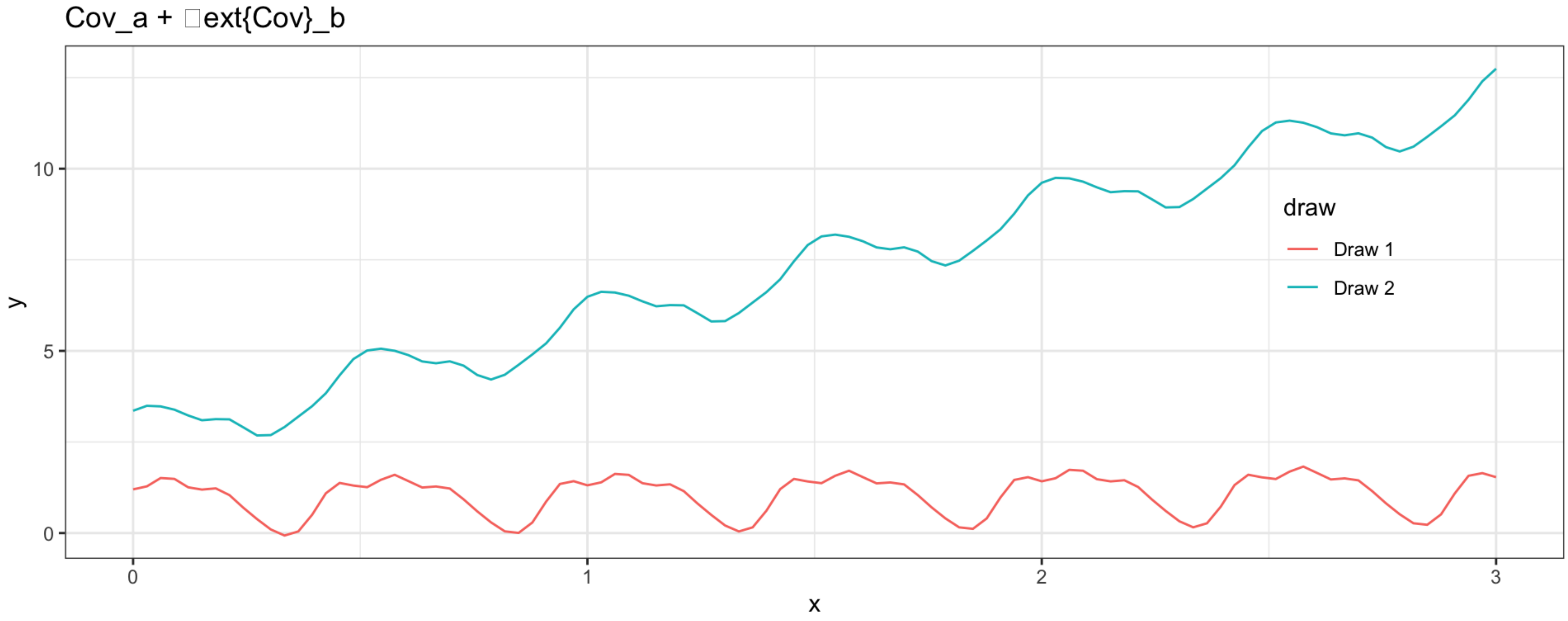
$$Cov_b(y_{t_i}, y_{t_j}) = \exp\left(-2\ \sin^2(2\pi\, h)\right)$$



Linear + Periodic

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

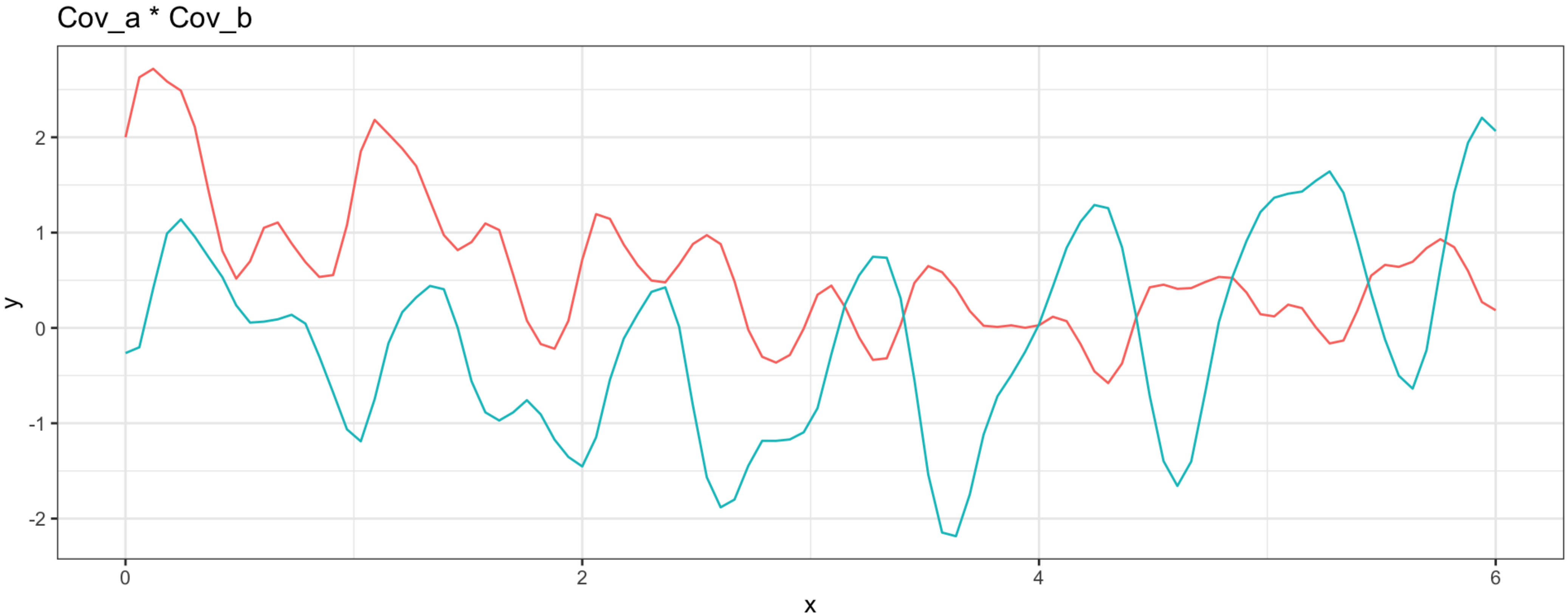
$$\text{Cov}_b(y_{t_i}, y_{t_j}) = \exp\left(-2\ \sin^2(2\pi\, h)\right)$$



Sq Exp × Periodic → Locally Periodic

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

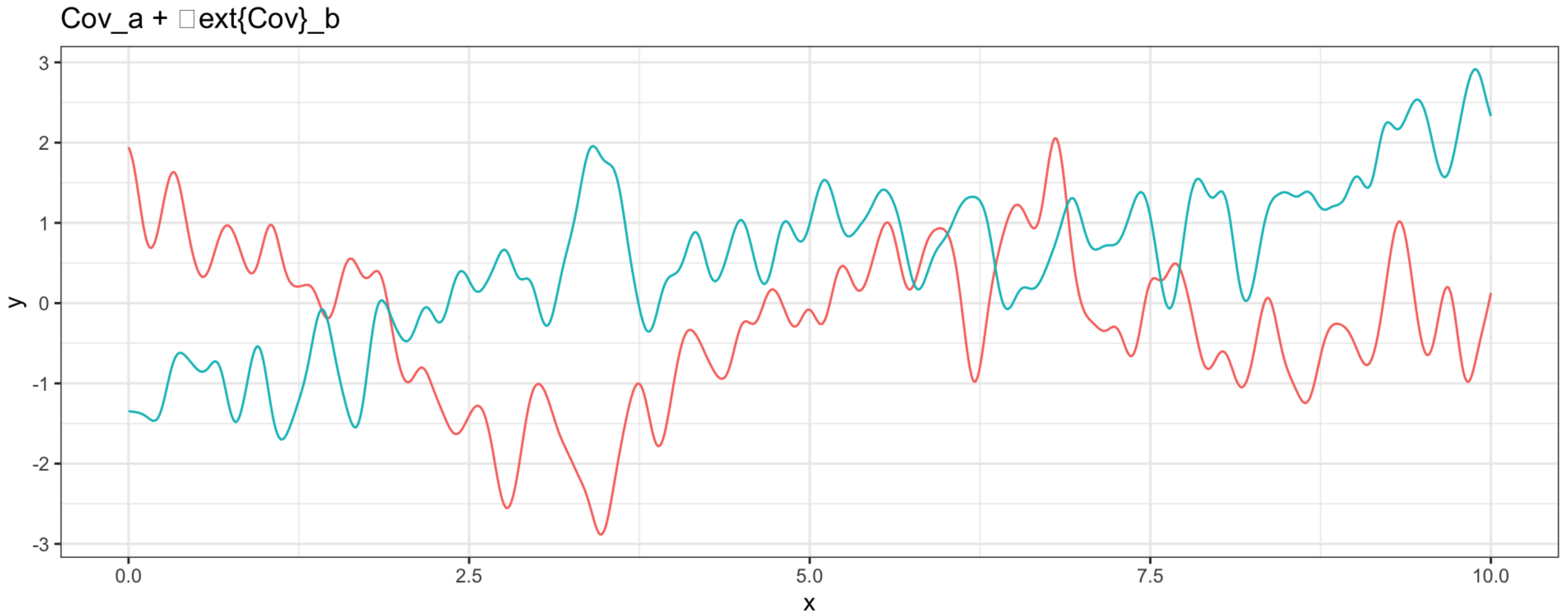
$$\text{Cov}_b(y_{t_i}, y_{t_j}) = \exp\left(-2\ \sin^2(\pi\, h)\right)$$



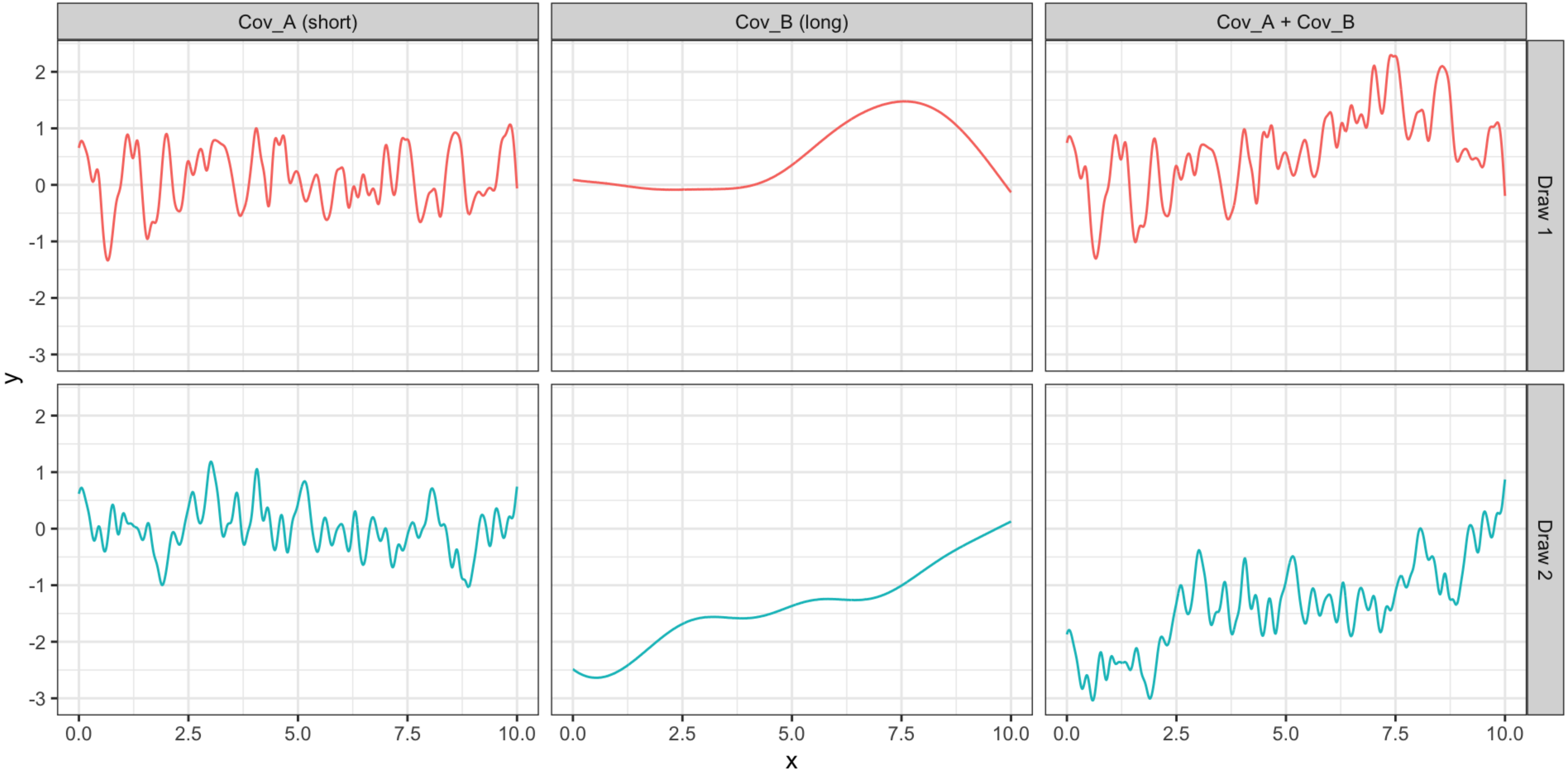
Sq Exp (short) + Sq Exp (long)

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

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

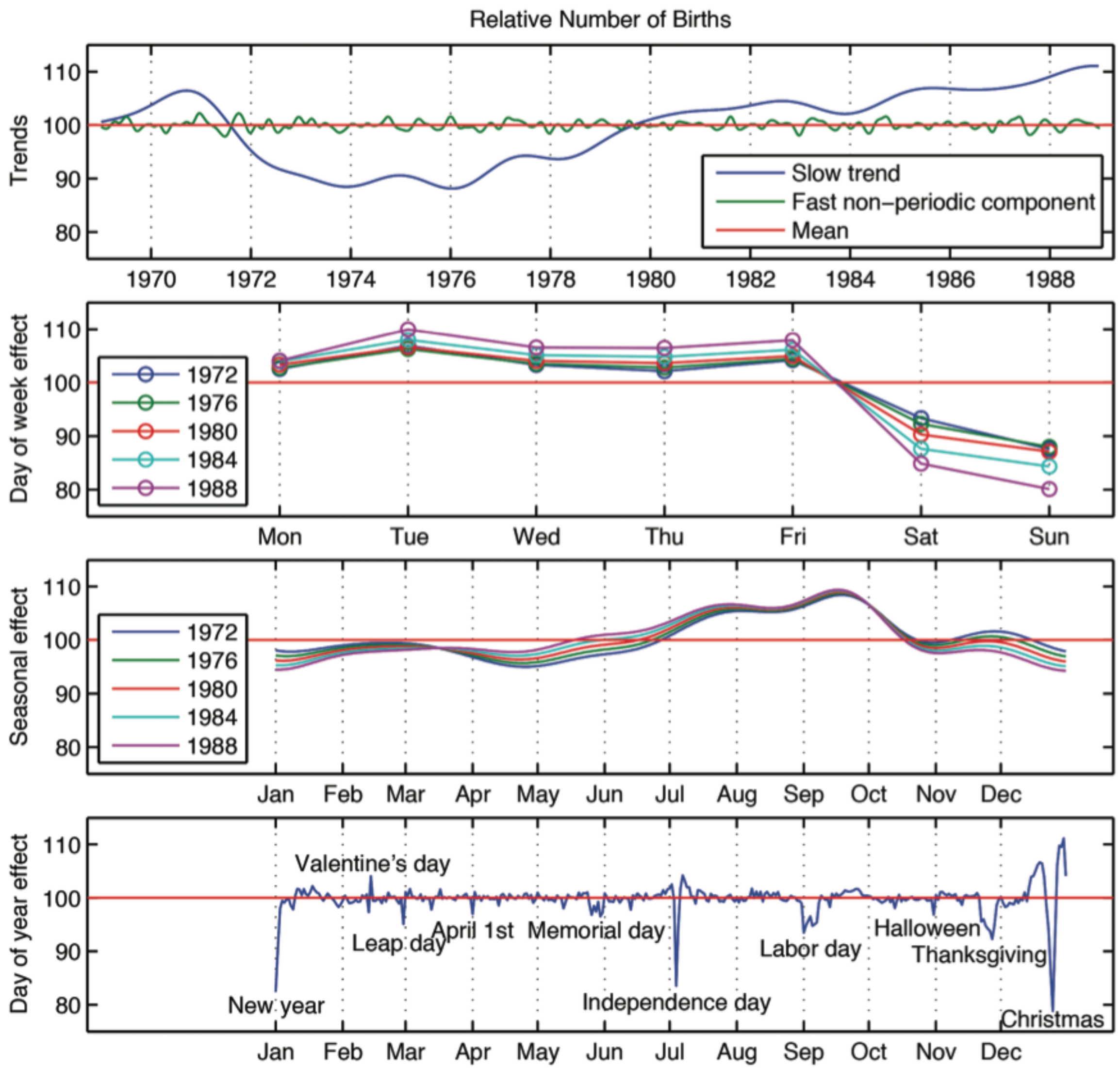


Seen another way



BDA3 example

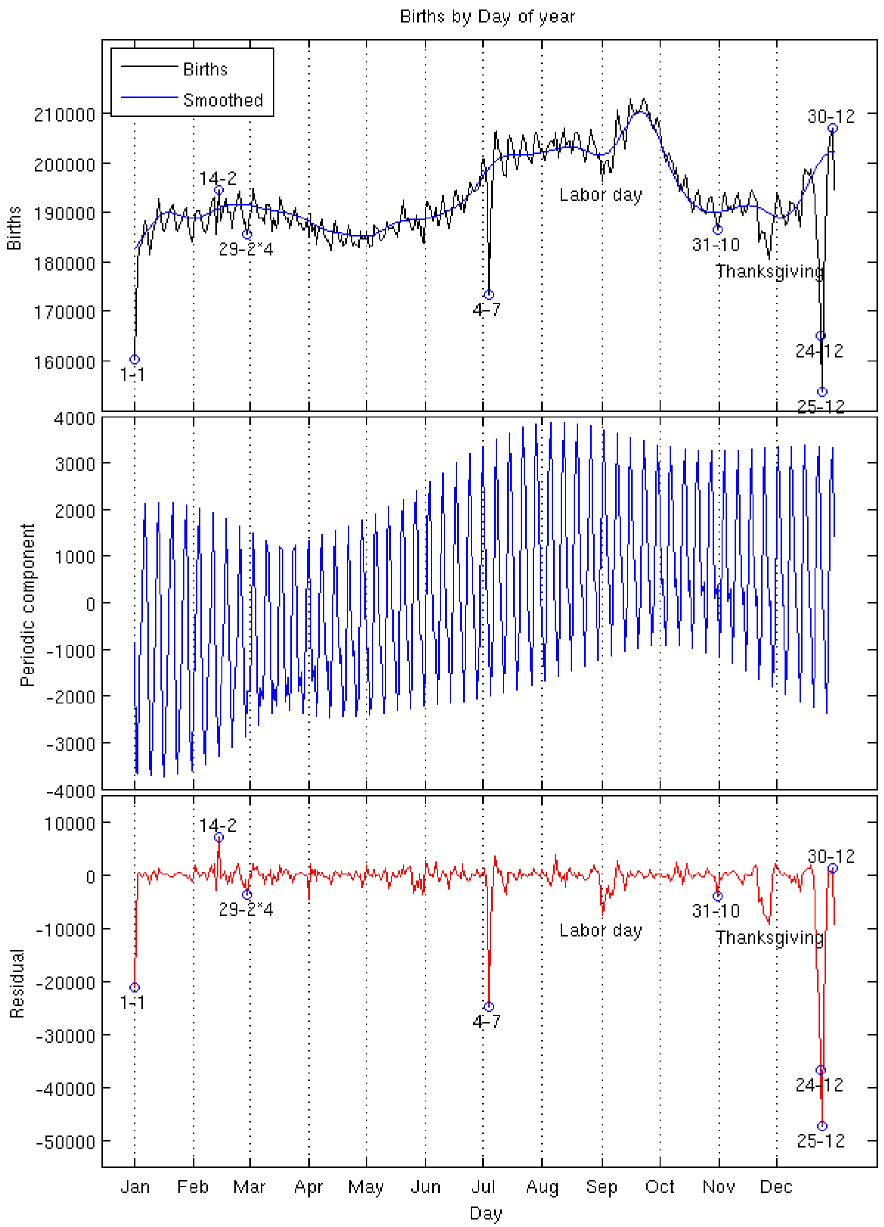
BDA3



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

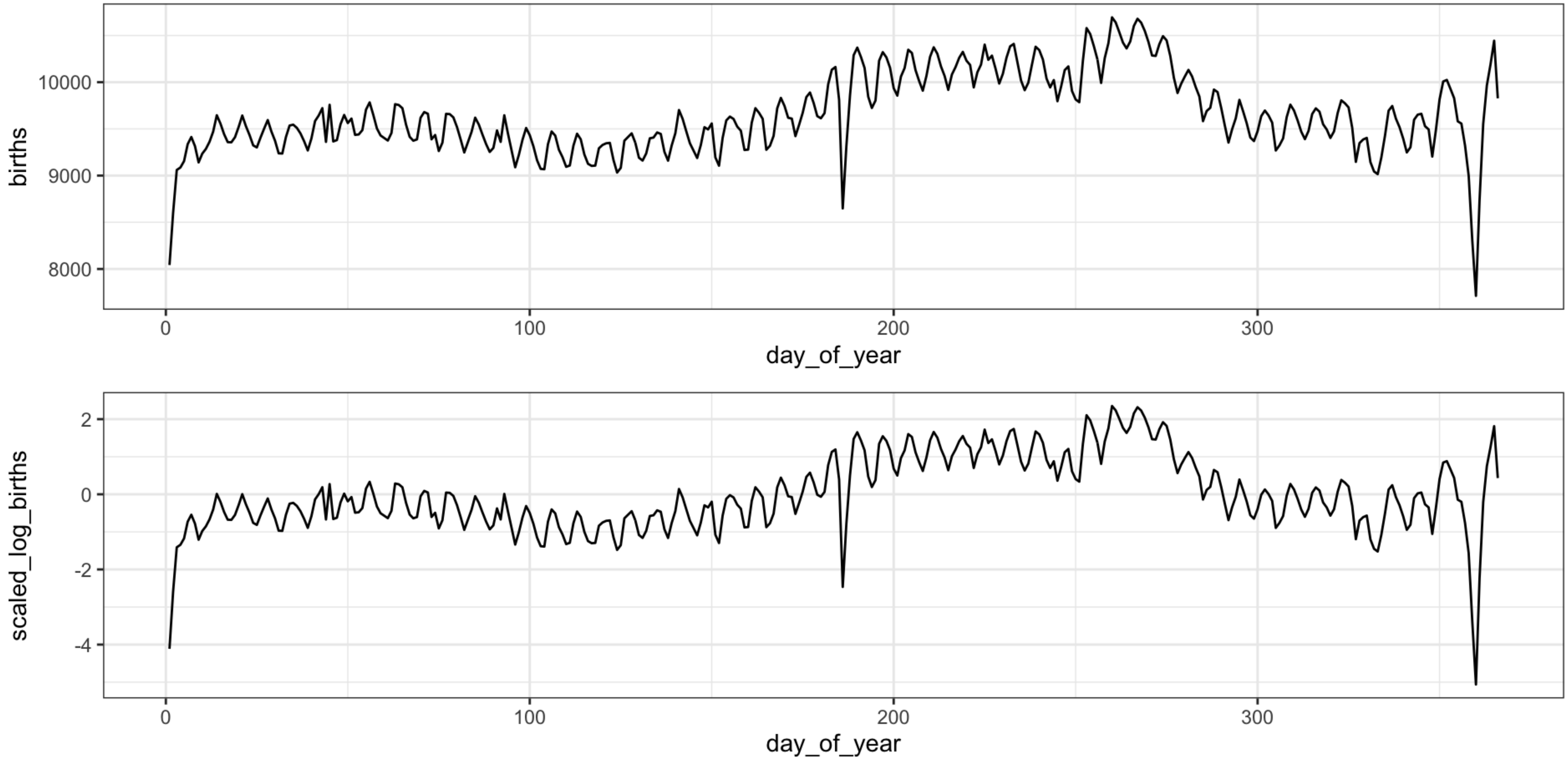
Source

Births (one year)



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

Birth data



births
10000
9000
8000
0
100
200
300
day_of_year
scaled_log_births
2
0
-2
-4
0
100
200
300
day_of_year

Model (JAGS)

```
model = "model{
  y ~ dmnorm(rep(0,N), 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] / per), 2

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

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

  for(i in 1:3){
    sigma2[i] ~ dt(0, 2.5, 1) T(0,)
    l[i] ~ dt(0, 2.5, 1) T(0,)
  }
}"
```

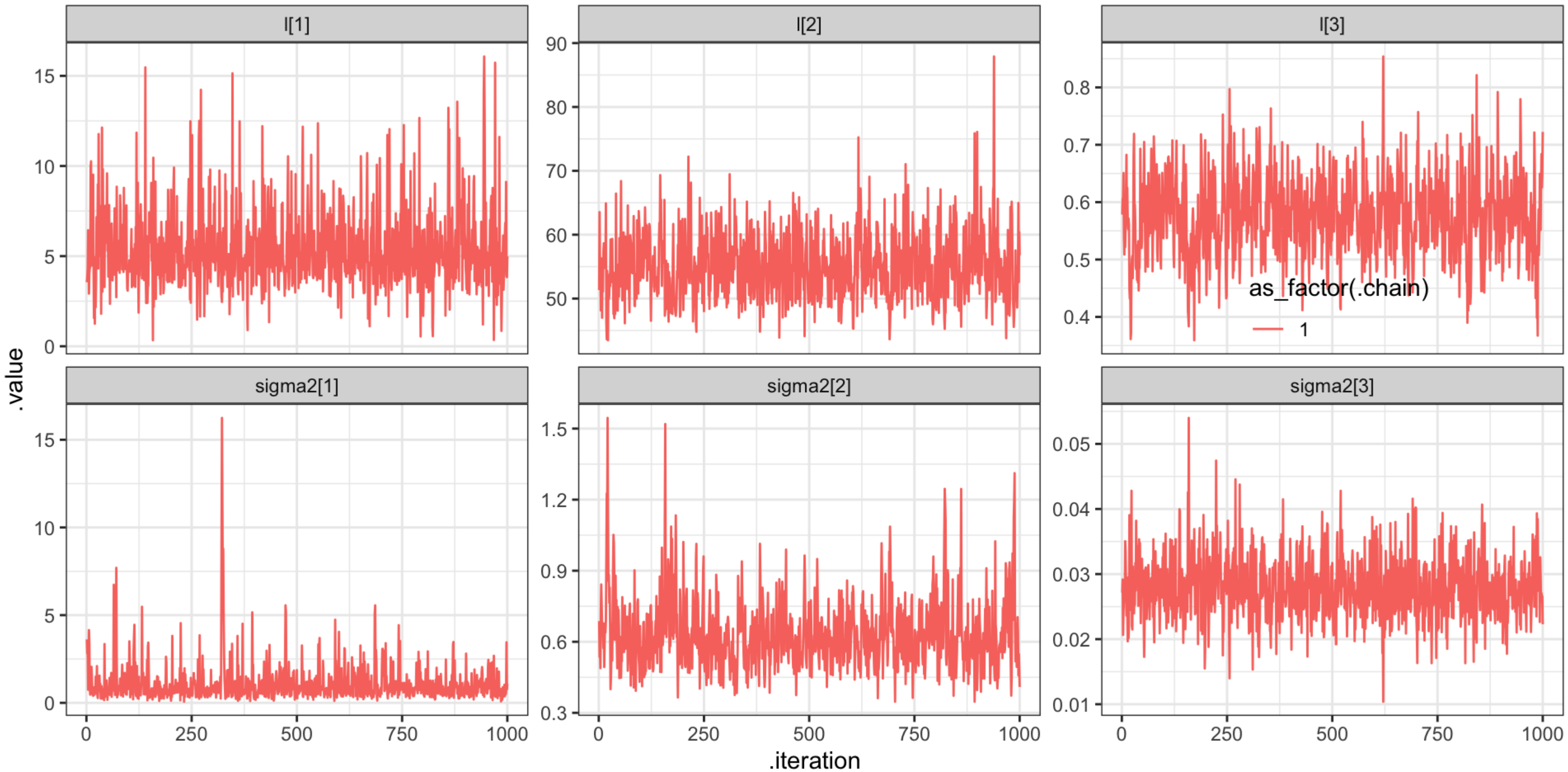
Model fitting

```
m = rjags::jags.model(
  textConnection(model),
  data = list(
    y = births$scaled_log_births,
    d = fields::rdist(births$day_of_year / ma
    per = 7 / max(births$day_of_year),
    pi = pi,
    N = nrow(births)
  ),
  n.adapt=5000,
  n.chains = 1
)
```

```
gp_coda = rjags::coda.samples(
  m, variable.names=c("sigma2", "l"),
  n.iter=5000,
  thin=5
)
```

Diagnostics

Traceplot | Density



Component Contributions

We can view our GP in the following ways (marginal form),

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

but with appropriate conditioning we can also think of $\boldsymbol{y}$ as being the sum of multiple independent GPs (latent form)

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathrm{w}_1(\boldsymbol{t}) + \mathrm{w}_2(\boldsymbol{t}) + \mathrm{w}_3(\boldsymbol{t})$$

where

$$\begin{aligned}
\mathrm{w}_1(\boldsymbol{t}) &\sim \mathrm{N}(0, \boldsymbol{\Sigma}_1) && \text{(sq exp covariance)} \\
\mathrm{w}_2(\boldsymbol{t}) &\sim \mathrm{N}(0, \boldsymbol{\Sigma}_2) && \text{(periodic x sq exp cov)} \\
\mathrm{w}_3(\boldsymbol{t}) &\sim \mathrm{N}(0, \sigma^2 \boldsymbol{I}) && \text{(nugget cov / white noise)}
\end{aligned}$$

Decomposition of Covariance Components

$$\begin{bmatrix} \mathrm{y} \\ \mathrm{w}_1 \\ \mathrm{w}_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_1 + \Sigma_2 + \sigma^2 \boldsymbol{I} & \Sigma_1 & \Sigma_2 \\ \Sigma_1 & \Sigma_1 & 0 \\ \Sigma_2 & 0 & \Sigma_2 \end{bmatrix} \right)$$

therefore, if we want to know the contribution of w_1 we have the following

$$\mathrm{w}_1 \mid \boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\mu}_{\text{cond}}, \ \boldsymbol{\Sigma}_{\text{cond}})$$

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

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

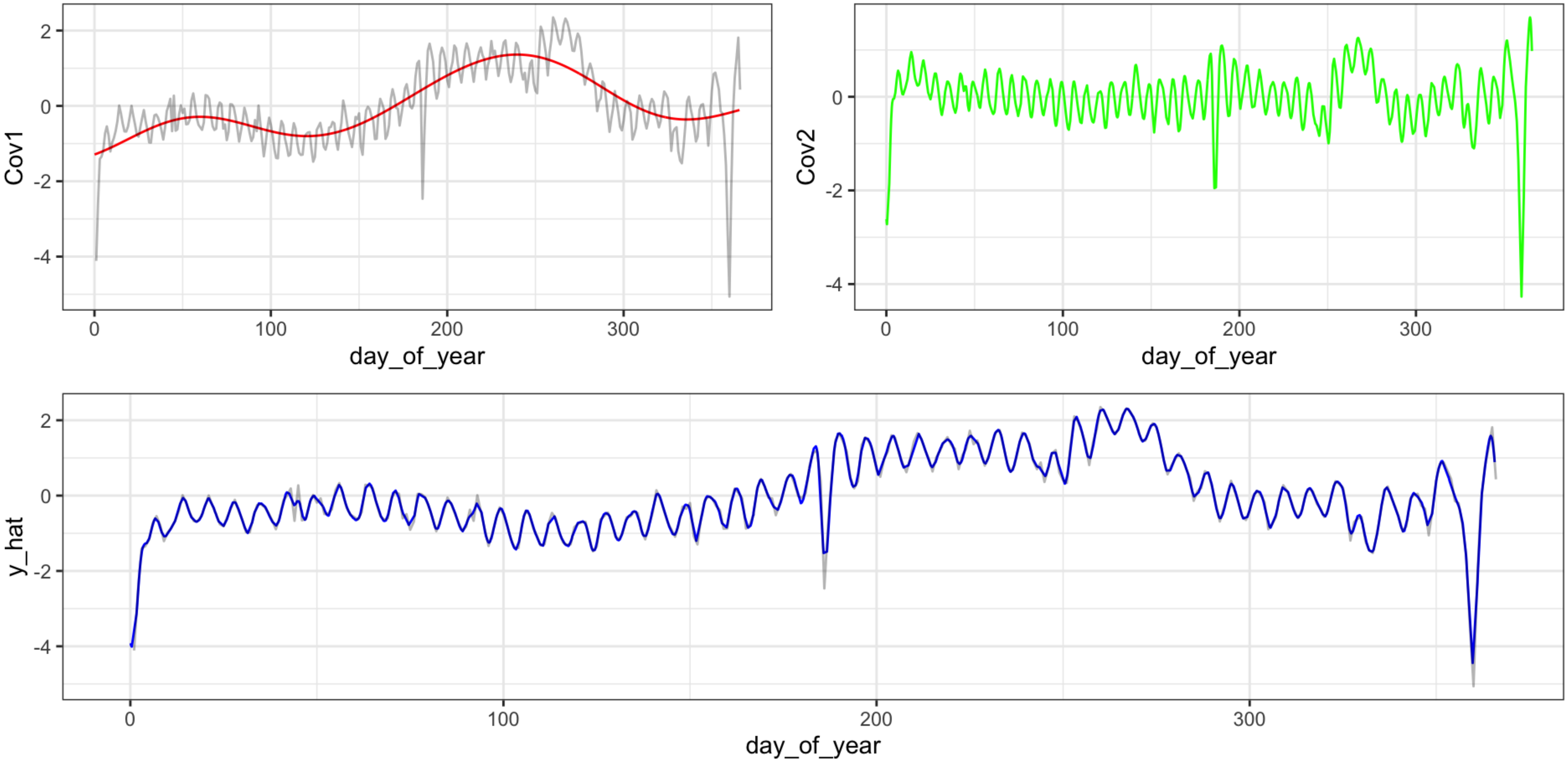
Covariance calculations

```r
cov_C1 = function(d, sigma2, l, per) {
 sigma2[1] * exp(- (l[1] * d)^2)
}

cov_C2 = function(d, sigma2, l, per) {
  sigma2[2] * exp(- (l[2] * d)^2 - 2 * (l[3] * sin(pi*d / per))^2)
}

cov_full = function(d, sigma2, l, per) {
  cov_C1(d, sigma2, l, per) +
  cov_C2(d, sigma2, l, per) +
  ifelse(abs(d)<1e-6, sigma2[3] + 1e-6, 0)
}
```

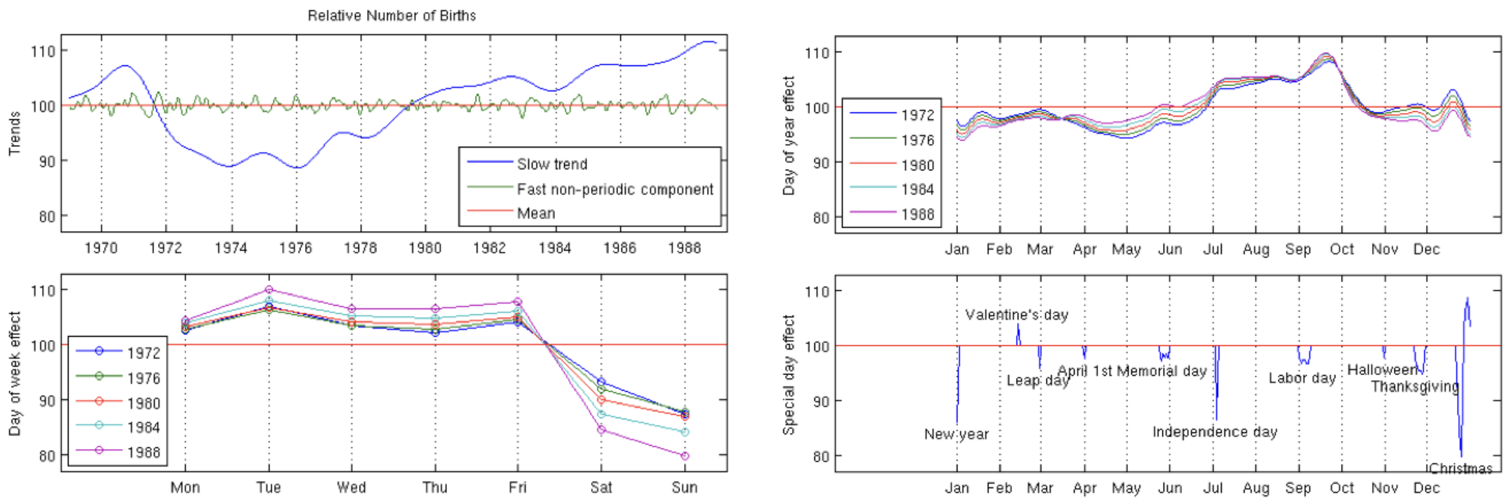
Conditional samples

```
full = dukestm::cond_predict(
  y=y, x=x, x_pred=x_pred, cov_full, sigma2 = post$sigma2, l=post$l, per=7 / max(births$day_of_ye
)

comp_C1 = dukestm::cond_predict(
  y=y, x=x, x_pred=x_pred, sigma2 = post$sigma2, l=post$l, per=7 / max(births$day_of_year),
  cov_f_o  = cov_full,
  cov_f_p  = cov_C1,
  cov_f_po = cov_C1
)

comp_C2 = dukestm::cond_predict(
  y=y, x=x, x_pred=x_pred, sigma2 = post$sigma2, l=post$l, per=7 / max(births$day_of_year),
  cov_f_o  = cov_full,
  cov_f_p  = cov_C2,
  cov_f_po = cov_C2
)
```

Results



Births (multiple years)

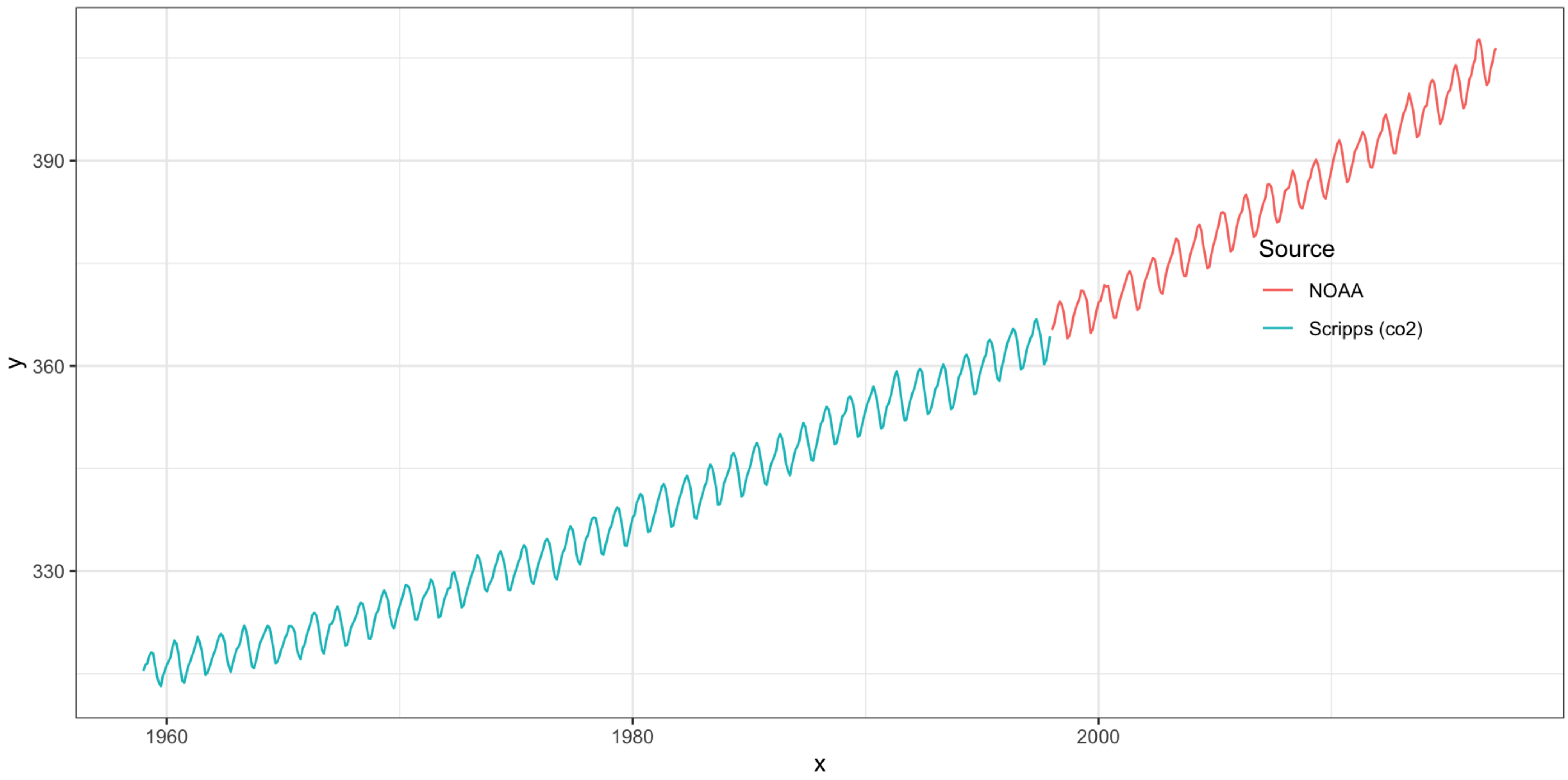


Full stan case study here with code here

1. slowly changing trend - yearly (*sq exp cov*)
2. small time scale trend - monthly (*sq exp cov*)
3. 7 day periodic - day of week effect (*periodic* $\times$ *sq exp cov*)
4. 365.25 day periodic - day of year effect (*periodic* $\times$ *sq exp cov*)
5. special days and interaction with weekends (*linear cov*)
6. independent Gaussian noise (*nugget cov*)
7. constant mean

Mauna Loa Example

Atmospheric CO_2



GP Model

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

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

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

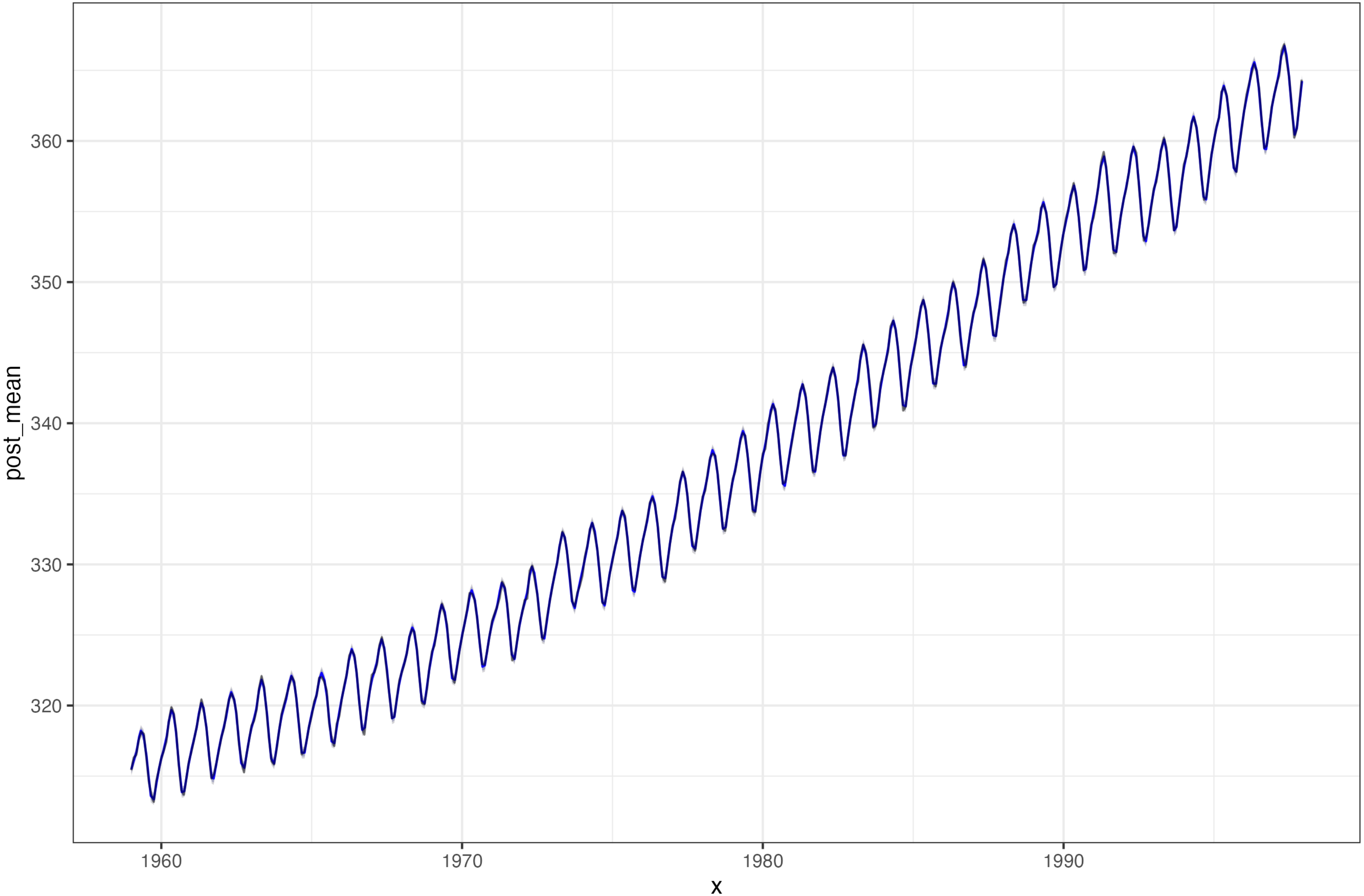
$$\{\boldsymbol{\Sigma}_1\}_{ij} = \sigma_1^2 \exp\left(-(l_1 \cdot d_{ij})^2\right) \qquad \text{smooth long term trend}$$

$$\{\boldsymbol{\Sigma}_2\}_{ij} = \sigma_2^2 \exp\left(-(l_2 \cdot d_{ij})^2\right) \exp\left(-2\,(l_3)^2 \sin^2(\pi\, d_{ij}/p)\right) \qquad \text{seasonal trend w/ decay}$$

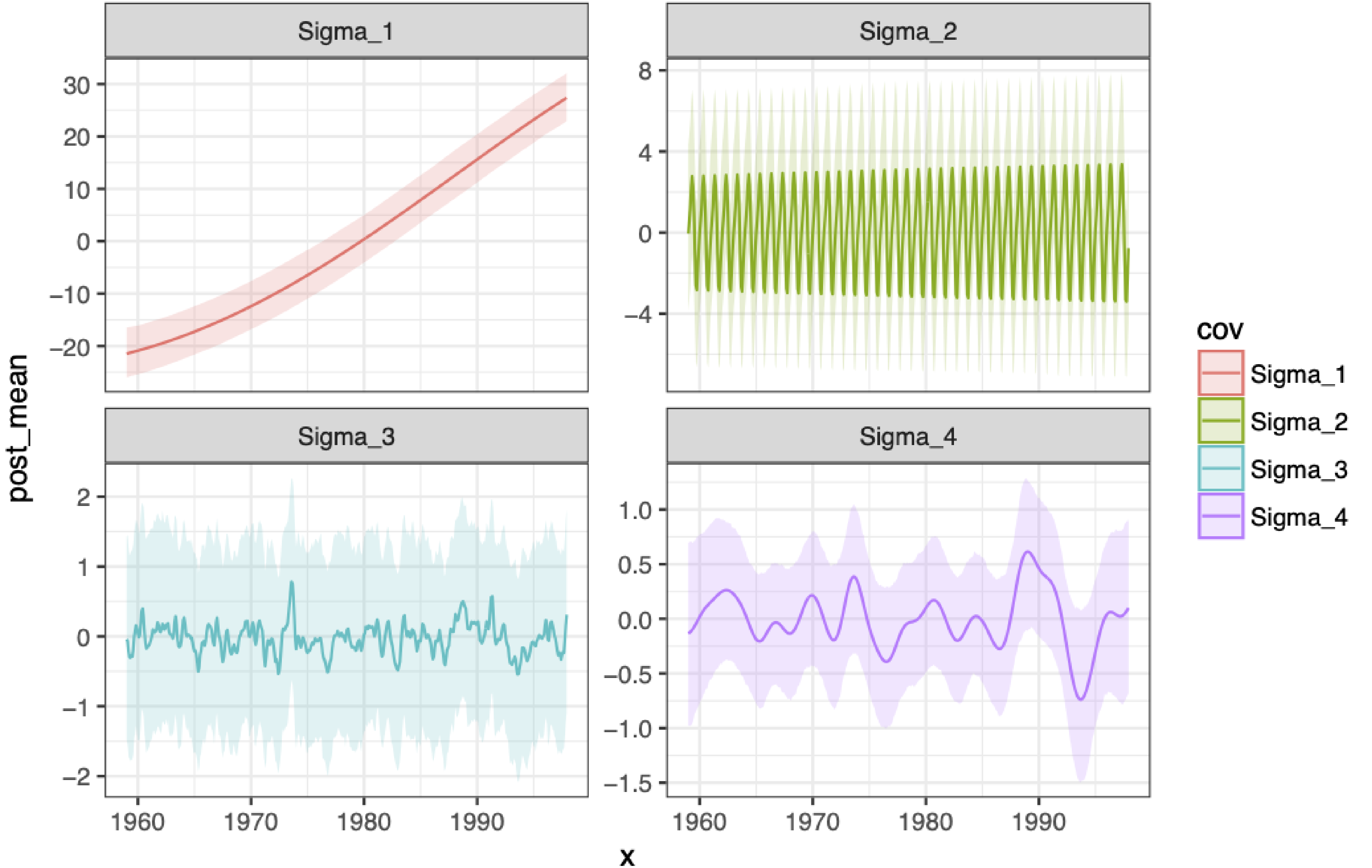
$$\{\boldsymbol{\Sigma}_3\}_{ij} = \sigma_3^2 \left(1 + \frac{(l_4 \cdot d_{ij})^2}{\alpha}\right)^{-\alpha} \qquad \text{small / medium term trend}$$

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

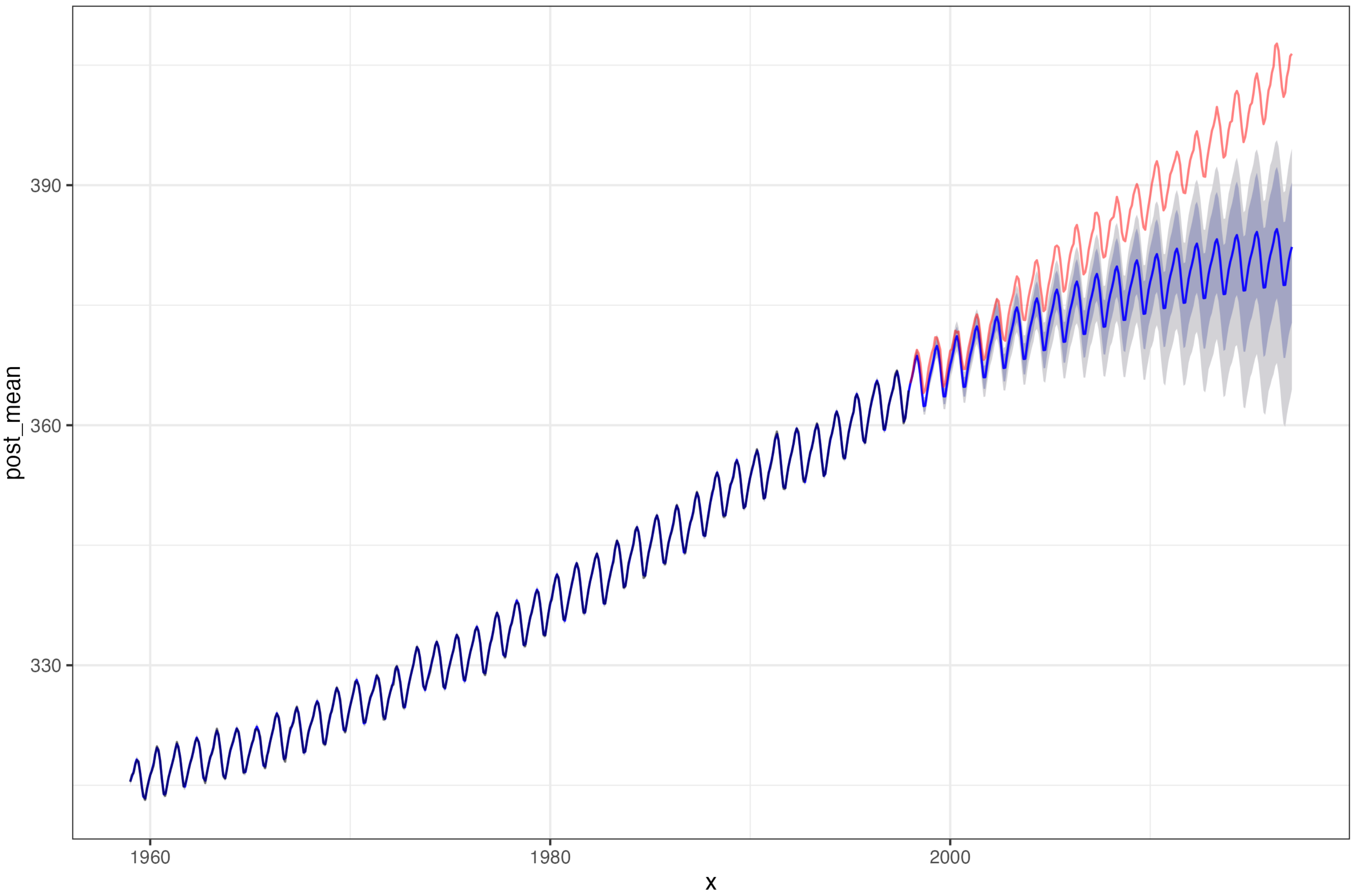
Model fit



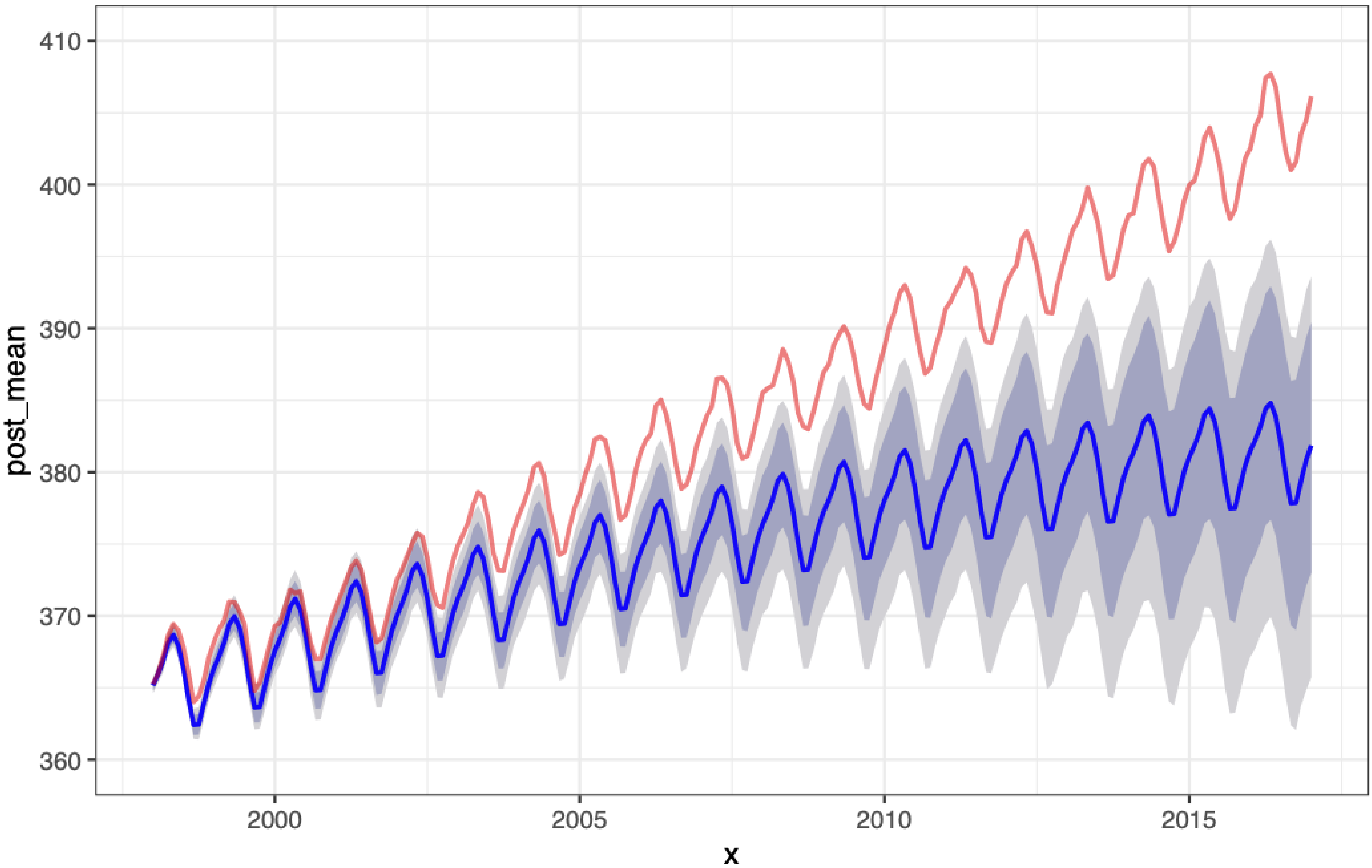
Fit Components



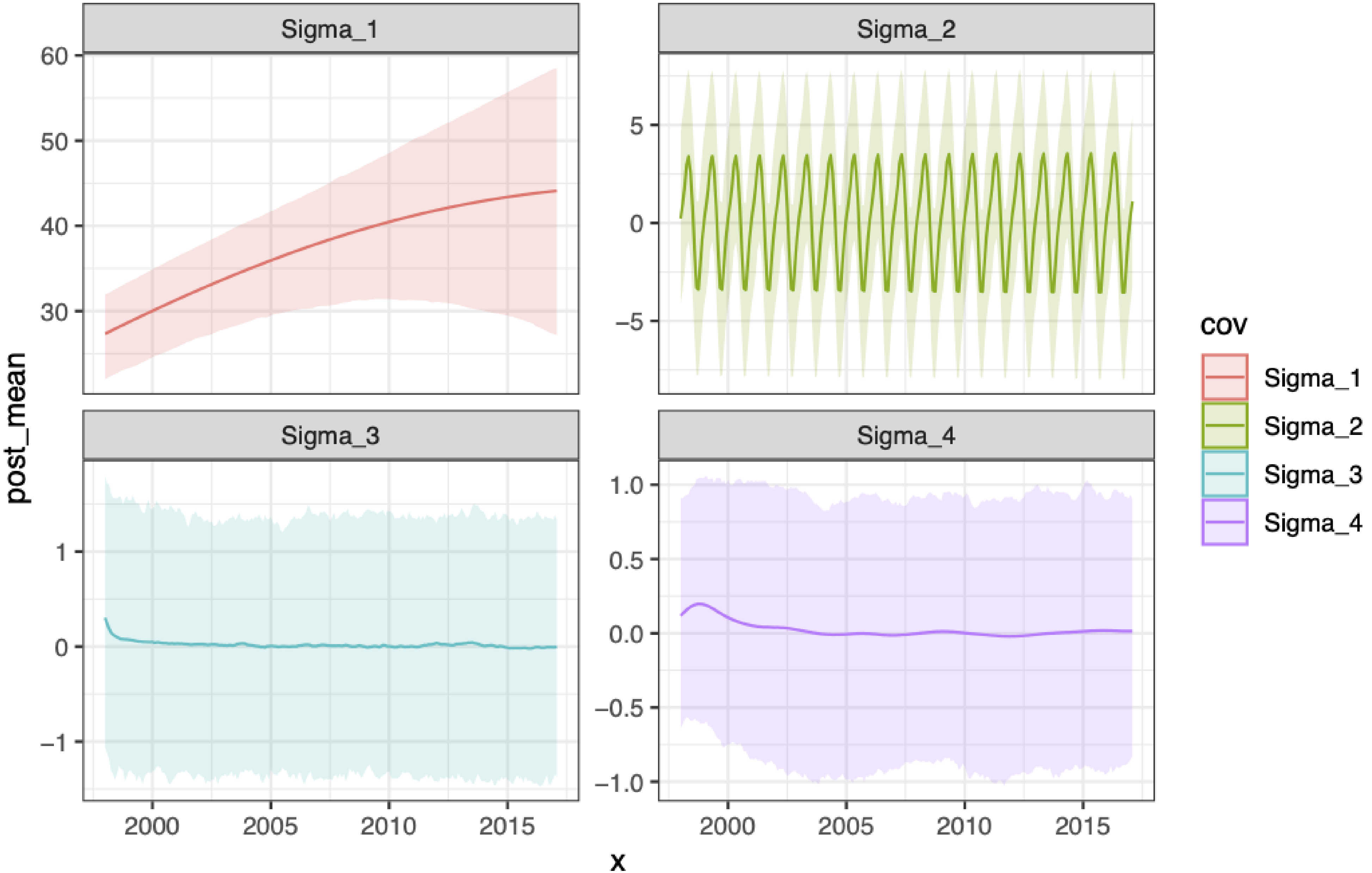
Model fit + forecast



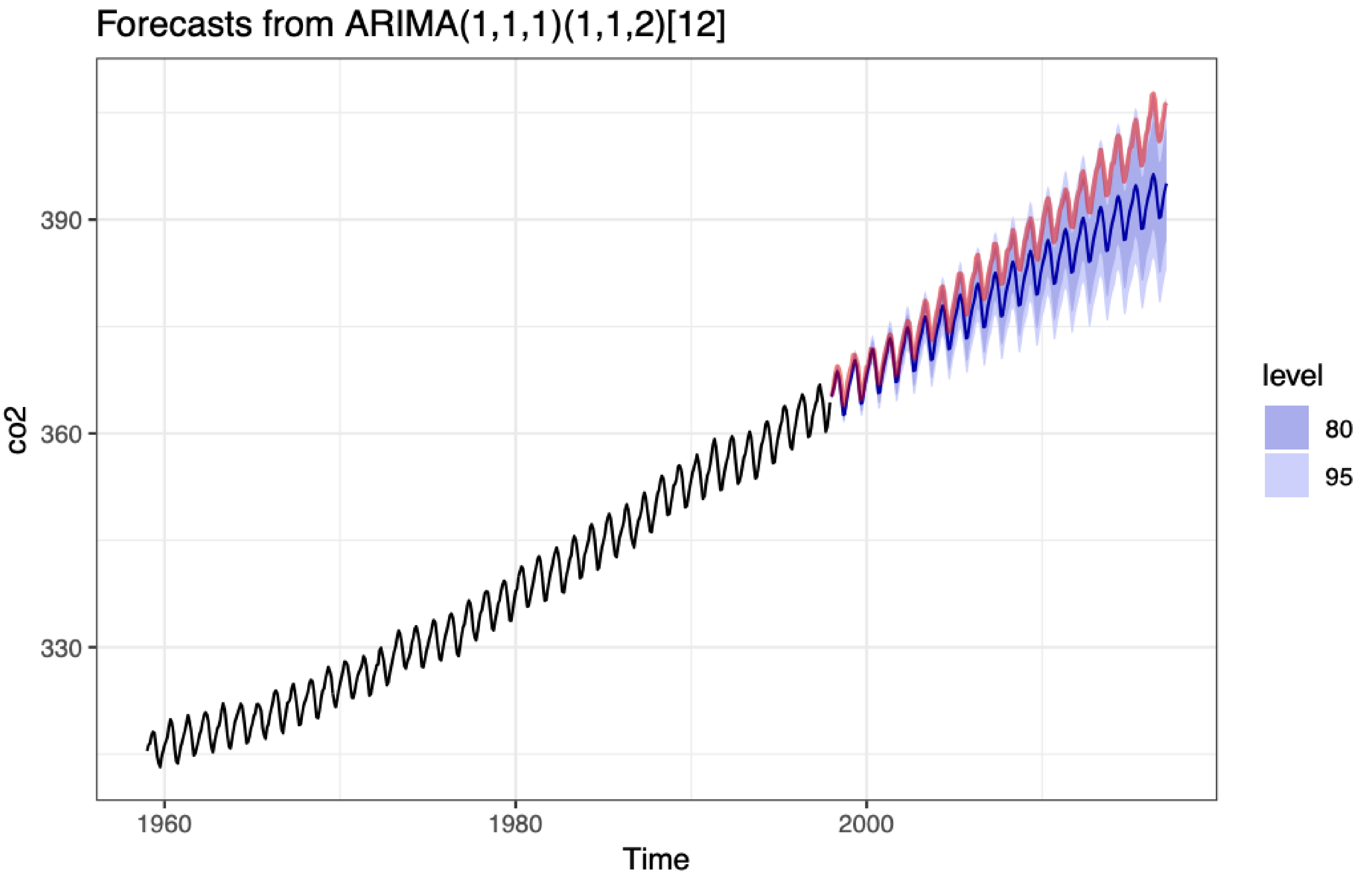
Forecast



Forecast components



ARIMA forecast



Model performance

Forecast dates	arima RMSE	gp RMSE
Jan 1998 - Jan 2003	1.10	1.91
Jan 1998 - Jan 2008	2.51	4.58
Jan 1998 - Jan 2013	3.82	7.71
Jan 1998 - Mar 2017	5.46	11.40