

Covariance Functions

Lecture 16

Dr. Colin Rundel

sta344 package

Basics

The package is based on the course organization and can be installed with,

```
1 remotes::install_github("sta344-fa22/sta344")
```

collects all of the utility functions from the host scripts (e.g. `util-crps.R`, `fix_pred_draw.R`, etc.)

```
1 library(sta344)
2 ls("package:sta344")
```

```
[1] "avg_temp"           "calc_crps"           "cond_predict"
[4] "emp_semivariogram" "epred_draws_fix"    "exp_cov"
[7] "exp_sv"             "gpglm"               "gplm"
[10] "linear_cov"        "matern_cov"         "normalize_weights"
[13] "nugget_cov"        "periodic_cov"      "pow_exp_cov"
[16] "pow_exp_sv"        "predicted_draws_fix" "residual_draws_fix"
[19] "rmvnorm"           "rquad_cov"          "sphere_cov"
[22] "sq_exp_cov"        "sq_exp_sv"          "strip_attrs"
```

From last time

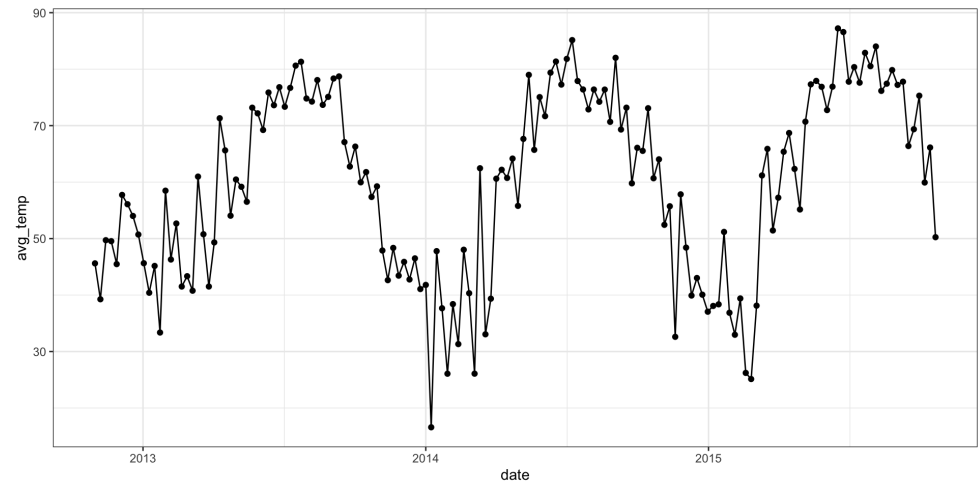
```
1 ( temp = readRDS("data/avg_temp_df.rds") ) %>%  
2   slice(1:(52*3)) %>%  
3   mutate(week = as.numeric(date - date[1]))
```

A tibble: 156 × 3

	date	avg_temp	week
	<date>	<dbl>	<dbl>
1	2012-10-31	45.6	0
2	2012-11-07	39.3	1
3	2012-11-14	49.7	2
4	2012-11-21	49.5	3
5	2012-11-28	45.5	4
6	2012-12-05	57.7	5
7	2012-12-12	56.1	6
8	2012-12-19	54.0	7
9	2012-12-26	50.7	8
10	2013-01-02	45.6	9

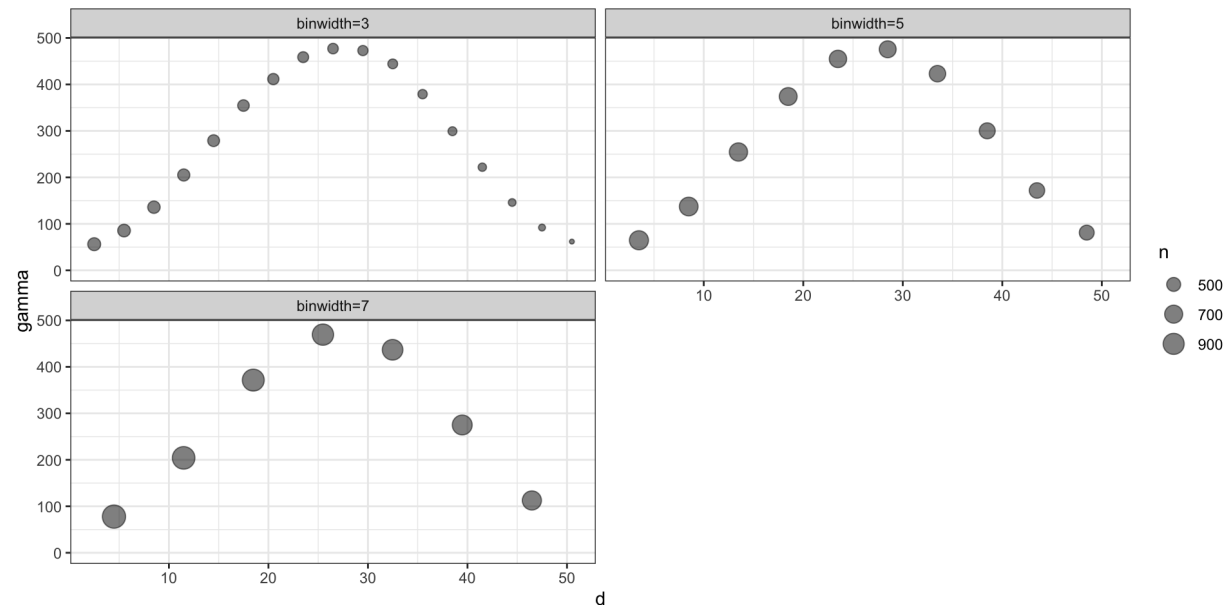
... with 146 more rows

```
1 ggplot(temp, aes(x=date, y=avg_temp)) +  
2   geom_line() +  
3   geom_point()
```



Empirical semi-variogram

```
1 emp_semivariogram(  
2   d=temp, y=avg_temp, x=week,  
3   bin=TRUE, binwidth=c(3,5,7)  
4 ) %>%  
5 ggplot(aes(x=d, y=gamma, size=n)) +  
6   geom_point(alpha=0.5) +  
7   facet_wrap(~paste0("binwidth=",bw), ncol=2) +  
8   ylim(0,NA)
```

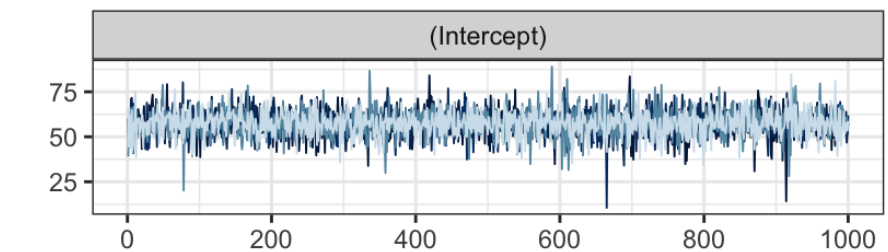
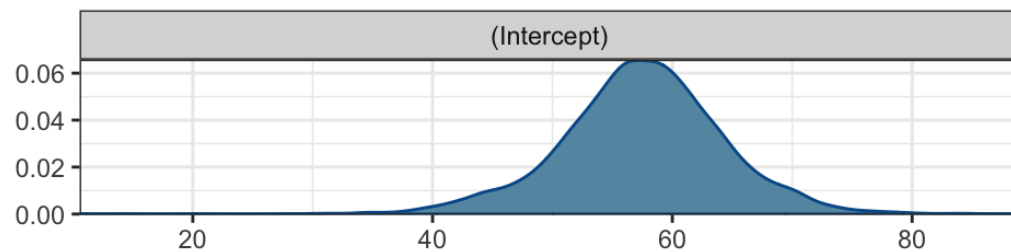
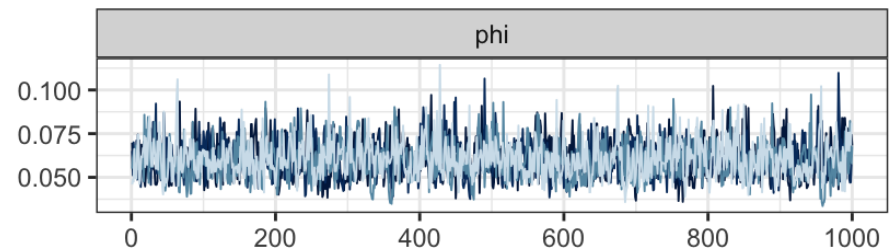
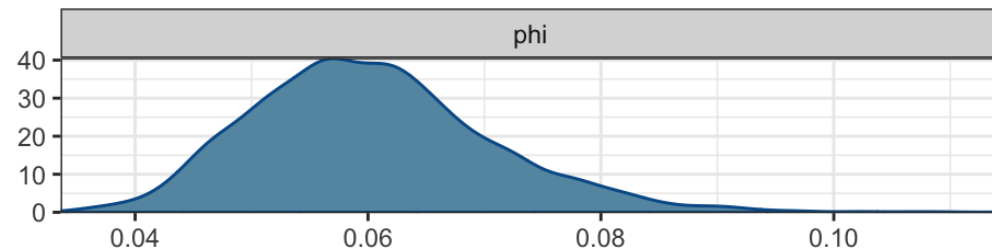
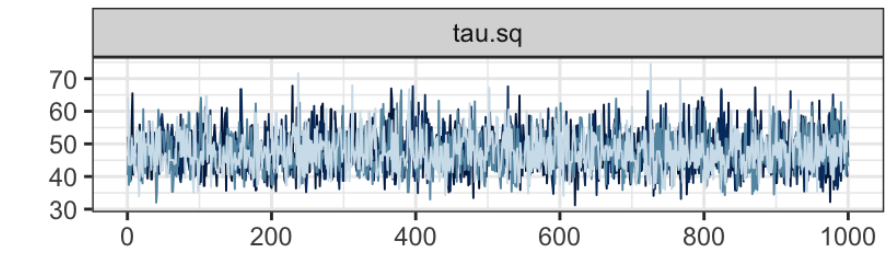
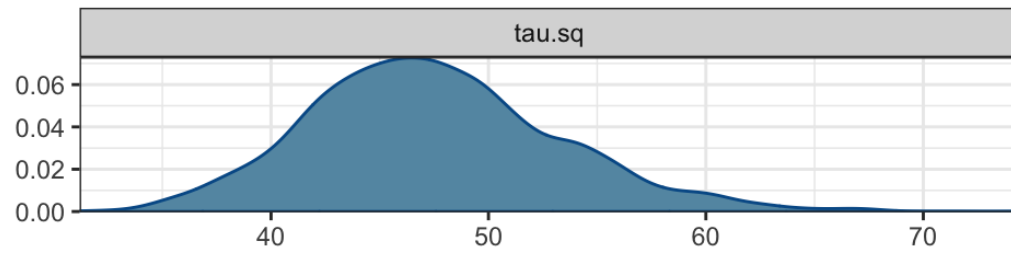
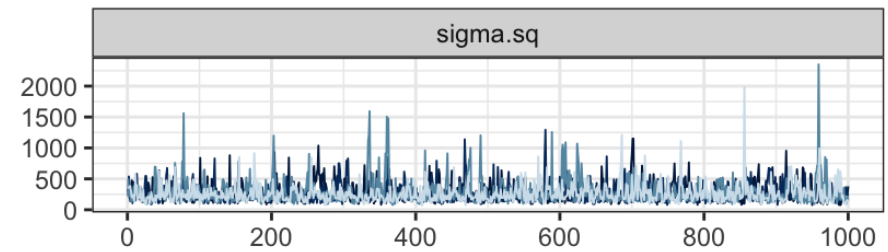
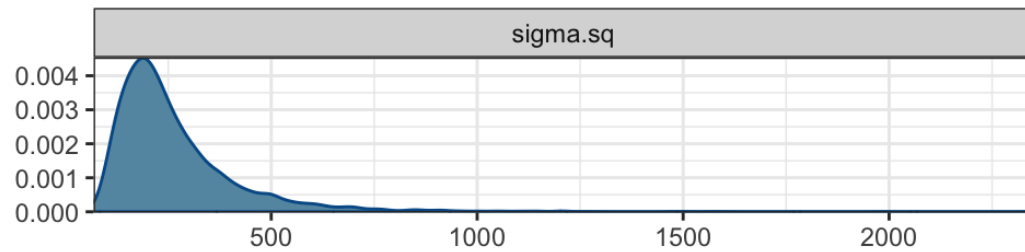


Model fitting

```
1 m = gplm(  
2   avg_temp~1, data = temp, coords = "week",  
3   cov_model = "gaussian",  
4   starting=list(  
5     "phi"=sqrt(3)/4, "sigma.sq"=1, "tau.sq"=1  
6   ),  
7   tuning=list(  
8     "phi"=1, "sigma.sq"=1, "tau.sq"=1  
9   ),  
10  priors=list(  
11    "phi.unif"=c(sqrt(3)/52, sqrt(3)/1),  
12    "sigma.sq.ig"=c(2, 1),  
13    "tau.sq.ig"=c(2, 1)  
14  ),  
15  thin = 5,  
16  n_batch = 100,  
17  batch_len = 100  
18 )
```

Model plots

```
1 plot(m)
```



Chain
— 1
— 2
— 3
— 4

Prediction

```
1 newdata = data.frame(  
2   week = seq(0, 3.5*52) |> jitter()  
3 )  
4  
5 (p = predict(m, newdata=newdata, coords = "week"))
```

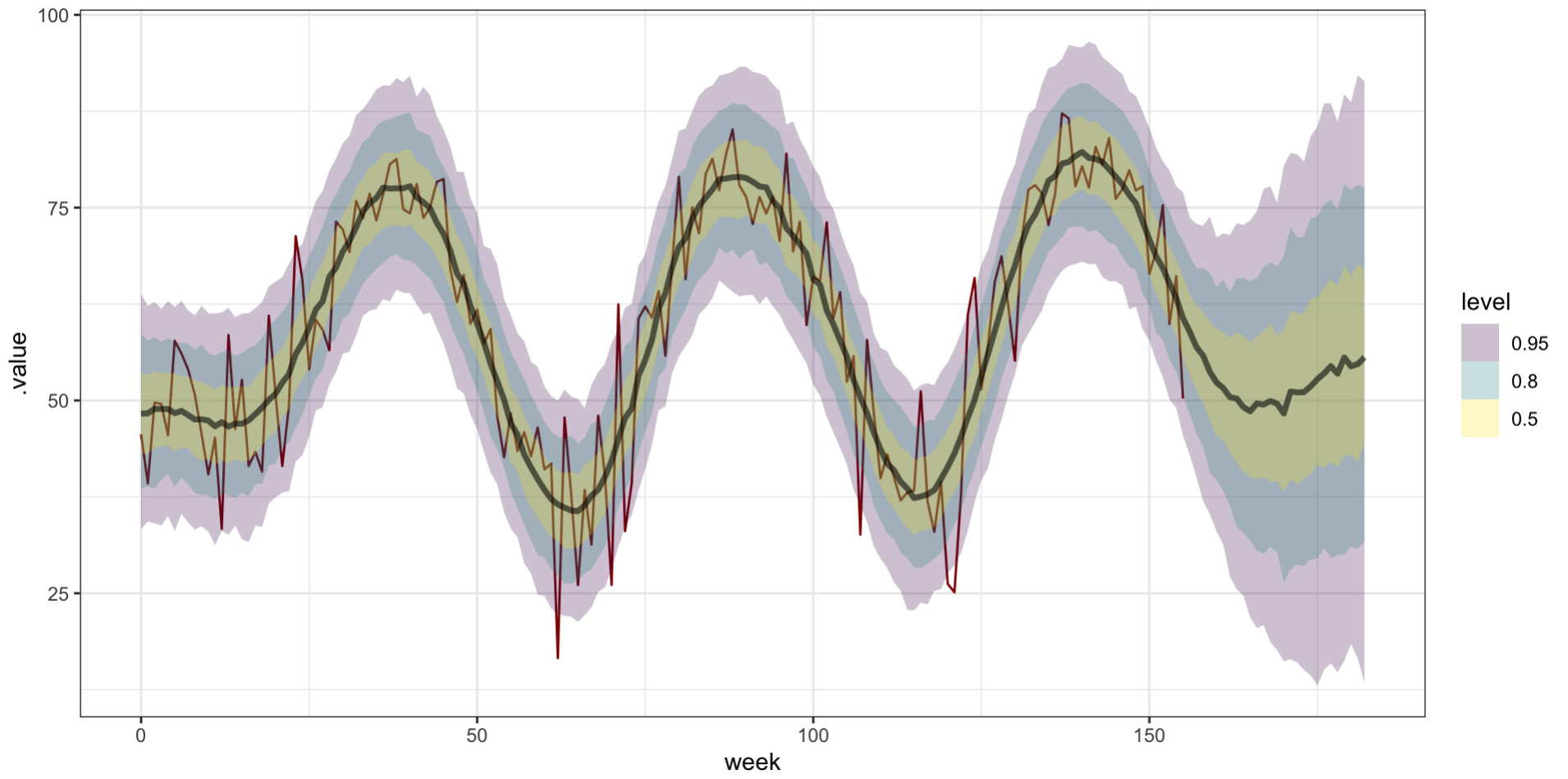
```
# A draws_matrix: 1000 iterations, 4 chains, and 183 variables
```

```
variable
```

```
draw y[1] y[2] y[3] y[4] y[5] y[6] y[7] y[8]  
1    38  45  57  42  43  47  43  55  
2    51  54  47  47  46  36  50  46  
3    42  56  45  57  36  54  48  60  
4    48  43  52  39  53  48  54  59  
5    39  38  39  46  50  50  49  48  
6    65  43  50  45  48  43  55  57  
7    43  52  56  60  54  71  40  58  
8    51  35  55  48  54  45  45  49  
9    50  48  42  44  38  52  58  40
```


Results

```
1 p |>
2   tidybayes::gather_draws(y[i]) |>
3   mutate(week = i-1) |>
4   filter(.chain == 1) |>
5   ggplot(aes(x=week, y=.value)) +
6     geom_line(data=temp, aes(y=avg_temp), color="darkred") +
7     tidybayes::stat_lineribbon(alpha=0.25)
```

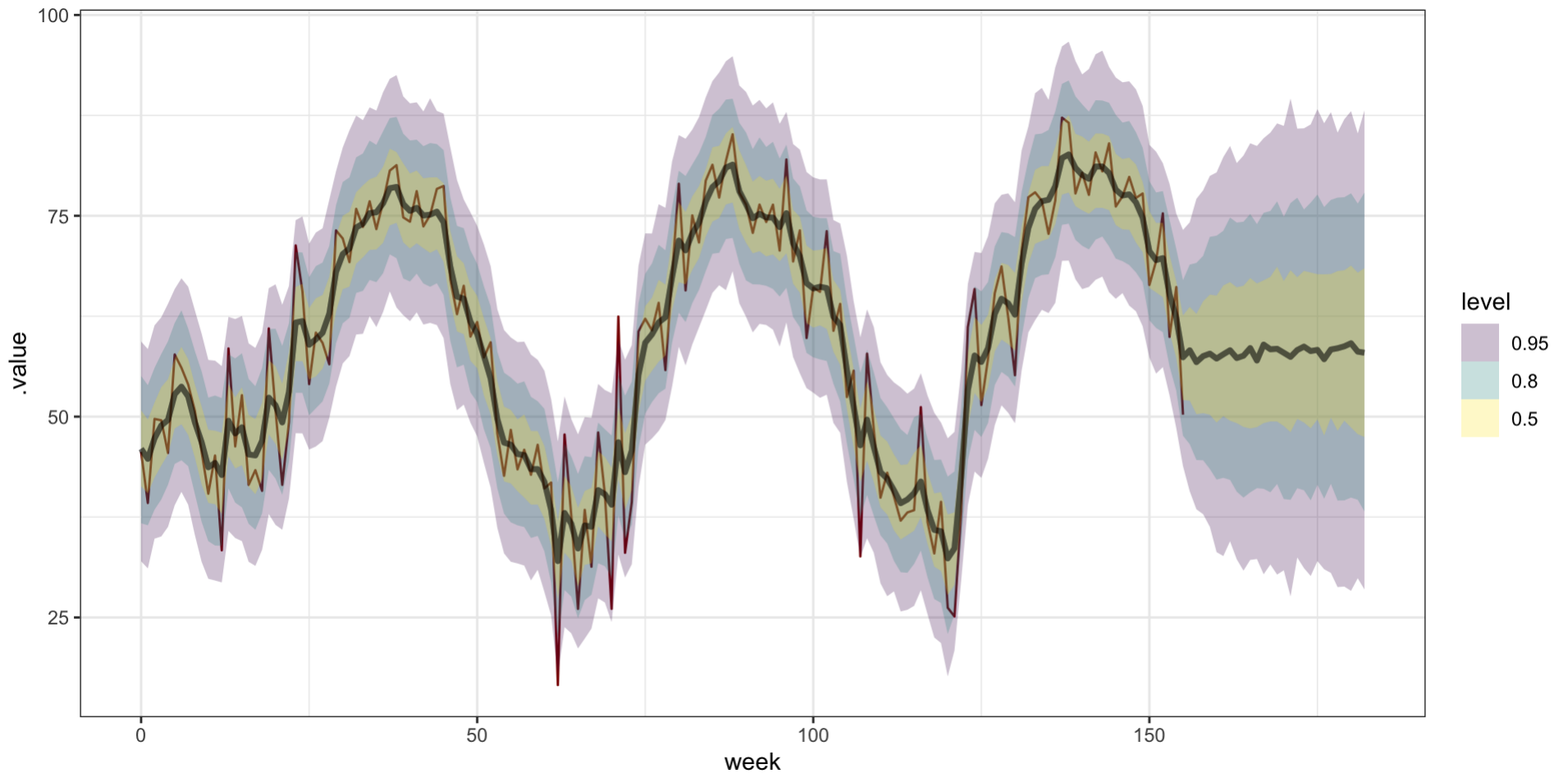


Exponential covariance

```
1 m_exp = gplm(  
2   avg_temp~1, data = temp, coords = "week",  
3   cov_model = "exponential",  
4   starting=list(  
5     "phi"=3/4, "sigma.sq"=1, "tau.sq"=1  
6   ),  
7   tuning=list(  
8     "phi"=1, "sigma.sq"=1, "tau.sq"=1  
9   ),  
10  priors=list(  
11    "phi.unif"=c(3/52, 3/1),  
12    "sigma.sq.ig"=c(2, 1),  
13    "tau.sq.ig"=c(2, 1)  
14  ),  
15  thin = 5,  
16  n_batch = 100,  
17  batch_len = 100  
18 )  
19  
20 p_exp = predict(m_exp, newdata=newdata, coords = "week")
```

Results

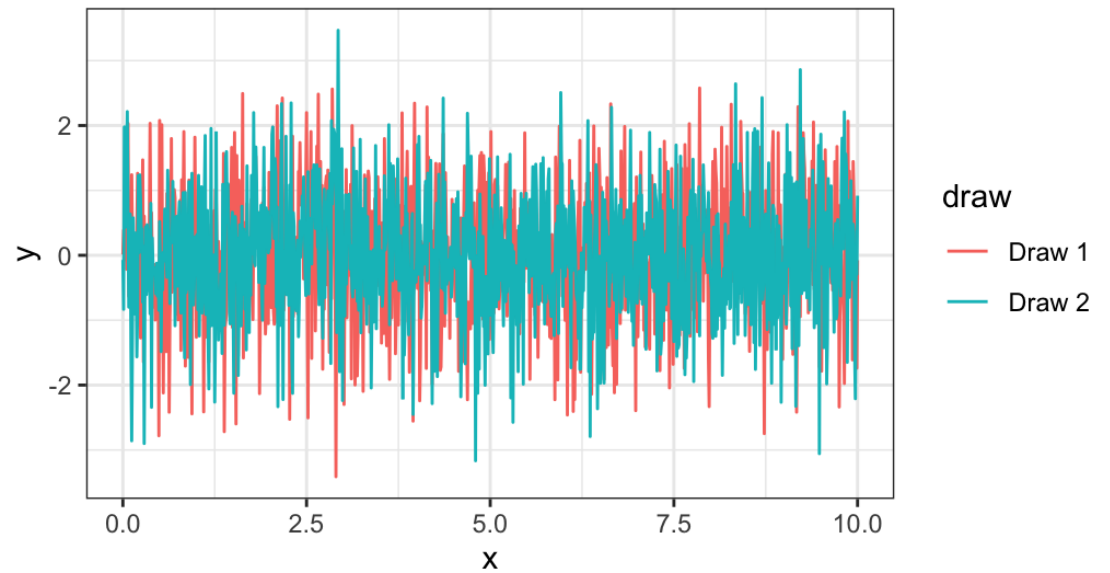
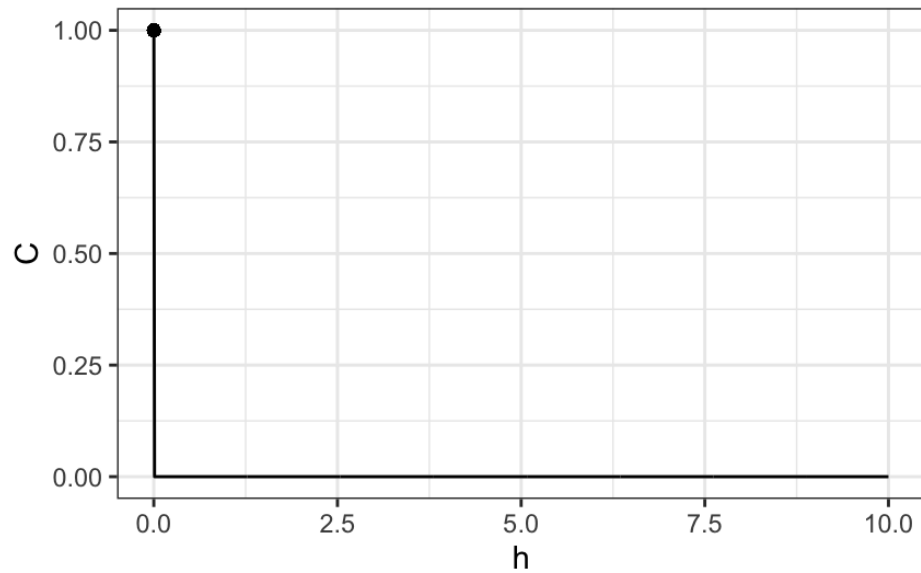
```
1 p_exp |>
2   tidybayes::gather_draws(y[i]) |>
3   mutate(week = i-1) |>
4   filter(.chain == 1) |>
5   ggplot(aes(x=week, y=.value)) +
6     geom_line(data=temp, aes(y=avg_temp), color="darkred") +
7     tidybayes::stat_lineribbon(alpha=0.25)
```



More Covariance Functions

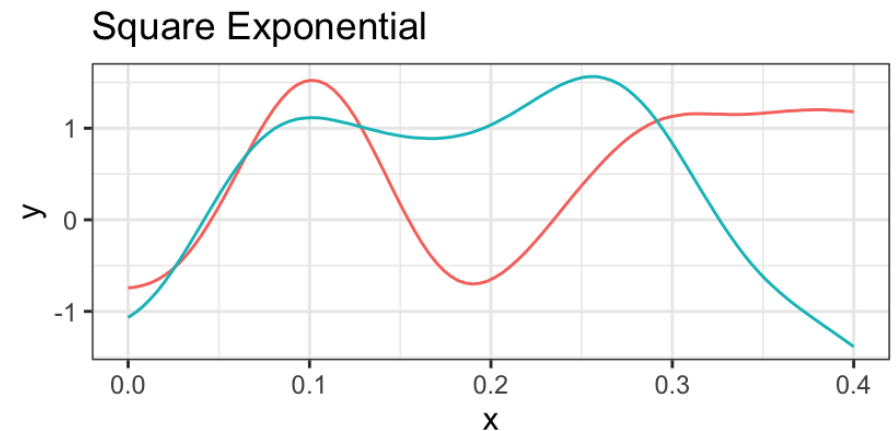
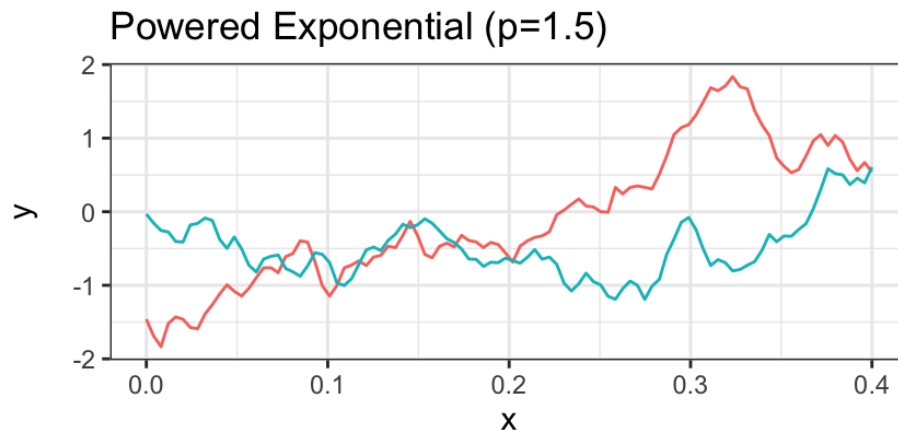
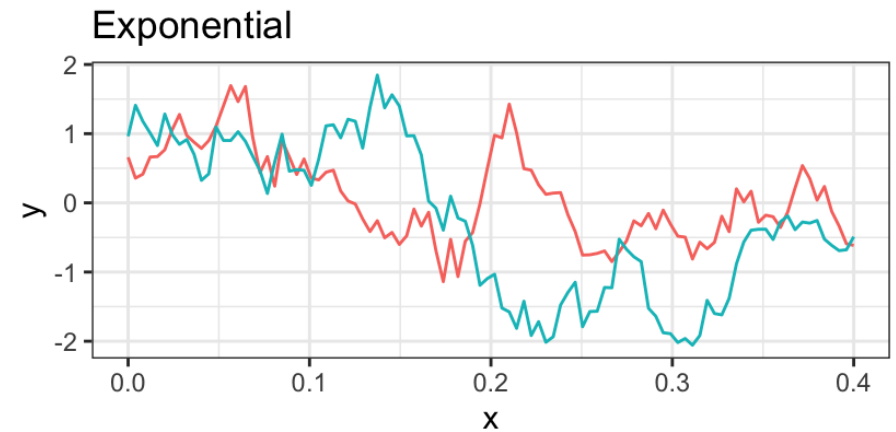
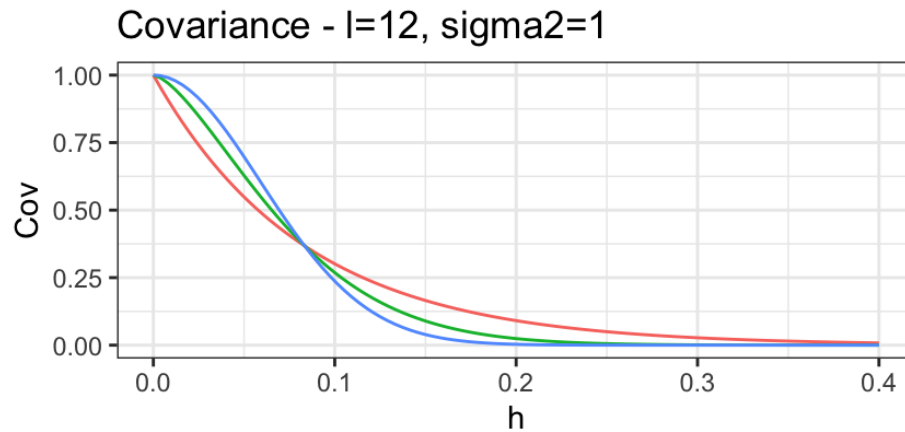
Nugget Covariance

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



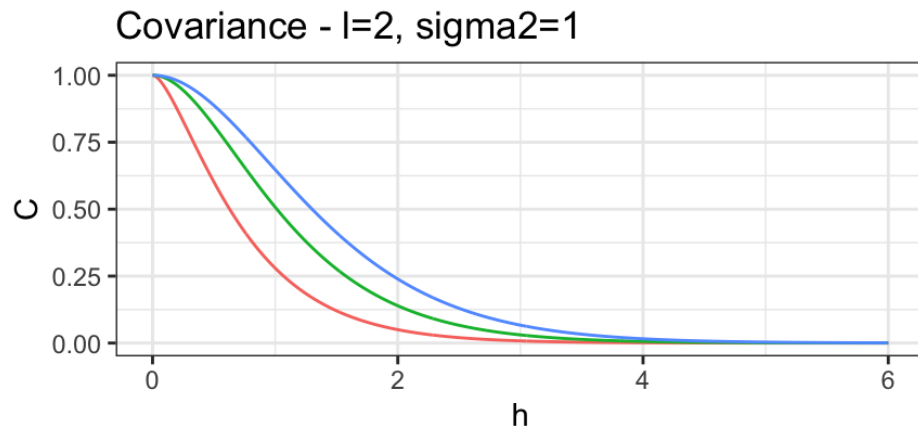
(- / Power / Square) Exponential Covariance

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

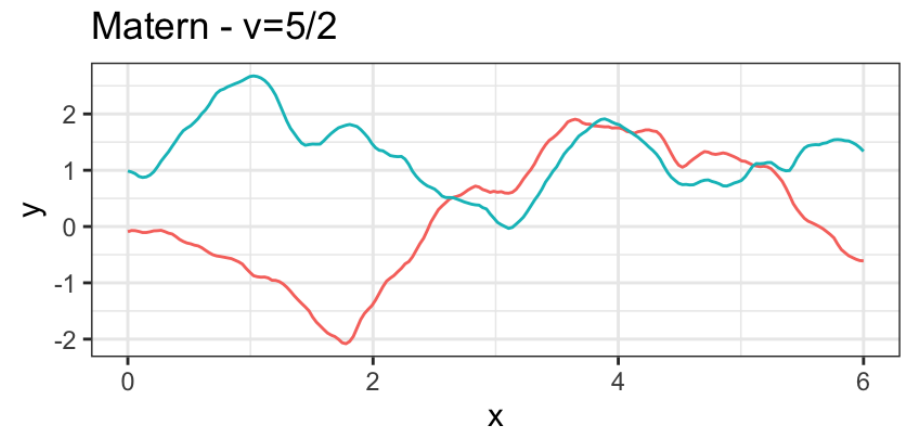
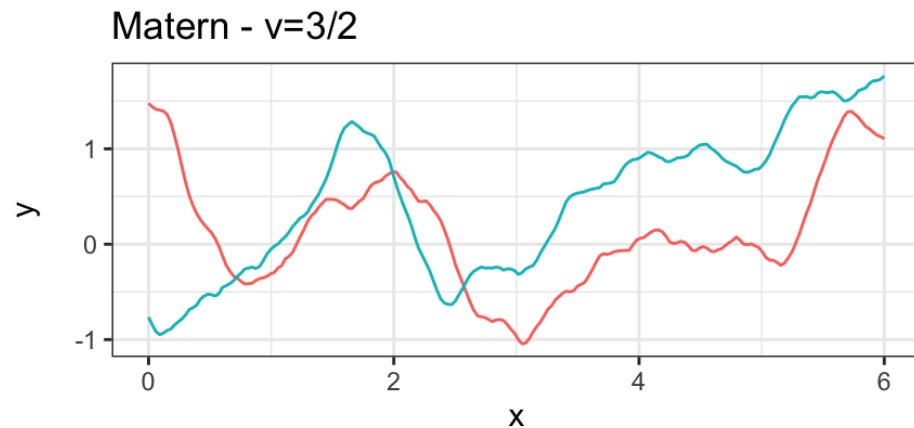
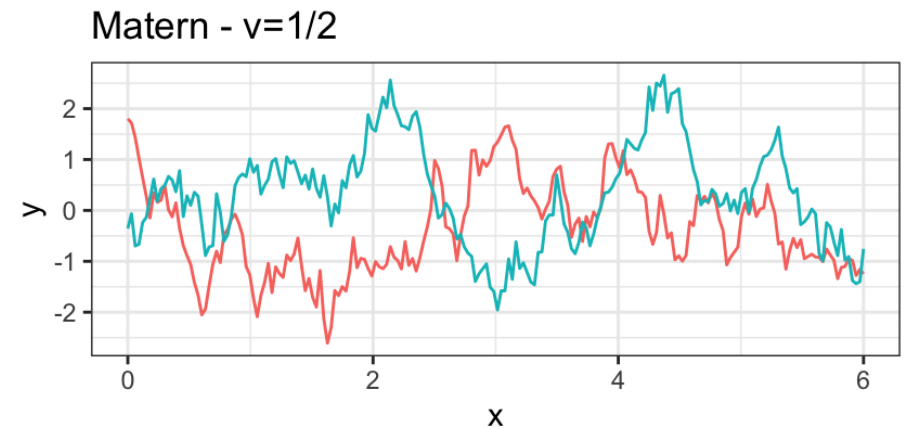


Matern Covariance

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



— v=1/2
— v=3/2
— v=5/2



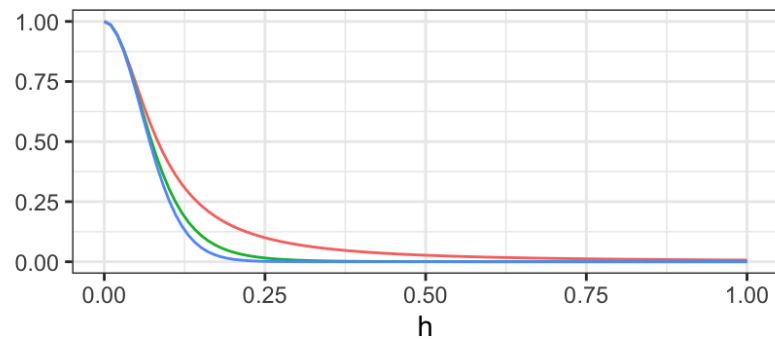
Matern Covariance

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

Rational Quadratic Covariance

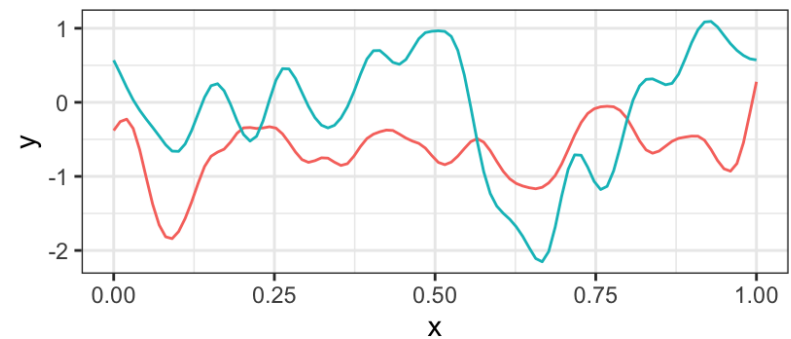
$$\text{Cov}(y_{t_i}, y_{t_j}) = \sigma^2 \left(1 + \frac{h^2 l^2}{\alpha} \right)^{-\alpha} \quad \text{where } h = |t_i - t_j|$$

Covariance - l=12, sigma2=1

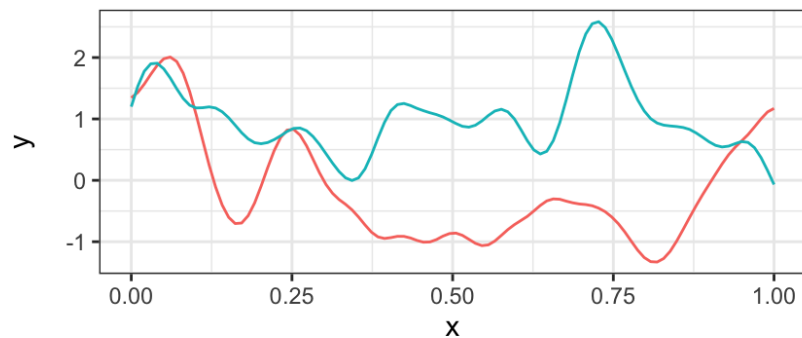


— alpha=1
— alpha=3
— alpha=10

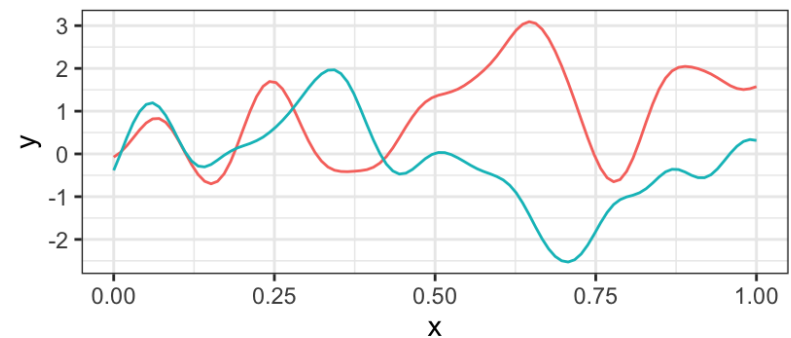
Rational Quadratic - alpha=1



Rational Quadratic - alpha=3



Rational Quadratic - alpha=10

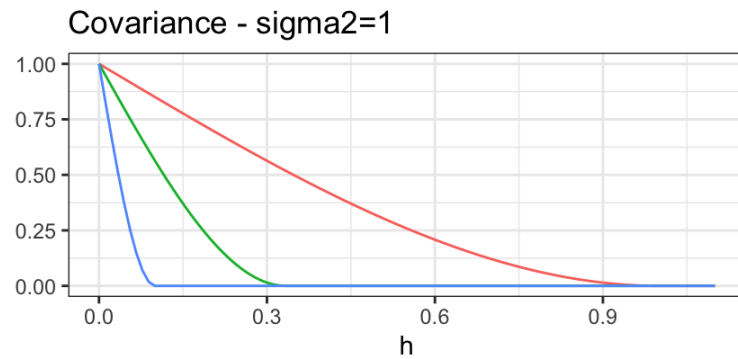


Rational Quadratic Covariance

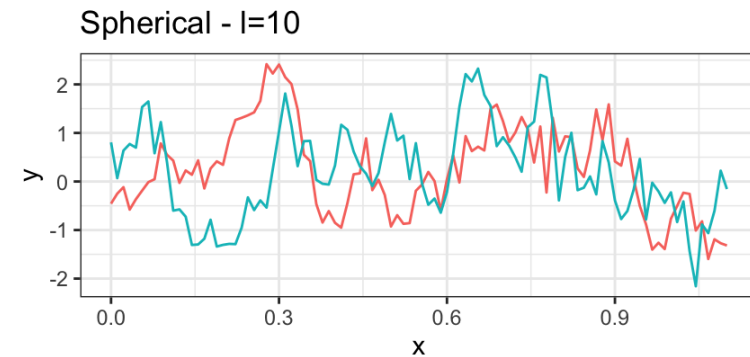
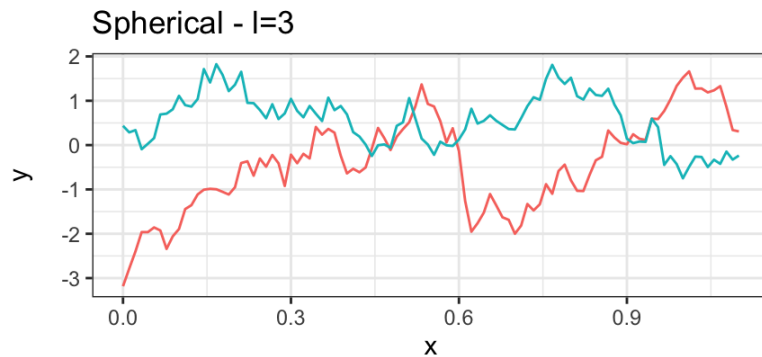
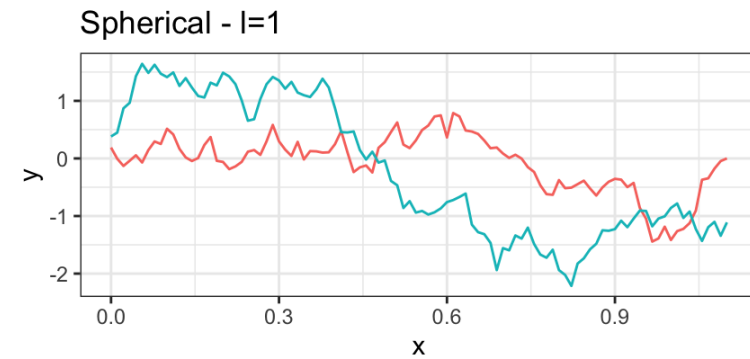
- is a scaled mixture of squared exponential covariance functions with different characteristic length-scales (l).
- As $\alpha \rightarrow \infty$ the rational quadratic converges to the square exponential covariance.
- Has sample functions that are infinitely differentiable for any value of α

Spherical Covariance

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



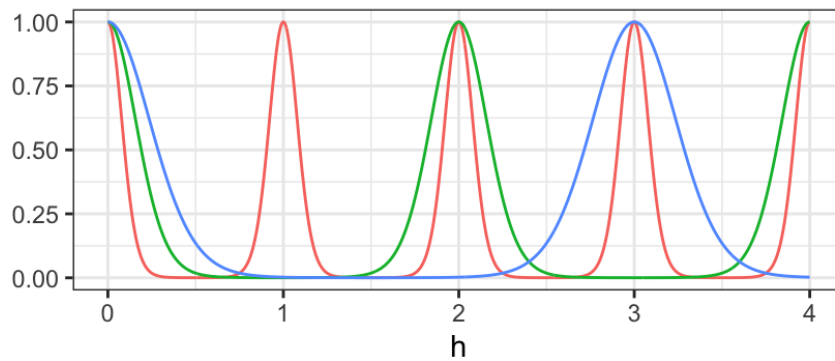
— l=1
— l=3
— l=10



Periodic Covariance

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

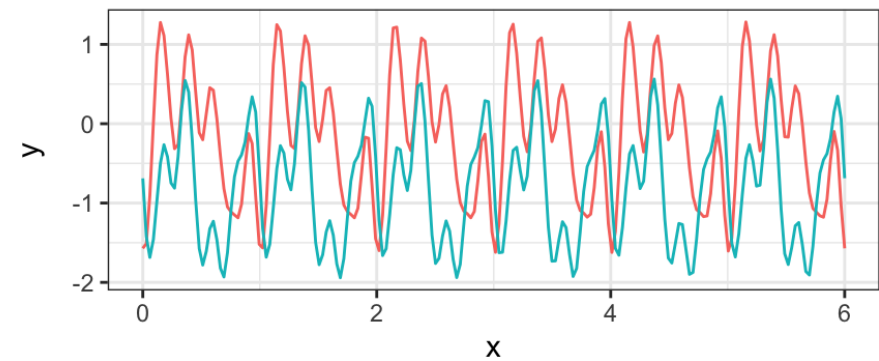
Covariance - l=2, sigma2=1



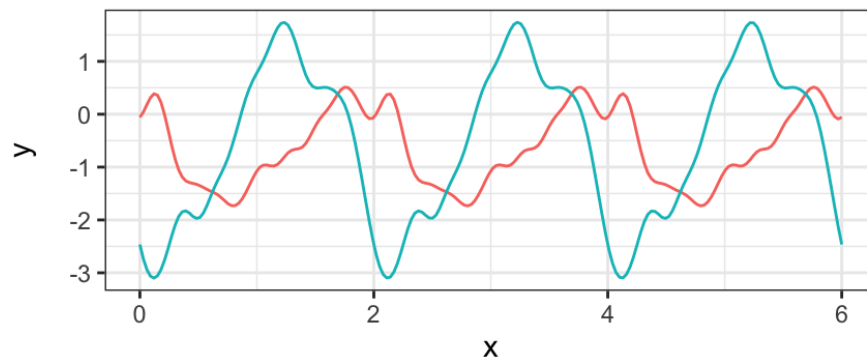
Period

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

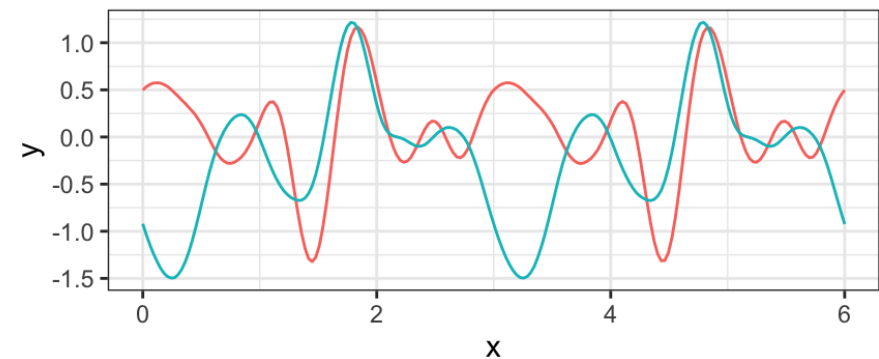
Periodic - p=1



Periodic - p=2

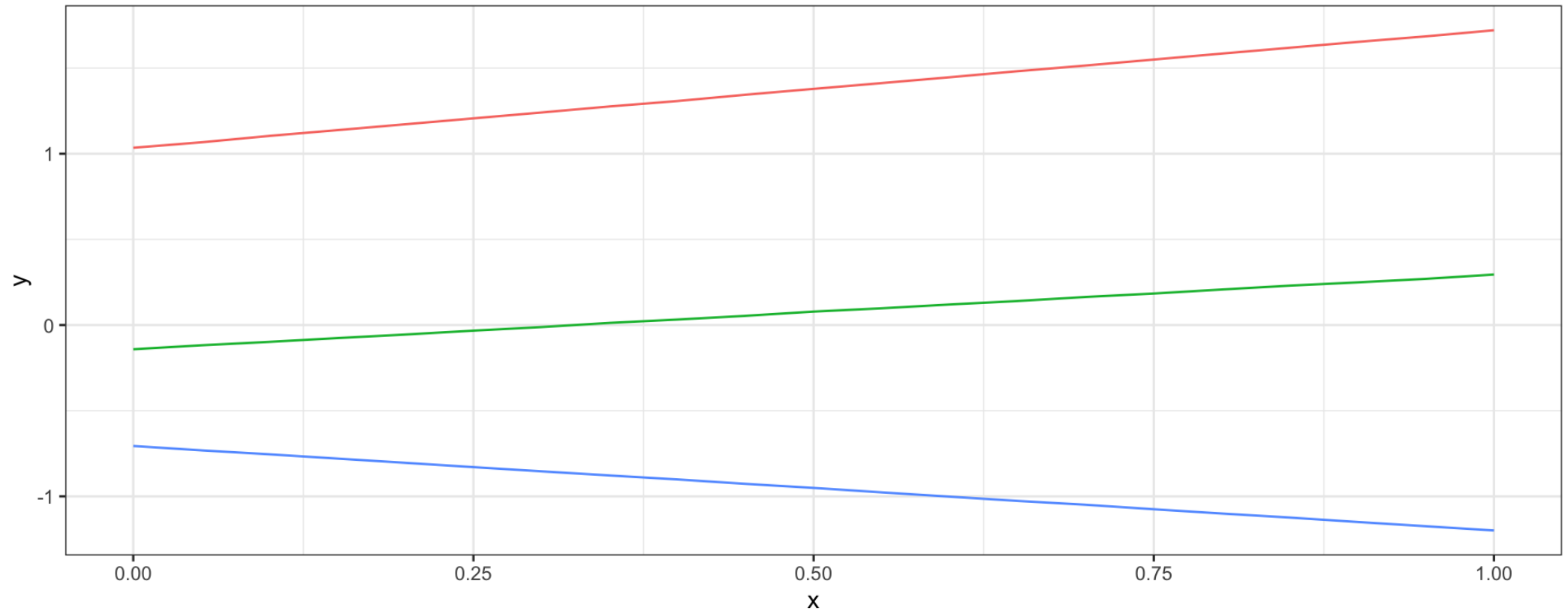


Periodic - p=3



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

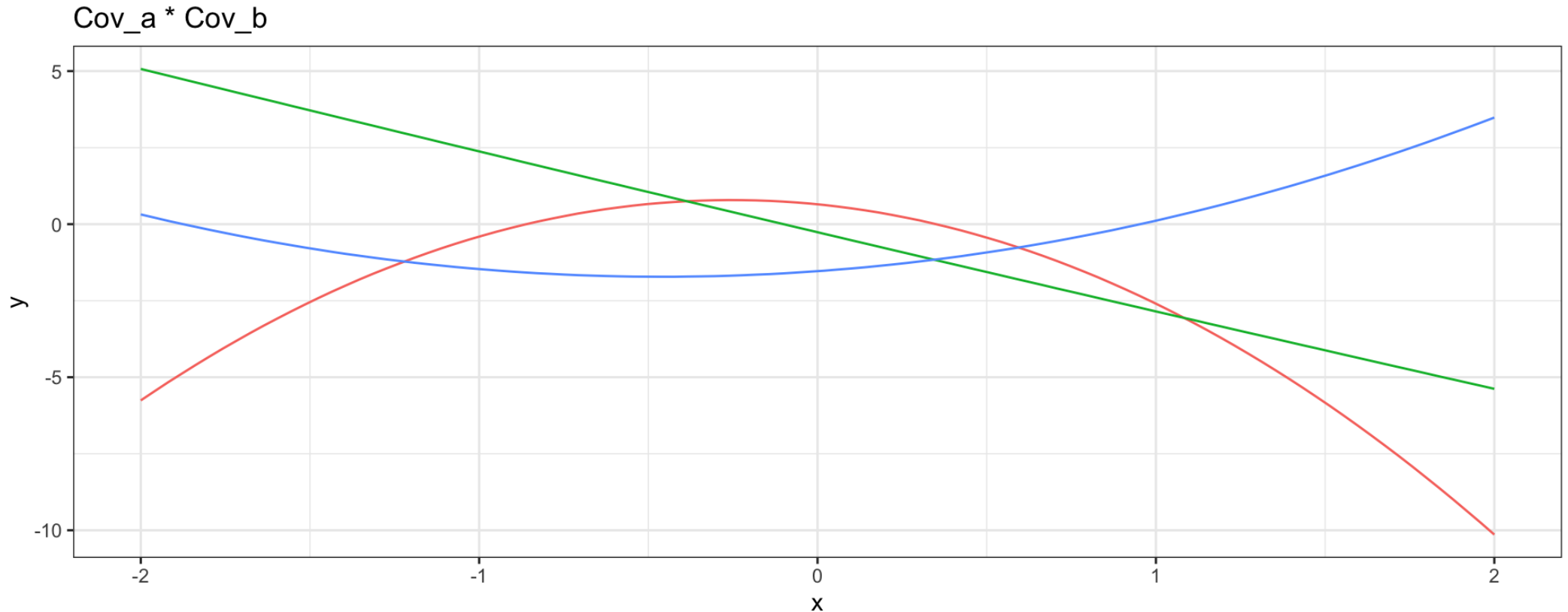
$$\text{Cov}_a(y_{t_i}, y_{t_j}) + \text{Cov}_b(y_{t_i}, y_{t_j})$$

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

Linear \times Linear \rightarrow Quadratic

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

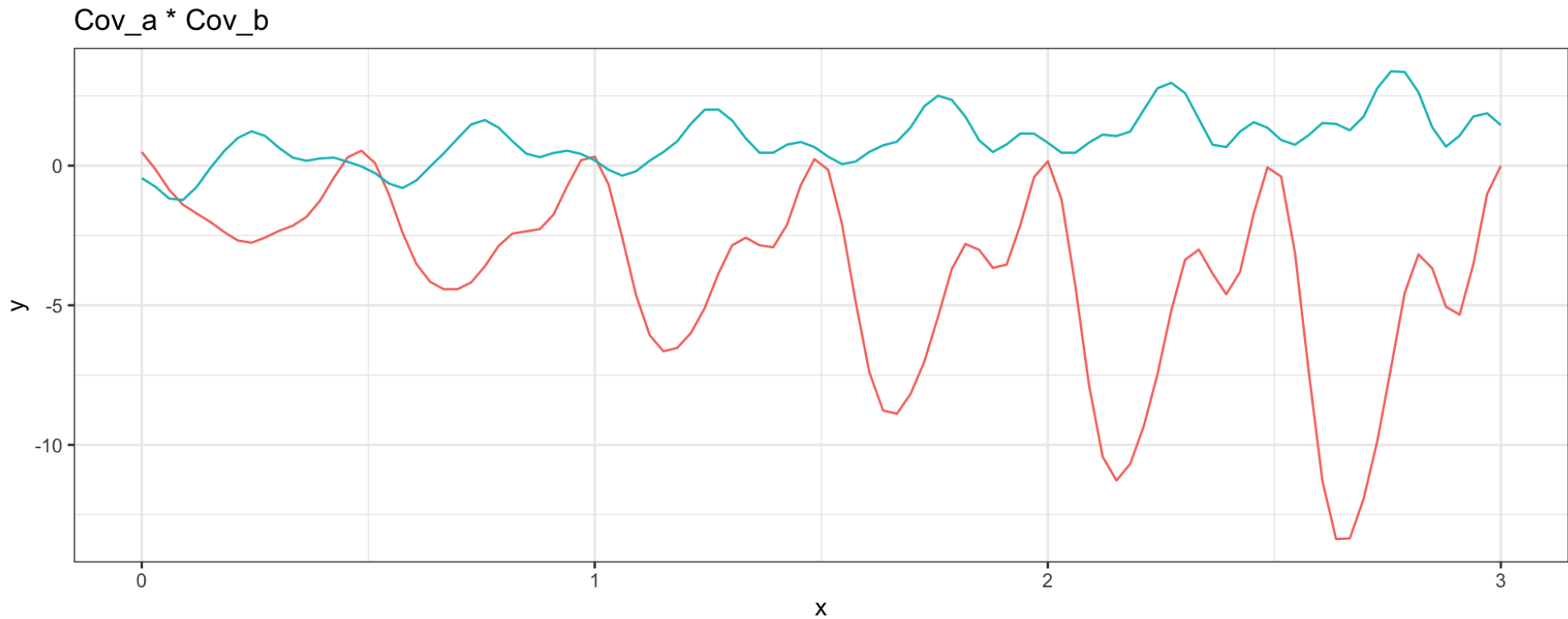
$$\text{Cov}_b(y_{t_i}, y_{t_j}) = 2 + 1(t_i \times t_j)$$



Linear \times Periodic

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

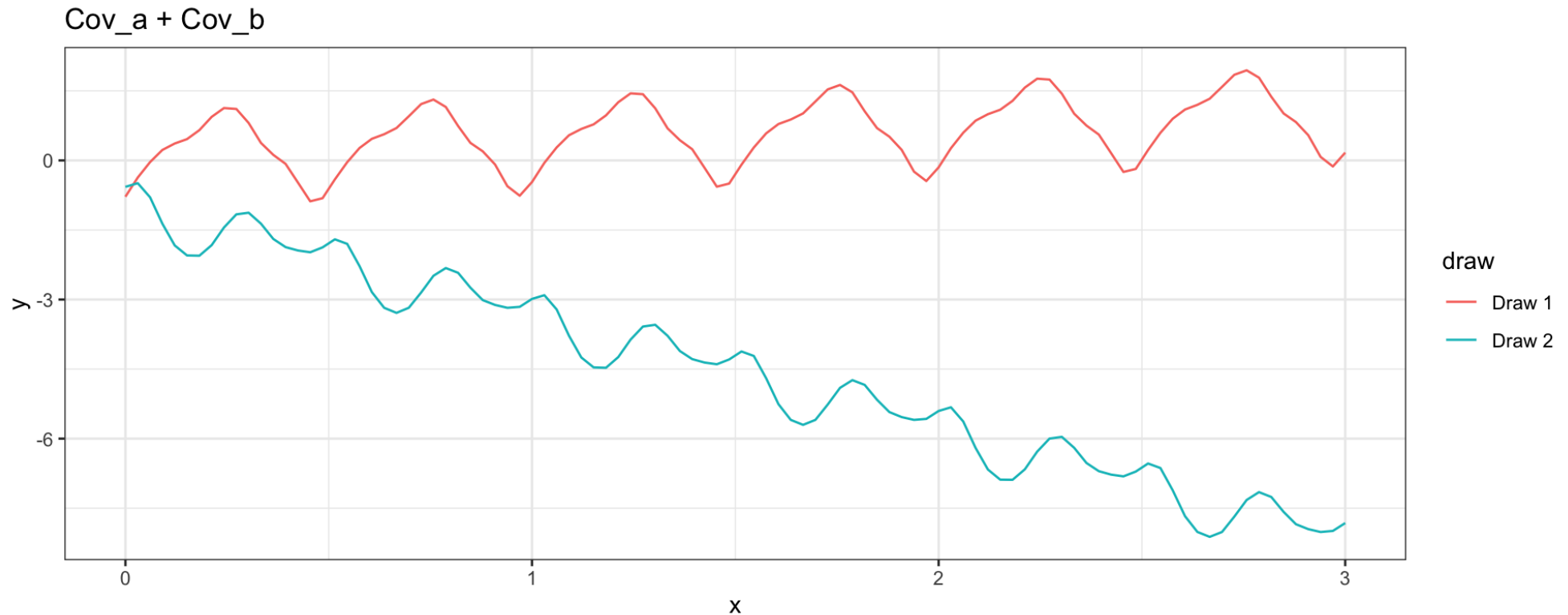
$$\text{Cov}_b(y_{t_i}, y_{t_j}) = \exp(-2 \sin^2(2\pi h))$$



Linear + Periodic

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

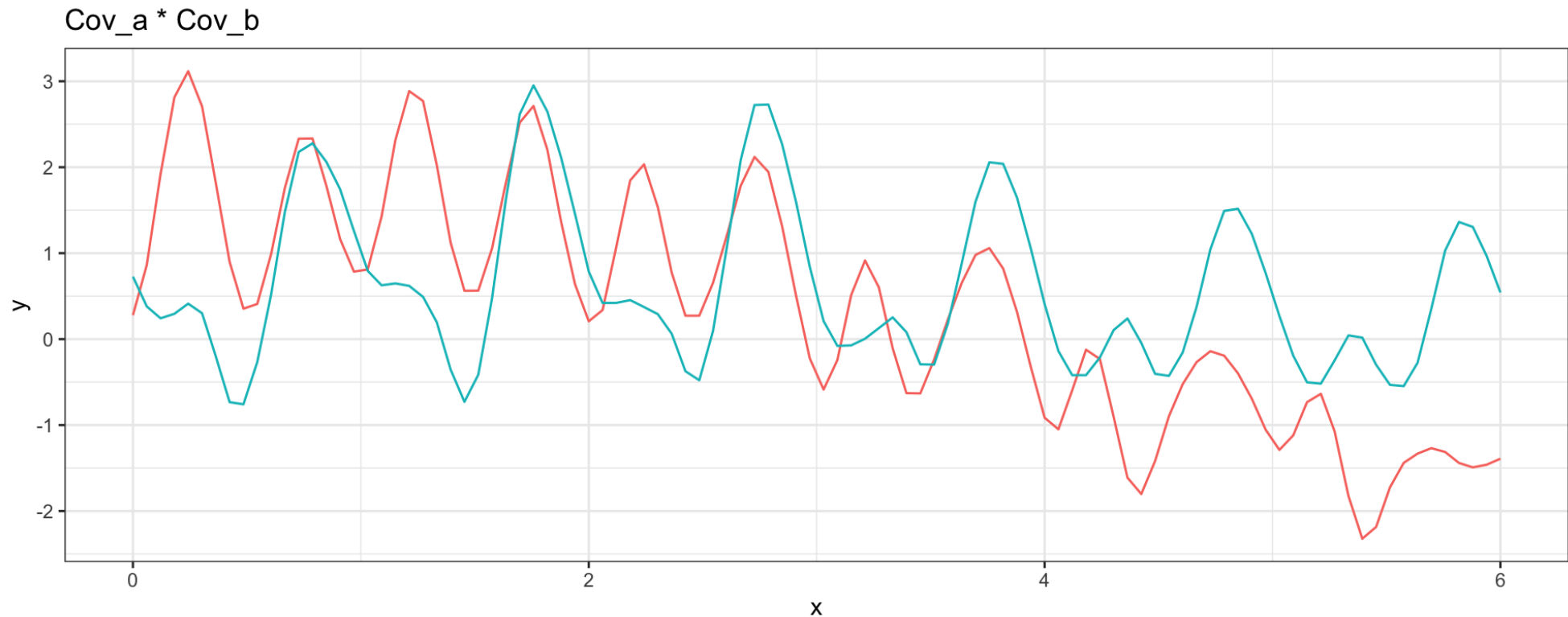
$$\text{Cov}_b(h = |t_i - t_j|) = \exp(-2 \sin^2(2\pi h))$$



Sq Exp \times Periodic \rightarrow Locally Periodic

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

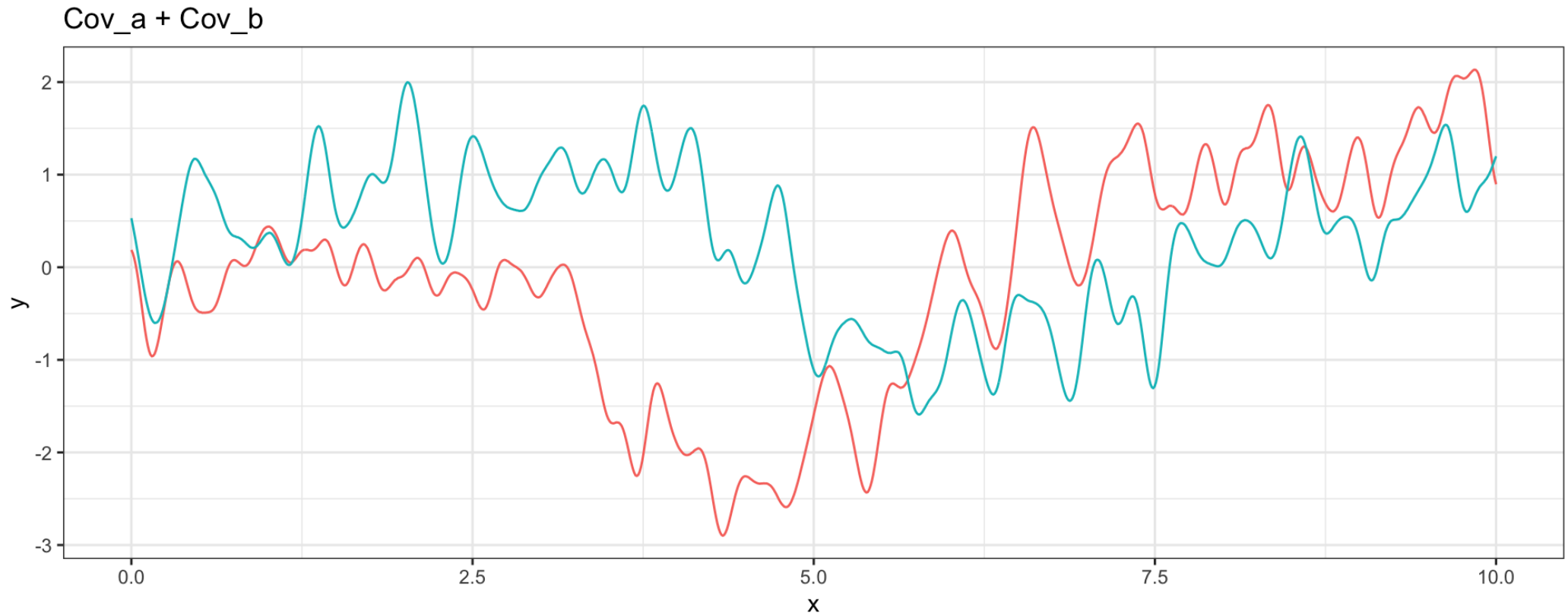
$$\text{Cov}_b(h = |t_i - t_j|) = \exp(-2 \sin^2(\pi h))$$



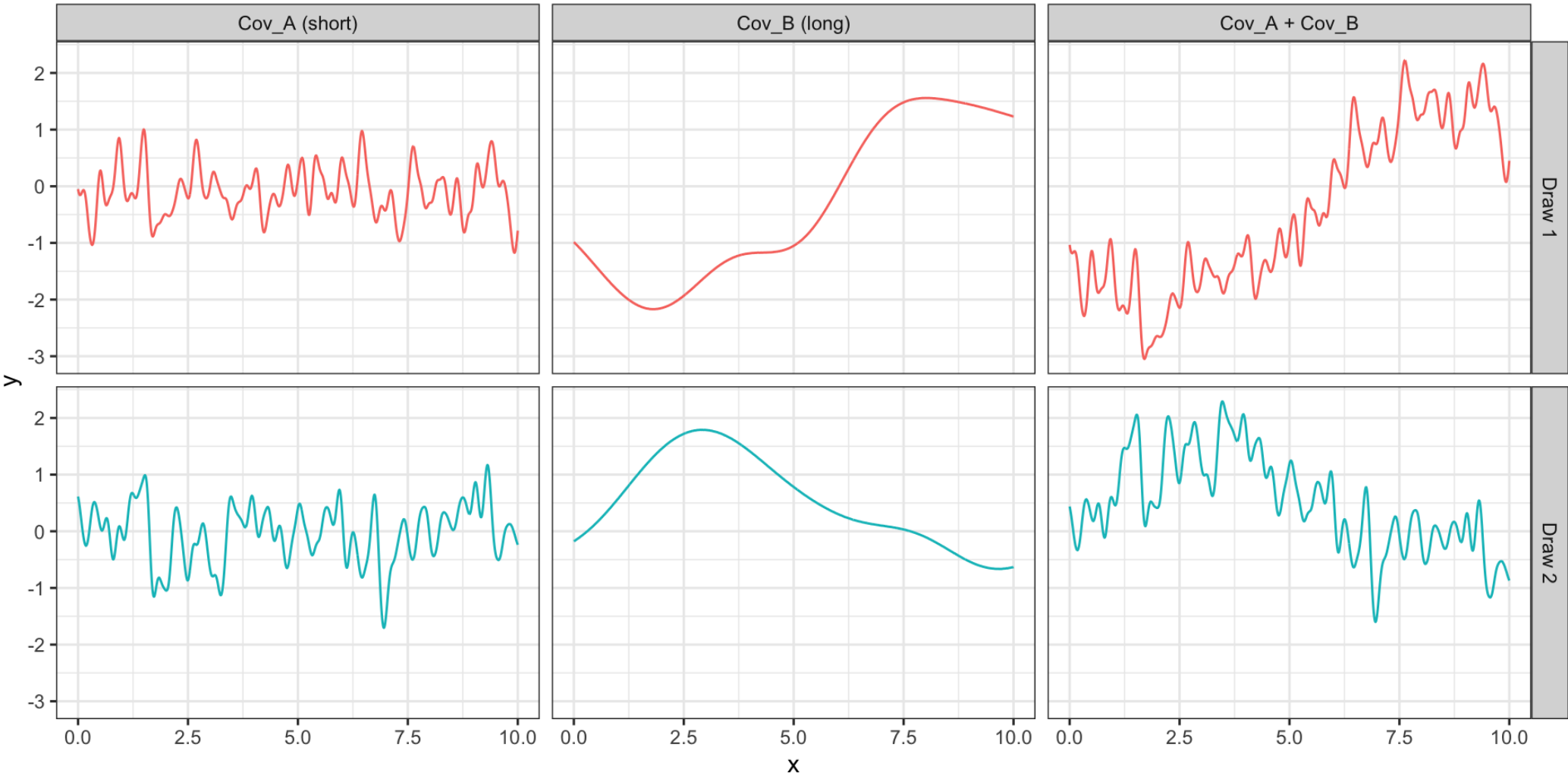
Sq Exp (short) + Sq Exp (long)

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

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

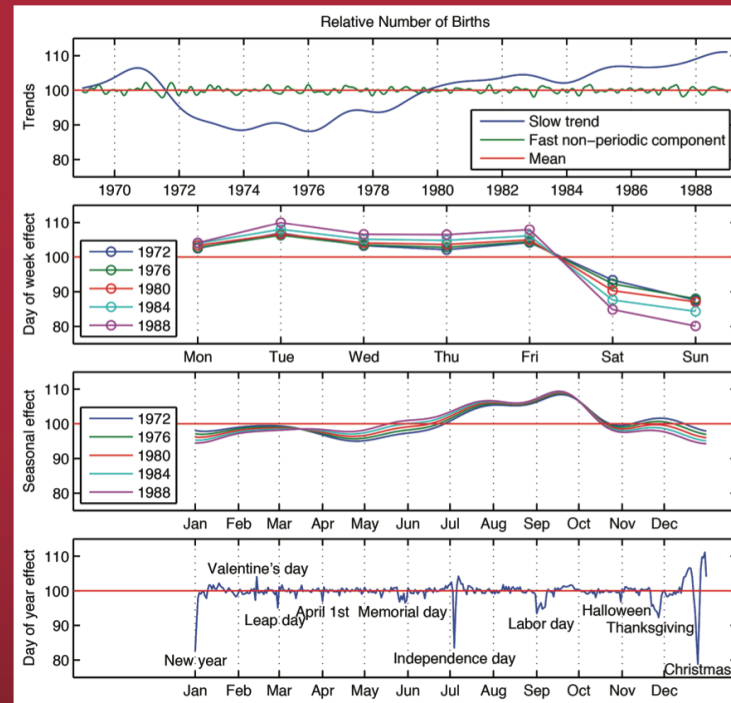


Seen another way



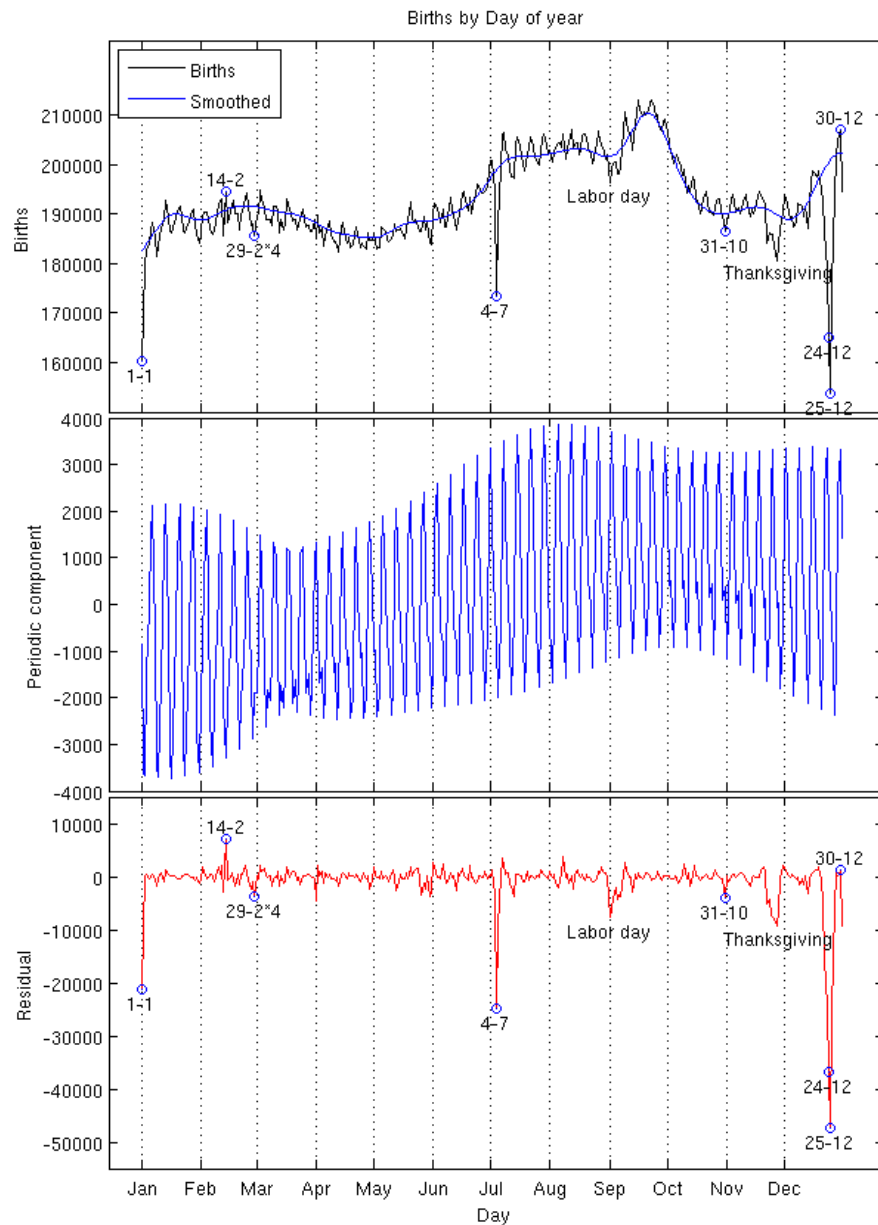
BDA3 example

Bayesian Data Analysis Third Edition



Andrew Gelman, John B. Carlin, Hal S. Stern,
David B. Dunson, Aki Vehtari, and Donald B. Rubin

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

Component Contributions

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

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

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

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

where

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

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

$$\mathbf{w}_3(\mathbf{t}) \sim \mathbf{N}(0, \sigma^2 \mathbf{I})$$

Decomposition of Covariance Components

$$\begin{bmatrix} y \\ w_1 \\ w_2 \end{bmatrix} \sim \mathbf{N} \left(\begin{bmatrix} \boldsymbol{\mu} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \sigma^2 \mathbf{I} & \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_2 \\ & \boldsymbol{\Sigma}_1 & 0 \\ & & \boldsymbol{\Sigma}_2 \end{bmatrix} \right)$$

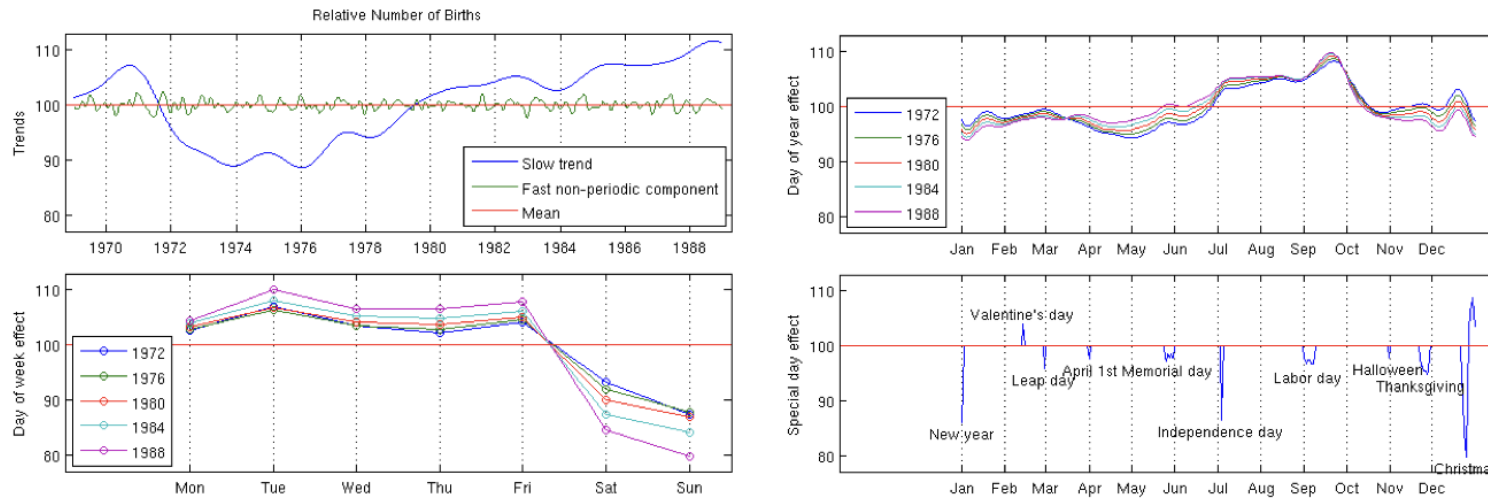
therefore

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

$$\boldsymbol{\mu}_{\text{cond}} = 0 + \boldsymbol{\Sigma}_1 (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

$$\boldsymbol{\Sigma}_{\text{cond}} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1 (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \sigma^2 \mathbf{I})^{-1} \boldsymbol{\Sigma}_1^t$$

Births (multiple years)

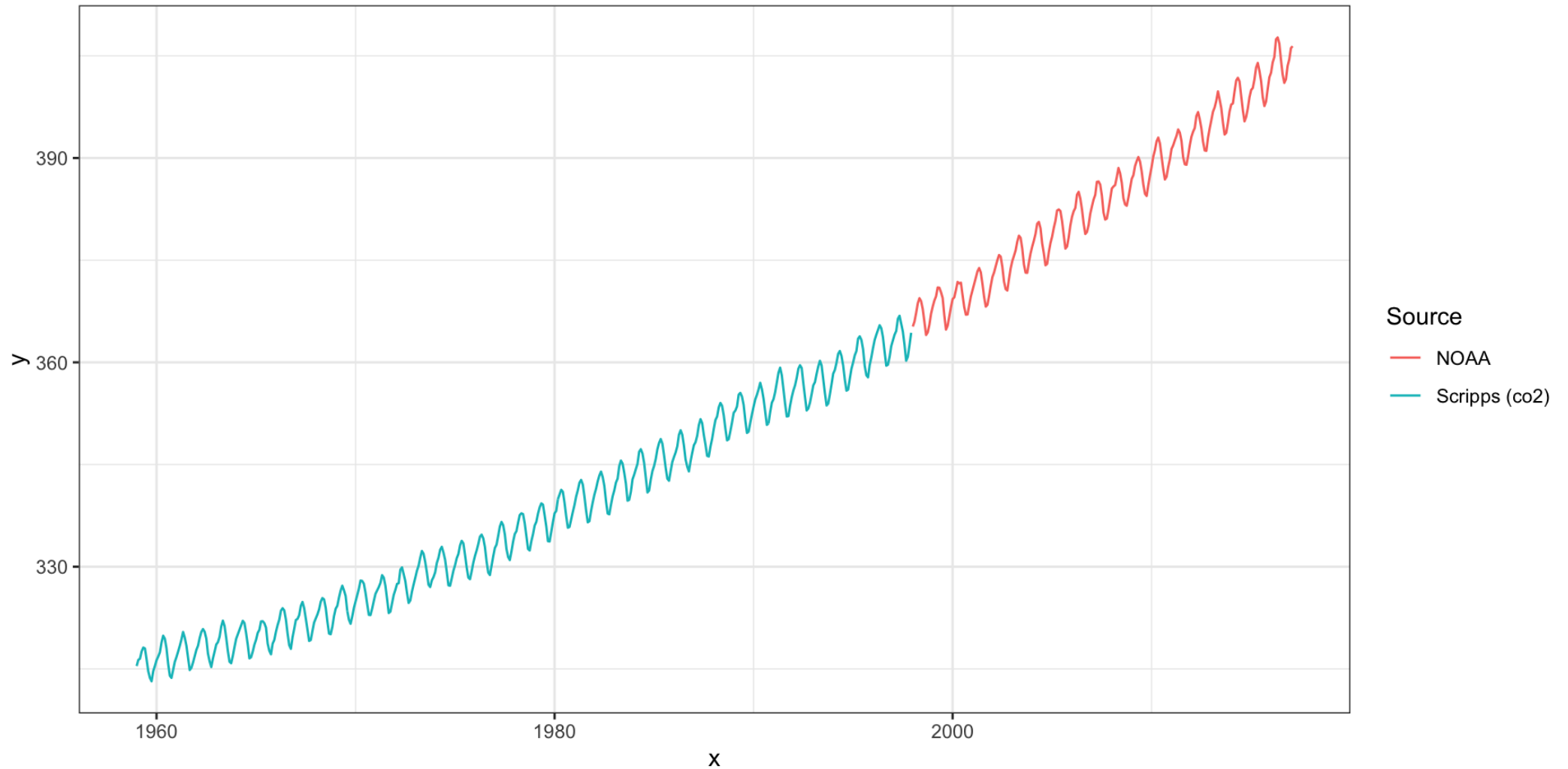


Full stan case study [here](#) with code [here](#)

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 CO₂



GP Model

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

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

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

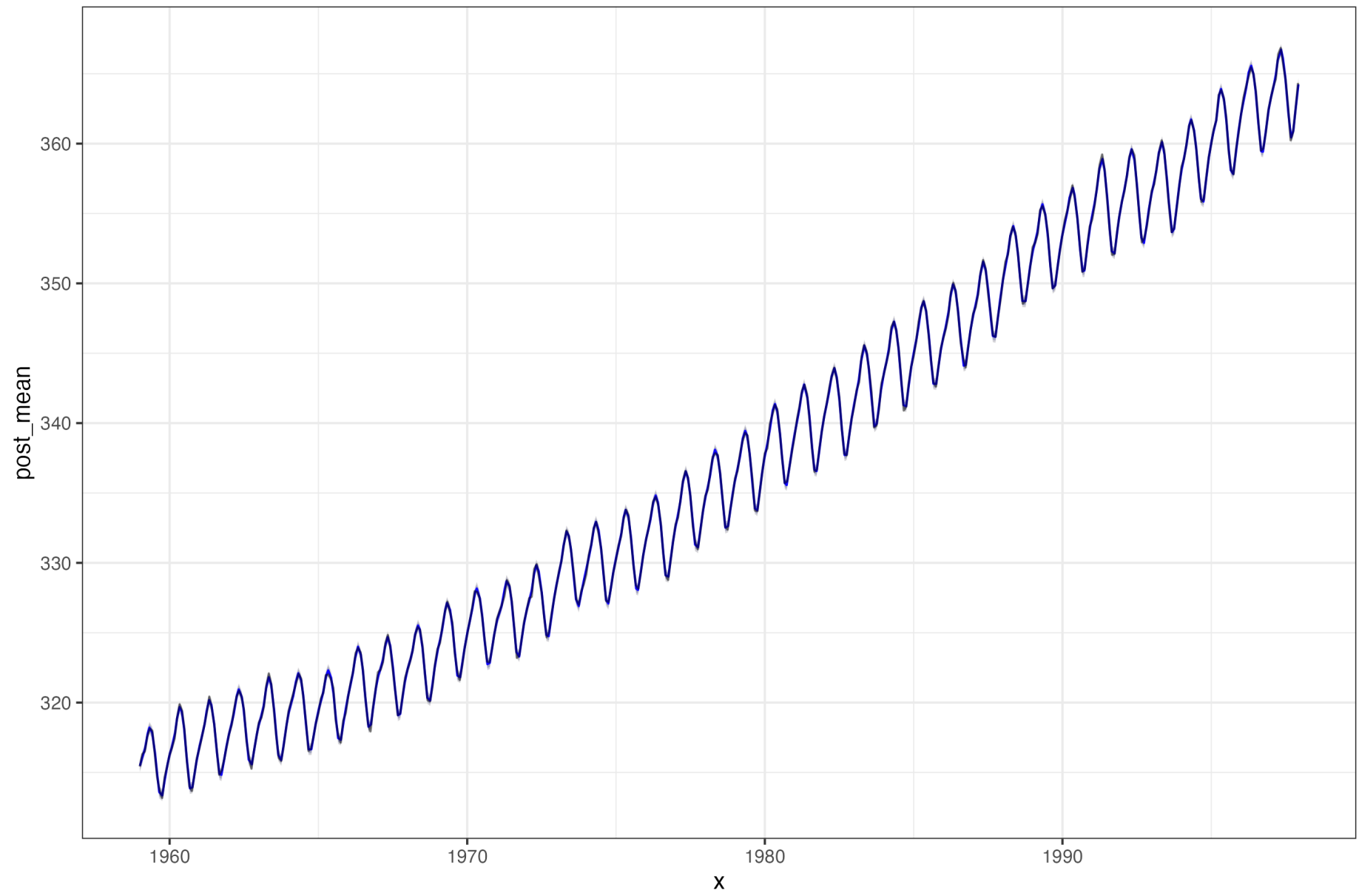
$$\{\boldsymbol{\Sigma}_1\}_{ij} = \sigma_1^2 \exp(-(l_1 \cdot d_{ij})^2)$$

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

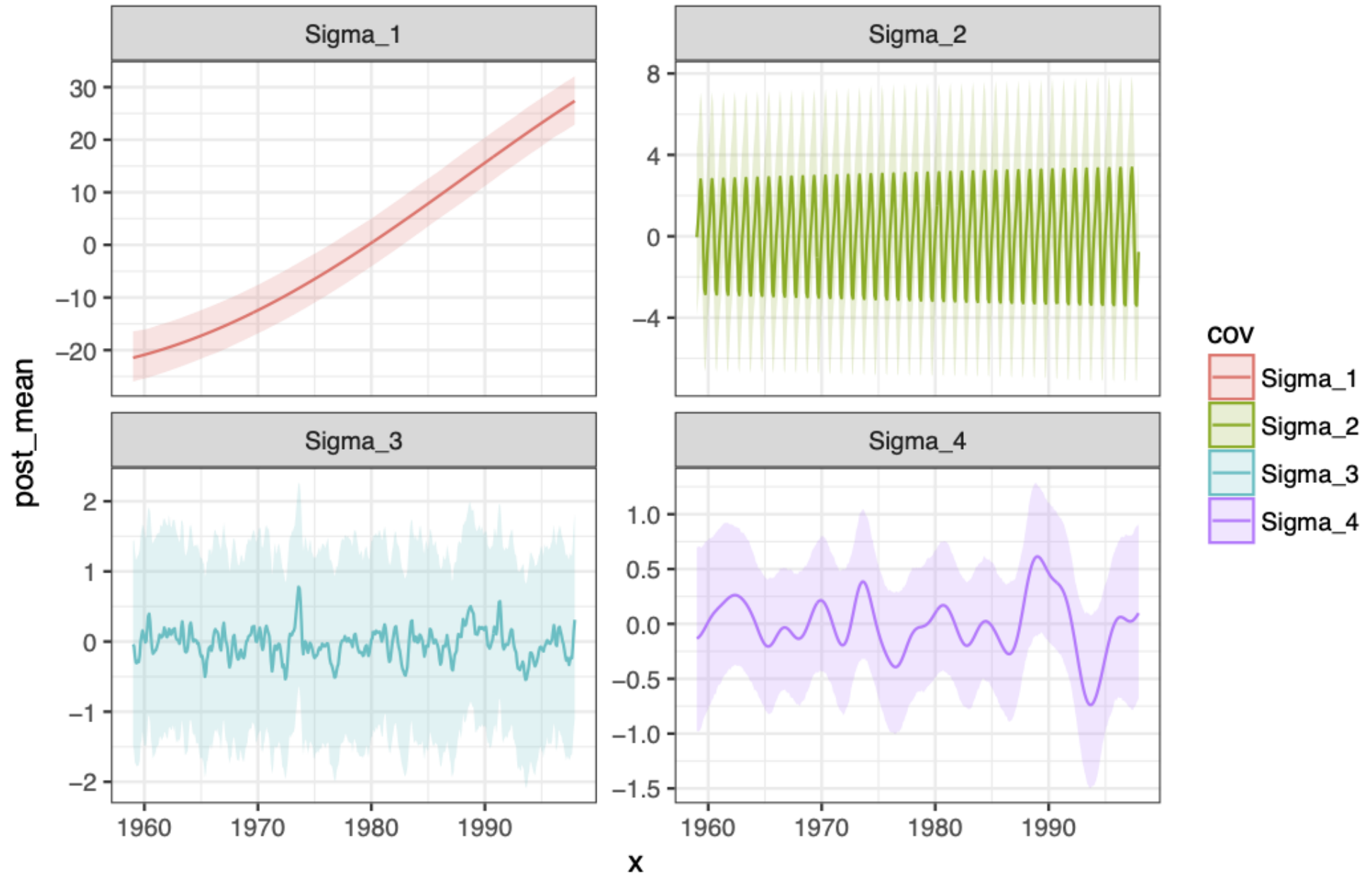
$$\{\boldsymbol{\Sigma}_3\}_{ij} = \sigma_3^2 \left(1 + \frac{(l_4 \cdot d_{ij})^2}{\alpha} \right)^{-\alpha}$$

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

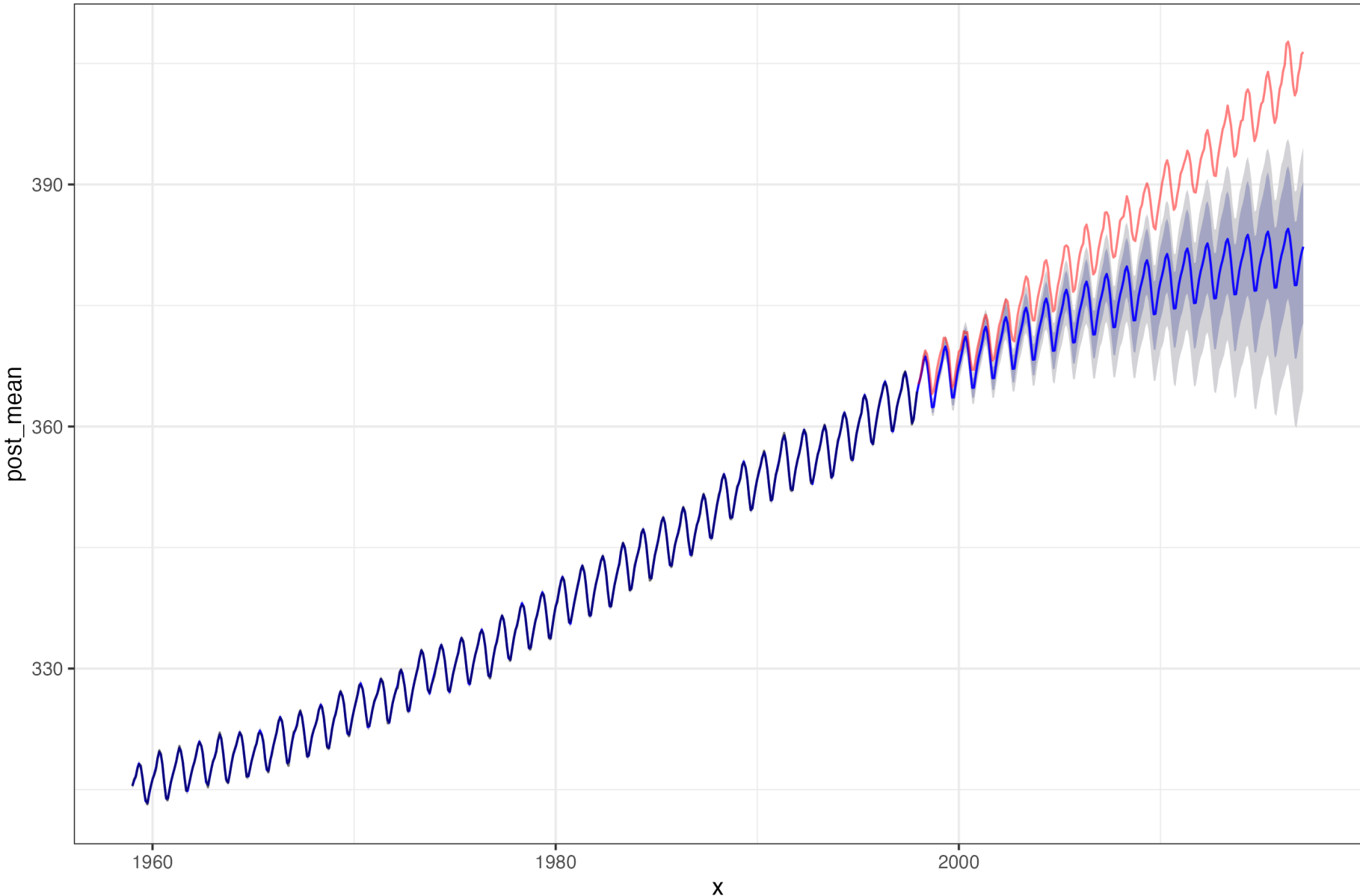
Model fit



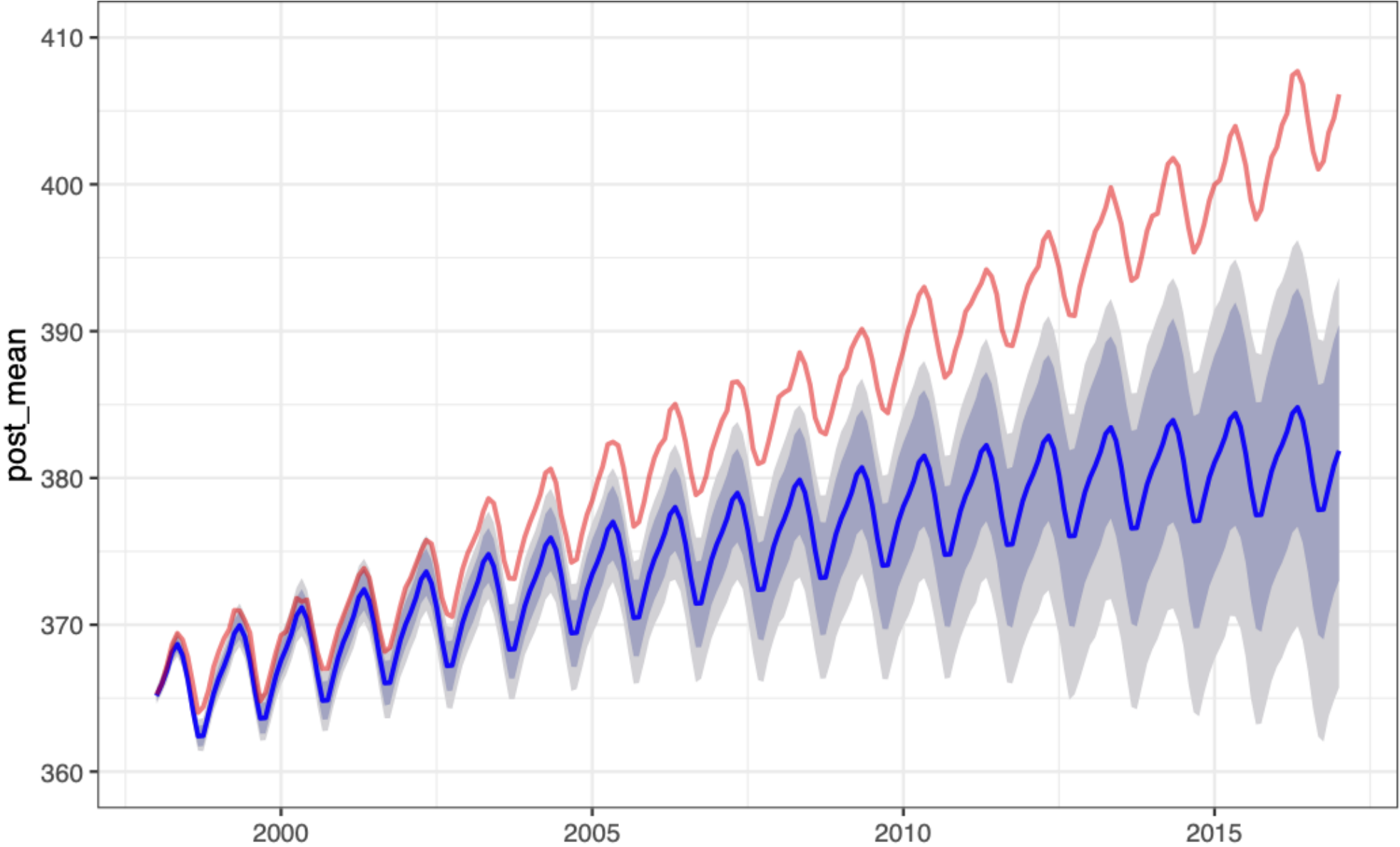
Fit Components



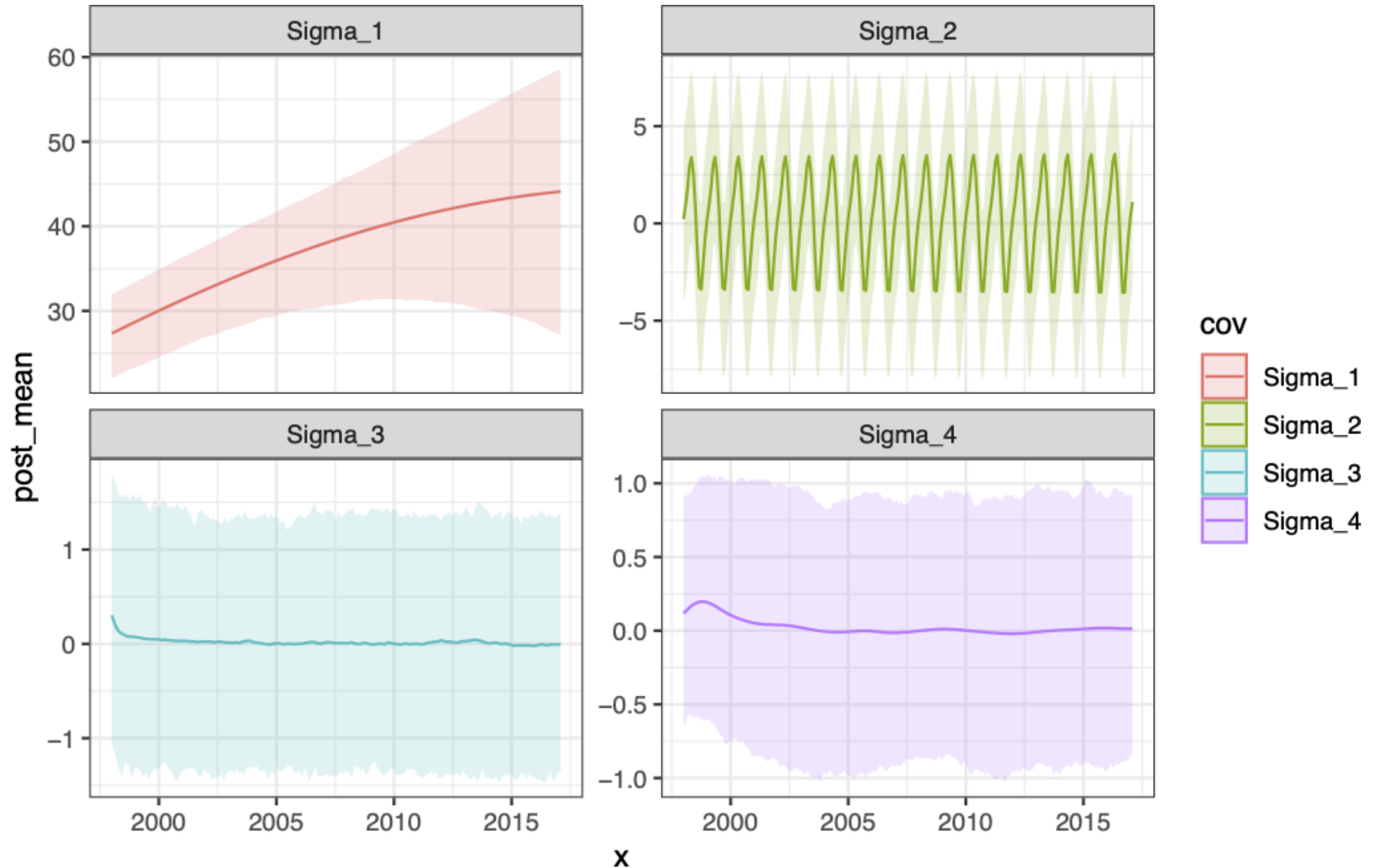
Model fit + forecast



Forecast

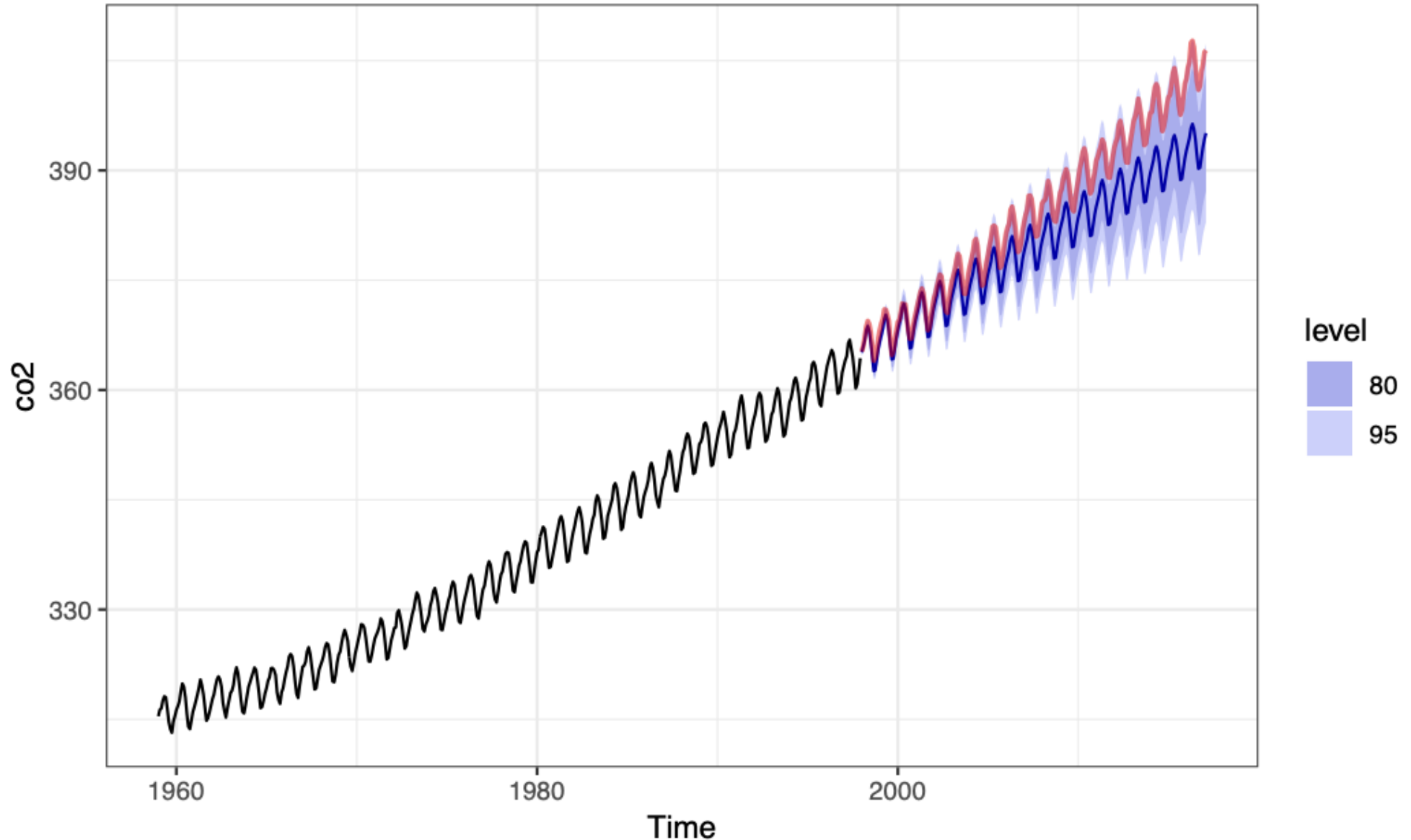


Forecast components



ARIMA forecast

Forecasts from ARIMA(1,1,1)(1,1,2)[12]



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