Lecture 14

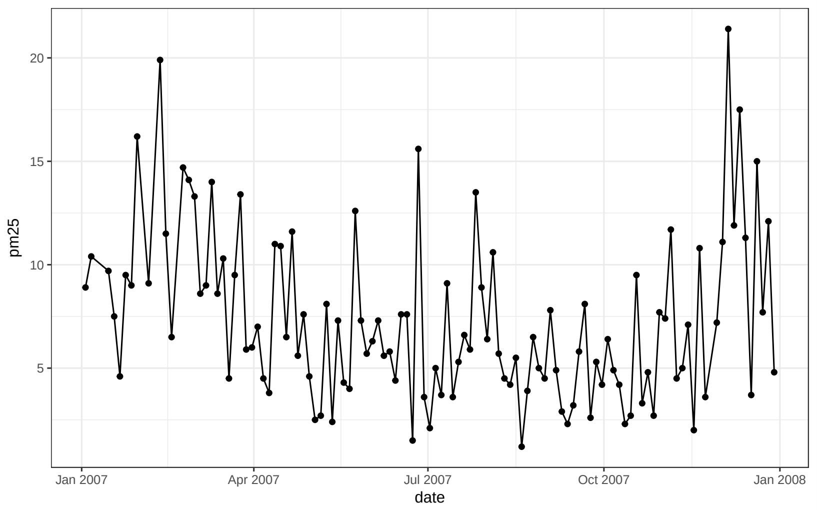
Full Posterior Pred & Covariance Functions

Colin Rundel

03/06/2017

Full Posterior Predictive Distribution

FRN Data



pm25
20
15
10
5
Jan 2007
Apr 2007
Jul 2007
Oct 2007
Jan 2008
date

```
## 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,)
## }
```

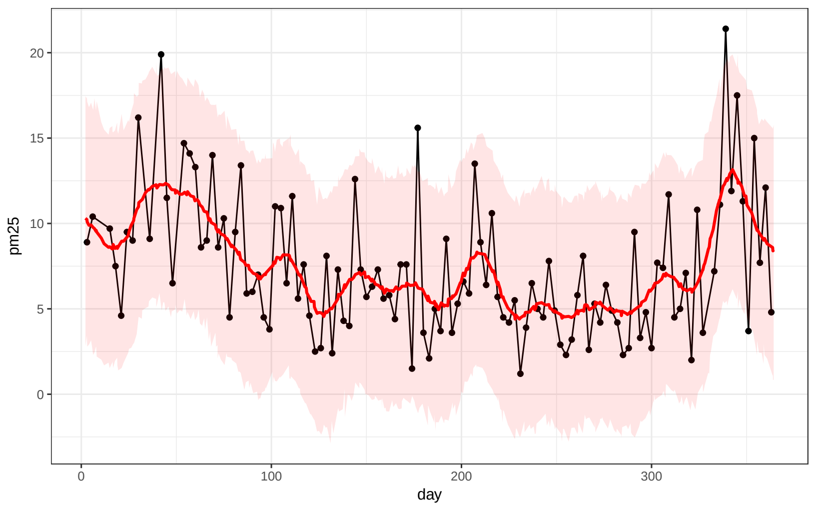
Posterior

param	post_mean	post_med	post_lower	post_upper
beta[1]	9.2136151	11.4359371	-0.4309078	15.2615892
beta[2]	-0.0361357	-0.0551308	-0.1012205	0.0849476
beta[3]	0.0001007	0.0001367	-0.0001924	0.0002552
l	0.8787410	0.0698553	0.0065124	7.0905582
sigma2	8.4807746	7.8609848	1.5342164	18.6524860
sigma2_w	9.7527513	10.4646243	2.2091857	14.8425142

Predicting

```
l = post %>% filter(param == 'l') %>% select(post_med) %>% unlist()
sigma2 = post %>% filter(param == 'sigma2') %>% select(post_med) %>% unlist()
sigma2_w = post %>% filter(param == 'sigma2_w') %>% select(post_med) %>% unlist()

beta0 = post %>% filter(param == 'beta[1]') %>% select(post_med) %>% unlist()
beta1 = post %>% filter(param == 'beta[2]') %>% select(post_med) %>% unlist()
beta2 = post %>% filter(param == 'beta[3]') %>% select(post_med) %>% unlist()

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 = rdist(x)
dist_p = rdist(x_pred)
dist_op = 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 = sigma2_w)
cov_p  = sq_exp_cov(dist_p,  sigma2 = sigma2, l = l) + nugget_cov(dist_p, sigma2 = sigma2_w)
cov_op = sq_exp_cov(dist_op, sigma2 = sigma2, l = l) + nugget_cov(dist_op, sigma2 = sigma2_w)
cov_po = sq_exp_cov(dist_po, sigma2 = sigma2, l = l) + nugget_cov(dist_po, sigma2 = sigma2_w)

cond_cov = cov_p - cov_po %*% solve(cov_o) %*% cov_op
cond_mu  = mu_pred + cov_po %*% solve(cov_o) %*% (y - mu)

pred = cond_mu %*% matrix(1, ncol=reps) + t(chol(cond_cov)) %*% matrix(rnorm(length(x_pred)*reps), ncol=reps)

pred_df = pred %>% t() %>% post_summary() %>% mutate(day=x_pred)
```

Predictions



Full Posterior Predictive Distribution

Our posterior consists of samples from

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

and for the purposes of generating the posterior predictions we sampled

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

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

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

In practice we should instead be sampling

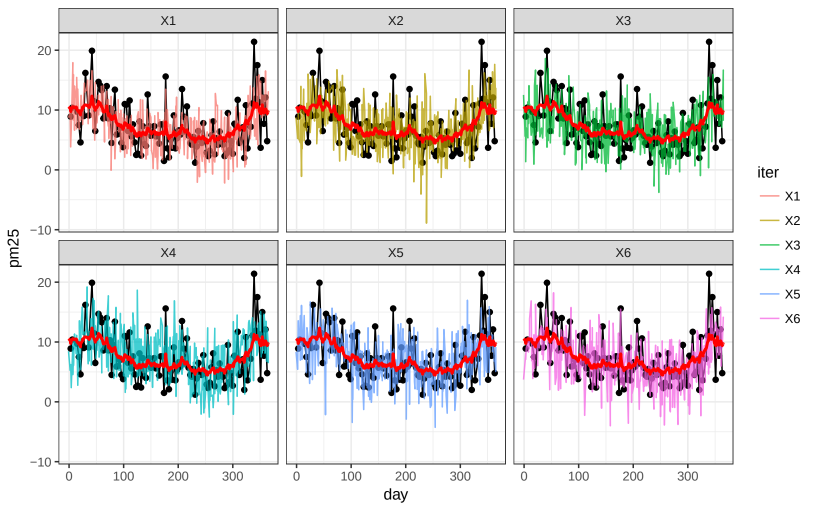
$$\mathbf{y}_{pred}^{(i)} \mid l^{(i)}, \sigma^{2}{}^{(i)}, \sigma_w^{2}{}^{(i)}, \beta_0{}^{(i)}, \beta_1{}^{(i)}, \beta_2{}^{(i)}, \mathbf{y}$$

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

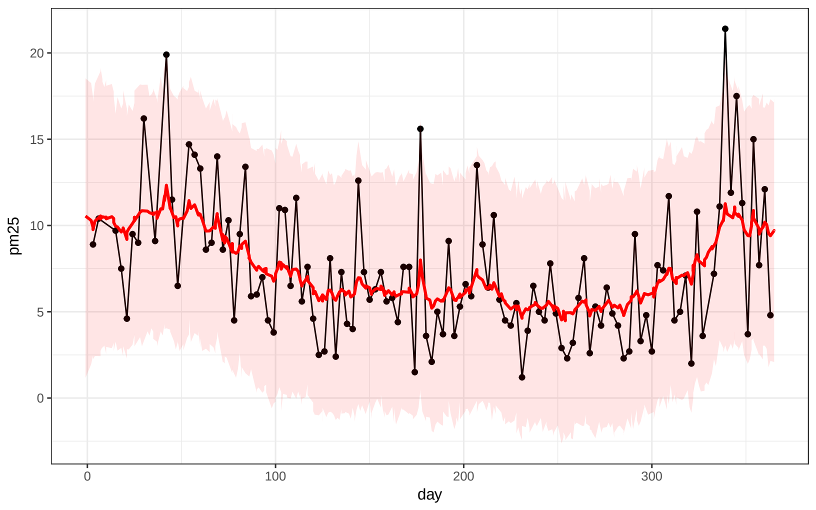
Full Posterior Predictive Distribution

```r
if (!file.exists("gp_pred.Rdata"))
{
  x = pm25$day; y = pm25$pm25

  n_post_samp = nrow(param)

  x_pred = 1:365 + rnorm(365, 0.01)
  y_pred = matrix(NA, nrow=n_post_samp, ncol=length(x_pred))
  colnames(y_pred) = paste0("Y_pred[", round(x_pred,0), "]")

  for(i in 1:n_post_samp)
  {
    l = param[i,'l']
    sigma2 = param[i,'sigma2']
    sigma2_w = param[i,'sigma2_w']
    beta0 = betas[i,"beta[1]"]
    beta1 = betas[i,"beta[2]"]
    beta2 = betas[i,"beta[3]"]

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

    dist_o = rdist(x)
    dist_p = rdist(x_pred)
    dist_op = 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 = sigma2_w)
    cov_p  = sq_exp_cov(dist_p,  sigma2 = sigma2, l = l) + nugget_cov(dist_p,  sigma2 = sigma2_w)
    cov_op = sq_exp_cov(dist_op, sigma2 = sigma2, l = l) + nugget_cov(dist_op, sigma2 = sigma2_w)
    cov_po = sq_exp_cov(dist_po, sigma2 = sigma2, l = l) + nugget_cov(dist_po, sigma2 = sigma2_w)

    cond_cov = cov_p - cov_po %*% solve(cov_o) %*% cov_op
    cond_mu  = mu_pred + cov_po %*% solve(cov_o) %*% (y - mu)

    y_pred[i,] = cond_mu + t(chol(cond_cov)) %*% matrix(rnorm(length(x_pred)), ncol=1)
  }
```

Full Posterior Predictive Distribution - Plots



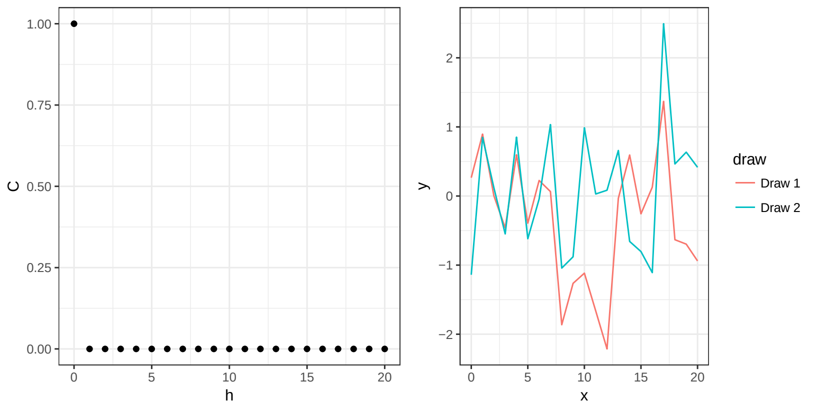
Full Posterior Predictive Distribution - Mean + CI



More on Covariance Functions

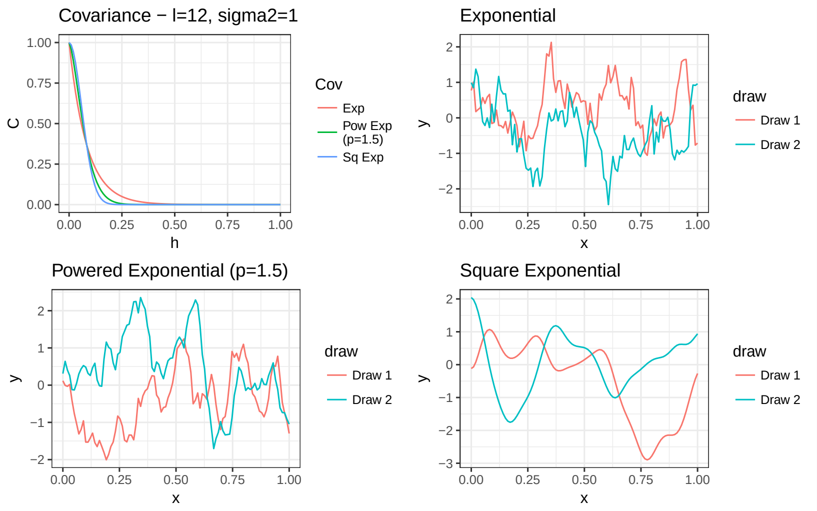
Nugget Covariance

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



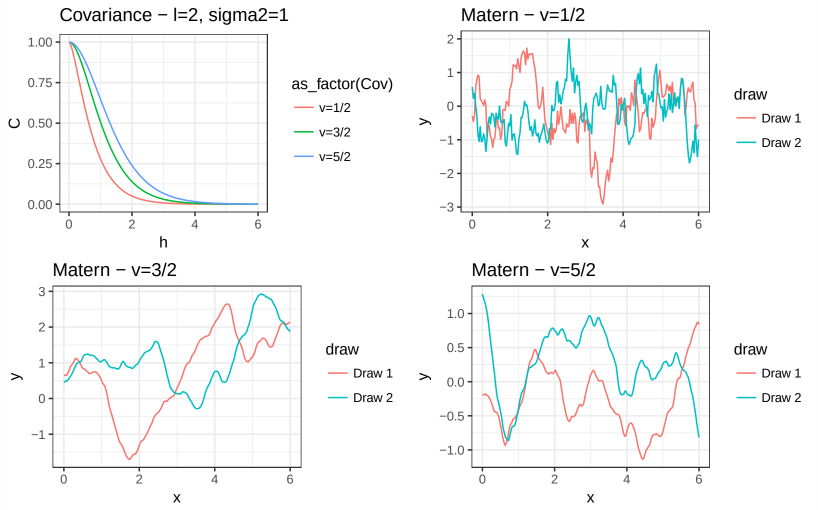
(- / Power / Square) Exponential Covariance

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



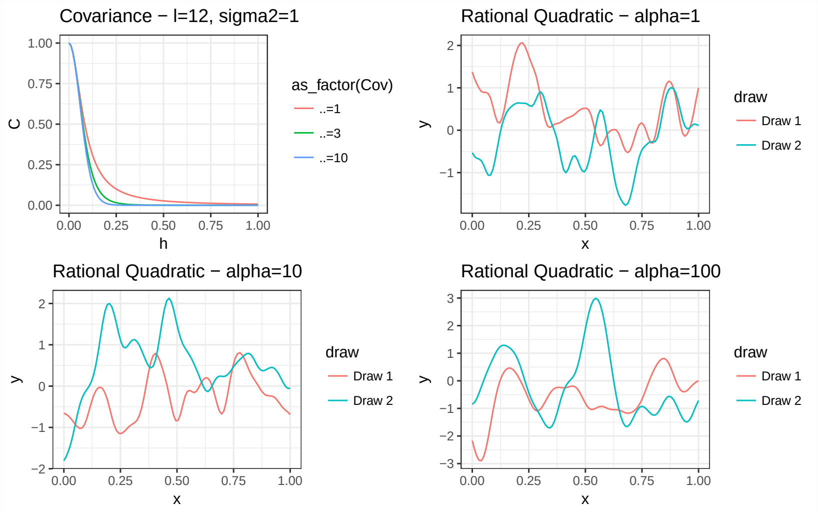
Matern Covariance

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



Rational Quadratic Covariance

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



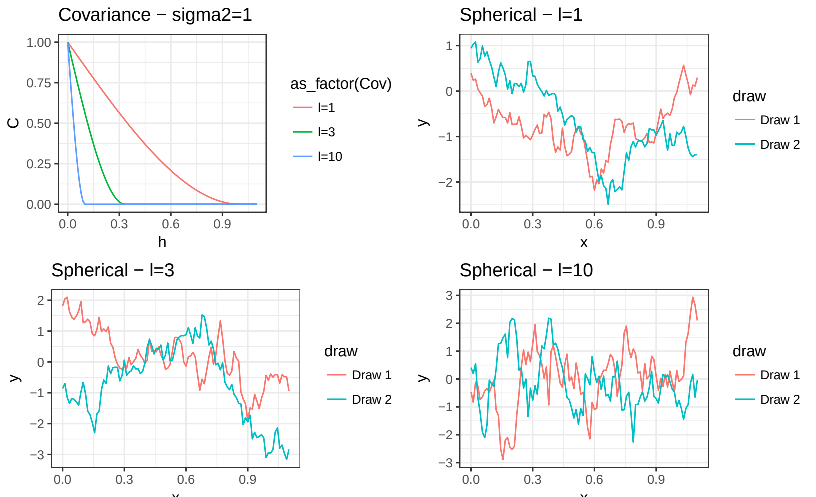
Some properties

- Matern Covariance
  - 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.

- Rational Quadratic Covariance
  - is a scale mixture (infinite sum) 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

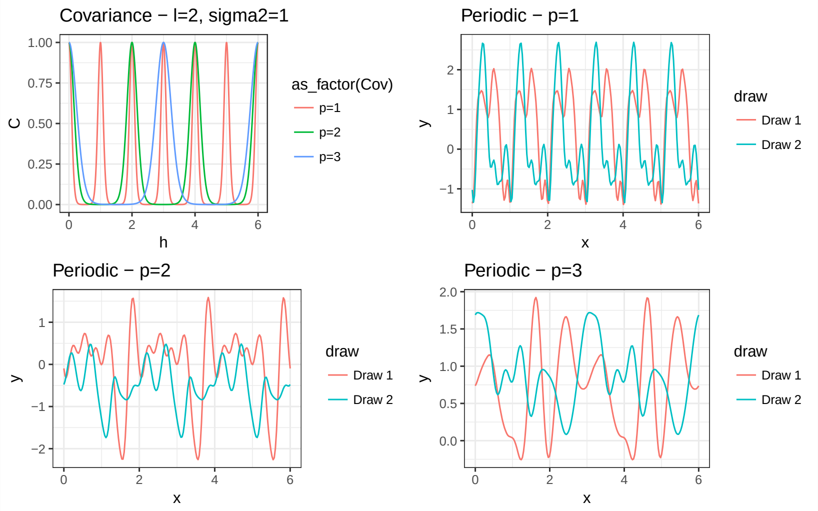
Spherical Covariance

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



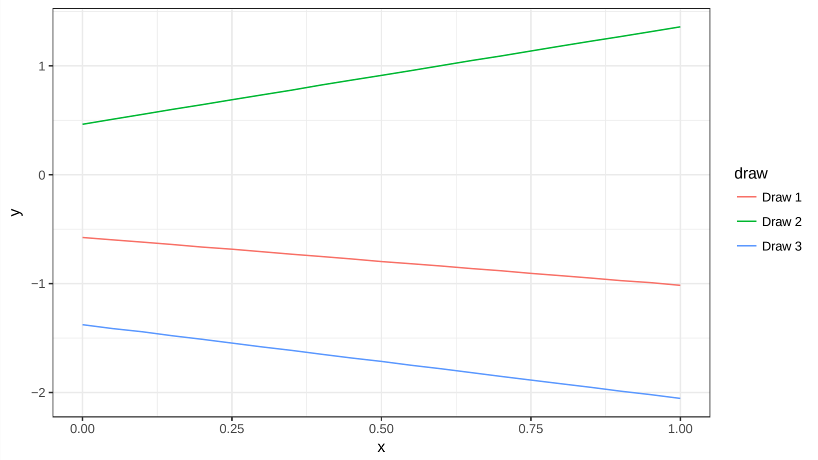
Periodic Covariance

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



Linear Covariance

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



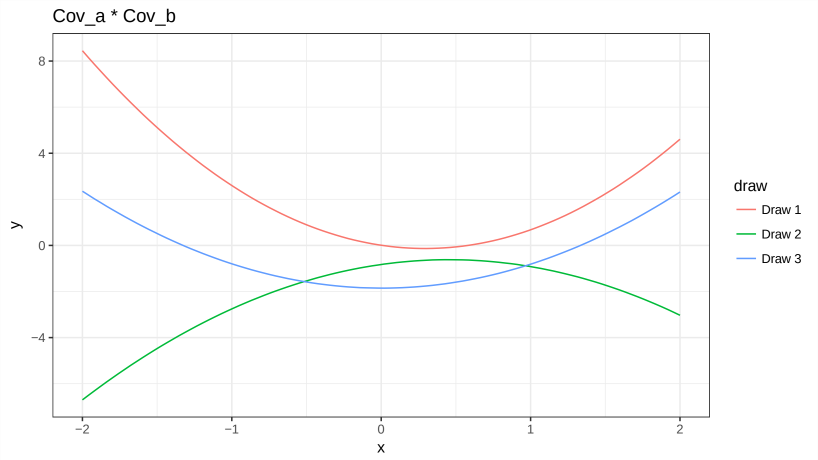
If we definite two valid covariance functions, $Cov_a(y_{t_i}, y_{t_j})$ and $Cov_b(y_{t_i}, y_{t_j})$ then the following are also valid covariance functions,

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

Linear × Linear → Quadratic

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

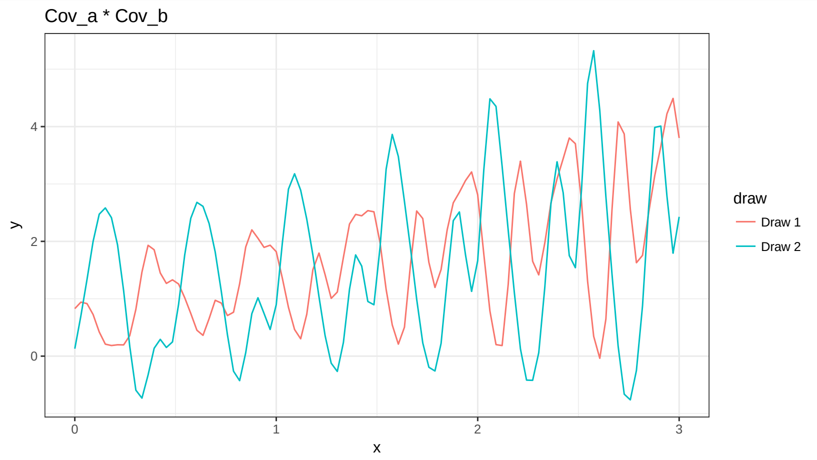
$$Cov_b(y_{t_i}, y_{t_j}) = 2 + 1\,(t_i \times t_j)$$



Linear × Periodic

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

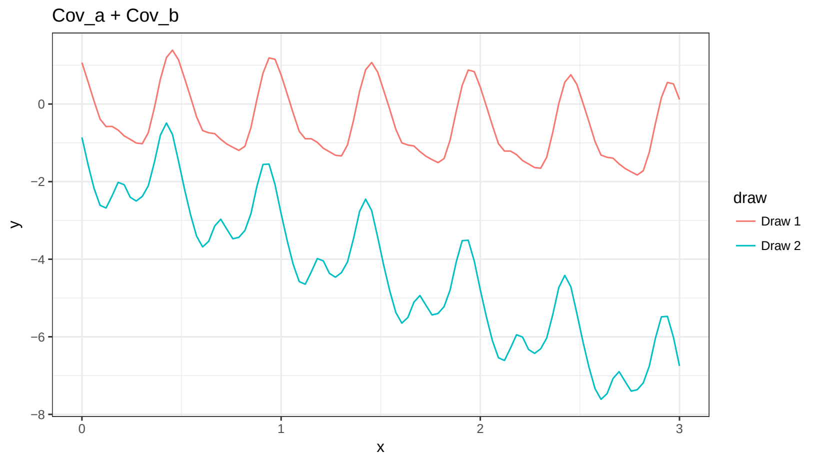
$$Cov_b(y_{t_i}, y_{t_j}) = Cov(h = |t_i - t_j|) = \exp\left(-2\,\sin^2\left(2\pi\,h\right)\right)$$



Linear + Periodic

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

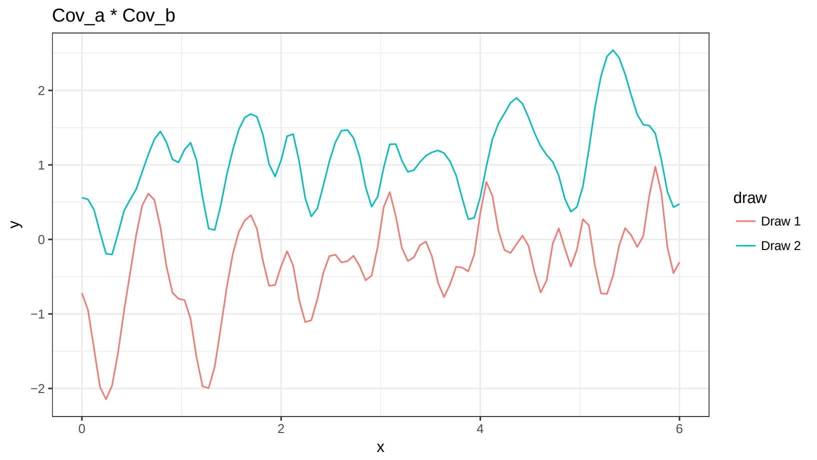
$$Cov_b(h = |t_i - t_j|) = \exp\left(-2\ \sin^2\left(2\pi\, h\right)\right)$$



Sq Exp × Periodic → Locally Periodic

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

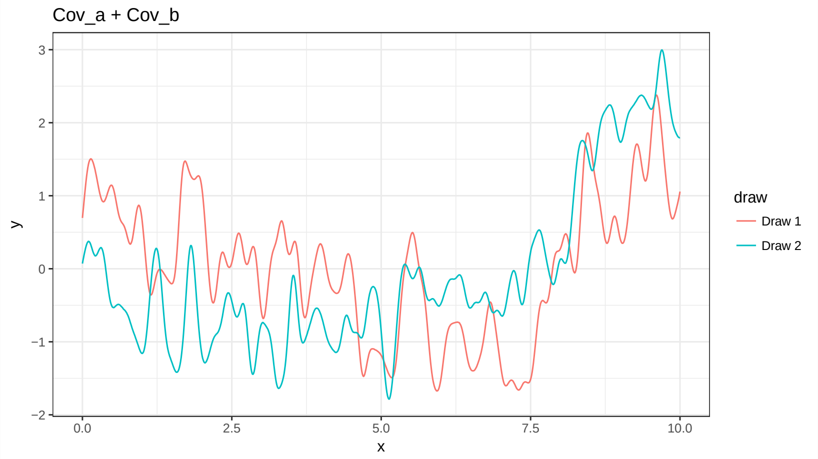
$$Cov_b(h = |t_i - t_j|) = \exp\left(-2\,\sin^2\left(\pi\, h\right)\right)$$



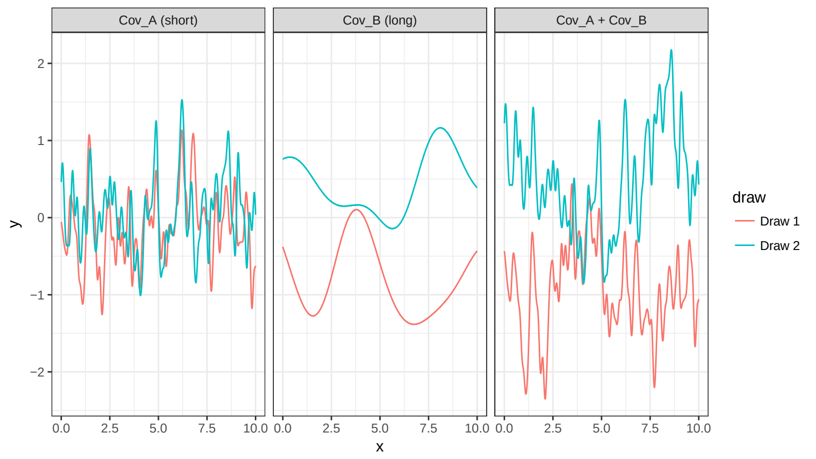
Sq Exp (short) + Sq Exp (long)

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

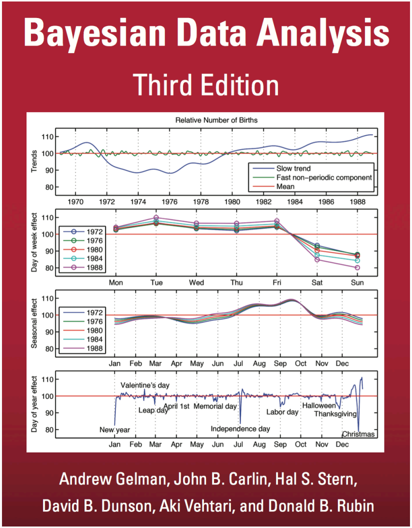
$$Cov_b(h = |t_i - t_j|) = \exp(-(\sqrt{3}/2)h^2)$$



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

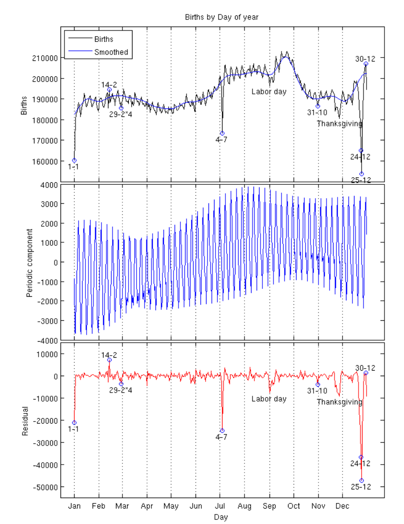


BDA3 example



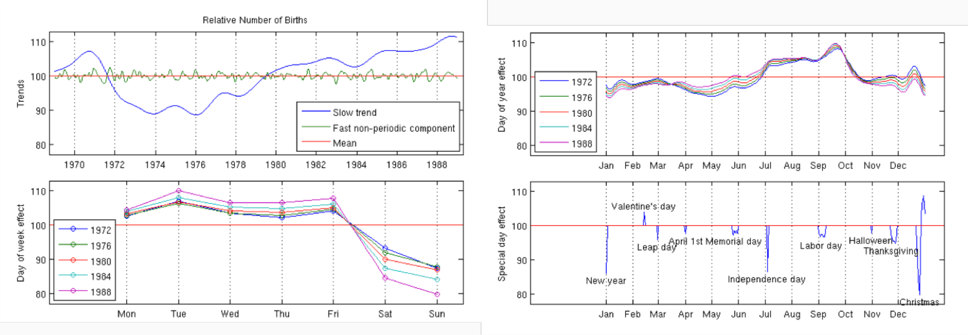
http://research.cs.aalto.fi/pml/software/gpstuff/demo_births.shtml

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
4. Student t observation model

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