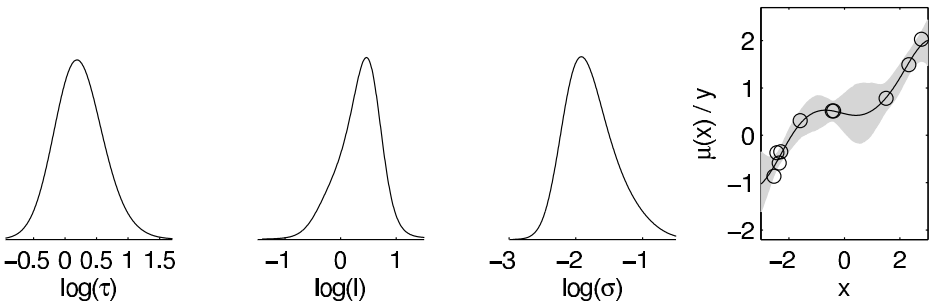


Figure 21.3 *Marginal posterior distributions for Gaussian process parameters $\tau$, $l$ and error scale $\sigma$, and posterior mean and pointwise 90% bands for $\mu(x)$, given the same ten data points from Figure 21.2.*

shows an example of estimated marginal posterior distributions for $\tau$, $l$, and $\sigma$, and the posterior mean and pointwise 90% bands for $\mu(x)$ using the same data as in Figure 21.2. We obtained the posterior simulations using slice sampling. The priors were $t_4^+(0,1)$ for $\tau$ and $l$, and log-uniform for $\sigma$.

## 21.2 Example: birthdays and birthdates

Gaussian processes can be directly fit to data, but more generally they can be used as components in a larger model. We illustrate with an analysis of patterns in birthday frequencies in a dataset containing records of all births in the United States on each day during the years 1969–1988. We originally read about these data being used to uncover a pattern of fewer births on Halloween and excess births on Valentine's Day (due, presumably, to choices involved in scheduled deliveries, along with decisions of whether to induce a birth for health reasons). We thought it would be instructive to fit a model to look not just at special days but also at day-of-week effects, patterns during the year, and longer-term trends.

*Decomposing the time series as a sum of Gaussian processes*

Based on the structural knowledge of the calendar and, we started with an additive model,

$$y_t(t) = f_1(t) + f_2(t) + f_3(t) + f_4(t) + f_5(t) + \epsilon_t,$$

where $t$ is time in days (starting with $t = 1$ on 1 January 1969), and the different terms represent variation with different scales and periodicity:

1. Long-term trends modeled by a Gaussian process with squared exponential covariance function:
$$f_1(t) \sim \mathrm{GP}(0, k_1), \quad k_1(t,t') = \sigma_1^2 \exp\left(-\frac{|t-t'|^2}{l_1^2}\right);$$
2. Shorter term variation using a GP with squared exponential covariance function with different amplitude and scale:
$$f_2(t) \sim \mathrm{GP}(0, k_2), \quad k_2(t,t') = \sigma_2^2 \exp\left(-\frac{|t-t'|^2}{l_2^2}\right);$$
3. Weekly quasi-periodic pattern (that is allowed to change over time) modeled as a product of periodic and squared exponential covariance function:
$$f_3(t) \sim \mathrm{GP}(0, k_3), \quad k_3(t,t') = \sigma_3^2 \exp\left(-\frac{2\sin^2(\pi(t-t')/7)}{l_{3,1}^2}\right) \exp\left(-\frac{|t-t'|^2}{l_{3,2}^2}\right);$$

4. Yearly smooth seasonal pattern using product of periodic and squared exponential covariance function (with period 365.25 to match the average length of the year):

$$f_4(t) \sim \mathrm{GP}(0, k_4), \quad k_4(s, s') = \sigma_4^2 \exp\left(-\frac{2\sin^2(\pi(s-s')/365.25)}{l_{4,1}^2}\right) \exp\left(-\frac{|s-s'|^2}{l_{4,2}^2}\right),$$

where $s = s(t) = t \bmod 365.25$, thus aligning itself with the calendar every four years .

5. Special days including an interaction term with weekend. Based on a combination of initial visual inspection and prior knowledge we chose the following special days: New Year's Day, Valentine's Day, Leap Day, April Fool's Day, Independence Day, Halloween, Christmas, and the days between Christmas and New Year's.

$$f_5(t) = I_{\text{special day}}(t)\beta_a + I_{\text{weekend}}(t) I_{\text{special day}}(t)\beta_b,$$

where $I_{\text{special day}}(t)$ is a row vector of 13 indicator variables corresponding to each of the special days (we can think of this vector of one row of an $n \times 13$ indicator matrix $I_{\text{special day}}$); $I_{\text{weekend}}(t)$ is an indicator variable that equals 1 if $t$ is a Saturday or Sunday, and 0 otherwise; and $\beta_a$ and $\beta_b$ are vectors, each of length 13, corresponding to the effects of special days when they fall on weekdays or weekends.

6. Finally, $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ represents the unstructured residuals.

We set weakly informative log-$t$ priors for the time-scale parameters $l$ (to improve identifiability of the model) and log-uniform priors for all the other hyperparameters. We normalized the number of daily births $y$ to have mean 0 and standard deviation 1.

The sum of Gaussian processes is also a Gaussian process, and the covariance function for the sum is

$$k(t, t') = k_1(t, t') + k_2(t, t') + k_3(t, t') + k_4(t, t') + k_5(t, t').$$

The inference for the model is then straightforward with basic Gaussian process equations.

We analytically determined the marginal likelihood and its gradients for hyperparameters as in (21.1), and we used the marginal posterior mode for the hyperparameters. As $n$ was relatively high (corresponding to all the days during a twenty-year period, that is $n \approx 20 \cdot 365.25$), this posterior mode was fine in practice. Central composite design (CCD) integration gave visually indistinguishable plots, and MCMC would have been too slow. The Gaussian process formulation with $O(n^3)$ computation time is not optimal for this kind of one-dimensional data, but computation time was still reasonable.

Figure 21.4 shows the slow trend, faster non-periodic correlated variation, weekly trend and its change through years, seasonal effect and its change through years, and day of year effects. All plots are on the same scale showing differences relative to a baseline of 100. Predictions for different additive components can be computed with the usual posterior equation (21.1) but using only one of the covariance functions to compute the covariance between training and the test data. For example, the mean of the slow trend is computed as

$$\mathrm{E}(\tilde{f}_1) = K_1(\tilde{x}, x)(K(x, x) + \sigma^2 I)^{-1} y. \tag{21.2}$$

The smooth seasonal effect has an inverse relation to the amount of daylight or the average temperature nine months before. The smaller number of births in weekends and smaller or larger number of births on special days can be explained by selective c-sections and induced births. The day-of-week patterns become more pronounced over time, which makes sense given the general increasing rate of these sorts of births.

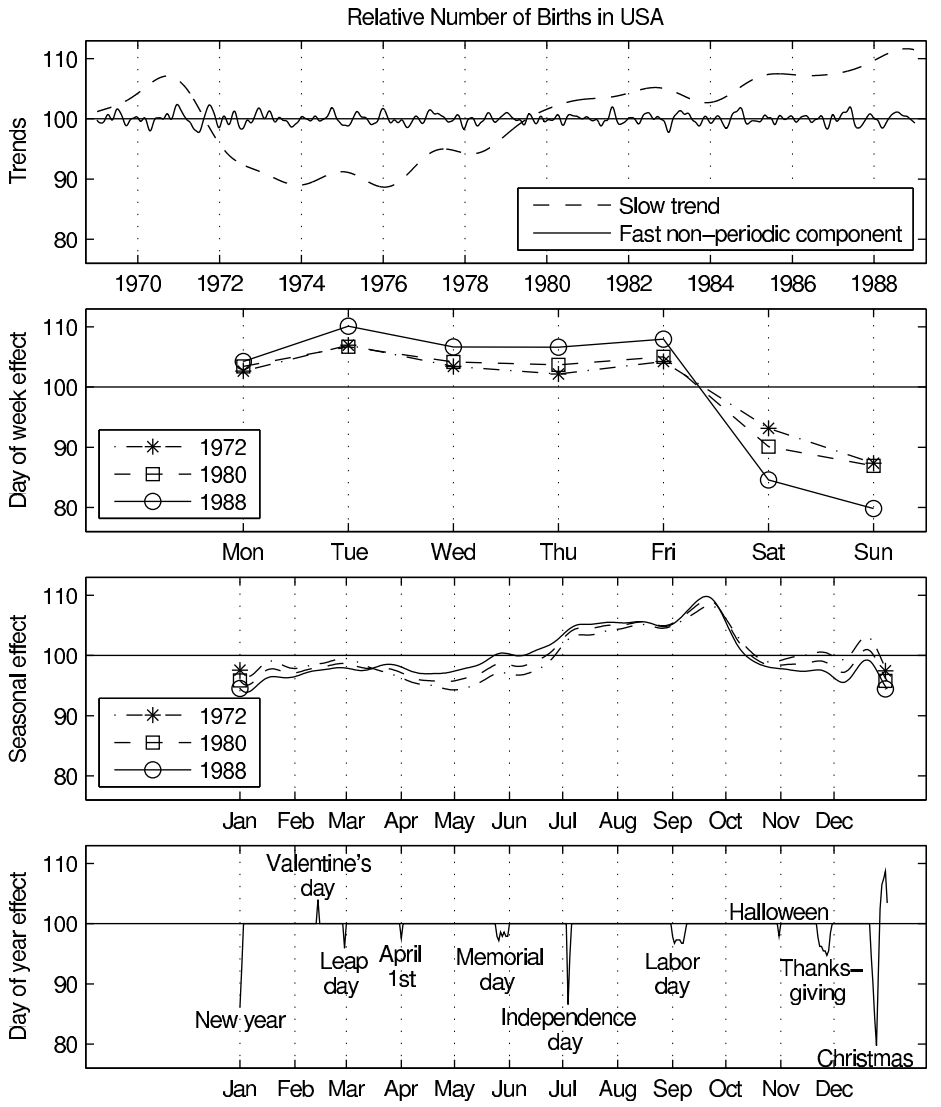


Figure 21.4 *Relative number of births in the United States based on exact data from each day from 1969 through 1988, divided into different components, each with an additive Gaussian process model. The estimates from an improved model are shown in Figure 21.5.*

*An improved model*

Statistical models are not built at once. Rather, we fit a model, notice problems, and improve it. In this case, selecting just some special days makes it impossible to discover other days having a considerable effect. Also we might expect to see a 'ringing' pattern with a distortion of births just before and after the special days (as the babies have to be born sometime).

To allow for these sorts of structures, we constructed a new model that allowed special effects for each day of the year. While analyzing the first model we also noticed that the residuals were slightly autocorrelated, so we added a very short time-scale non-periodic component to explain that. To improve yearly periodic components we also refined the

handling of the leap day. Our improved model has the form

$$y_t(t) = f_1(t) + f_2(t) + f_3(t) + f_4(t) + f_5(t) + f_6(t) + f_7(t) + f_8(t) + \epsilon_t :$$

1. Long-term trends modeled by a Gaussian process with squared exponential covariance function:
$$f_1(t) \sim \mathrm{GP}(0, k_1), \quad k_1(t, t') = \sigma_1^2 \exp\left(-\frac{|t - t'|^2}{l_1^2}\right);$$
2. Shorter term variation using a GP with squared exponential covariance function with different amplitude and scale:
$$f_2(t) \sim \mathrm{GP}(0, k_2), \quad k_2(t, t') = \sigma_2^2 \exp\left(-\frac{|t - t'|^2}{l_2^2}\right);$$
3. Weekly quasi-periodic pattern (that is allowed to change over time) modeled as a product of periodic and squared exponential covariance function:
$$f_3(t) \sim \mathrm{GP}(0, k_3), \quad k_3(t, t') = \sigma_3^2 \exp\left(-\frac{2\sin^2(\pi(t - t')/7)}{l_{3,1}^2}\right) \exp\left(-\frac{|t - t'|^2}{l_{3,2}^2}\right);$$
4. Yearly smooth seasonal pattern using product of periodic and squared exponential covariance function (with period 365.25 to match the average length of the year):
$$f_4(t) \sim \mathrm{GP}(0, k_4), \quad k_4(s, s') = \sigma_4^2 \exp\left(-\frac{2\sin^2(\pi(s - s')/365)}{l_{4,1}^2}\right) \exp\left(-\frac{|s - s'|^2}{l_{4,2}^2}\right),$$
$s = s(t)$ is now a modified time with time before and after leap day incremented by 0.5 day so that in $s$ the length of year is 365 also for leap years (making easier implementation of yearly periodicity).
5. Yearly fast changing pattern for weekdays (day-of-year effect) using a periodic covariance function:
$$f_5(t) \sim \mathrm{GP}(0, k_5), \quad k_5(s, s') = I_{\text{weekday}}(t)\sigma_5^2 \exp\left(-\frac{2\sin^2(\pi(s - s')/365)}{l_5^2}\right);$$
6. A similar pattern for weekends:
$$f_6(t) \sim \mathrm{GP}(0, k_6), \quad k_5(s, s') = I_{\text{weekend}}(t)\sigma_6^2 \exp\left(-\frac{2\sin^2(\pi(s - s')/365)}{l_6^2}\right);$$
7. Effects of special days whose dates are not constant from year to year (Leap Day, Memorial Day, Labor Day, Thanksgiving):
$$f_7(t) = I_{\text{special day}}(t)\beta,$$
where $I_{\text{special day}}(t)$ is now a row vector of 4 indicator variables corresponding to these floating holidays.
8. Short-term variation using a Gaussian process with squared exponential covariance function:
$$f_8(t) \sim \mathrm{GP}(0, k_8), \quad k_8(t, t') = \sigma_2^2 \exp\left(-\frac{|t - t'|^2}{l_8^2}\right);$$
9. Finally, $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ models the unstructured residual.

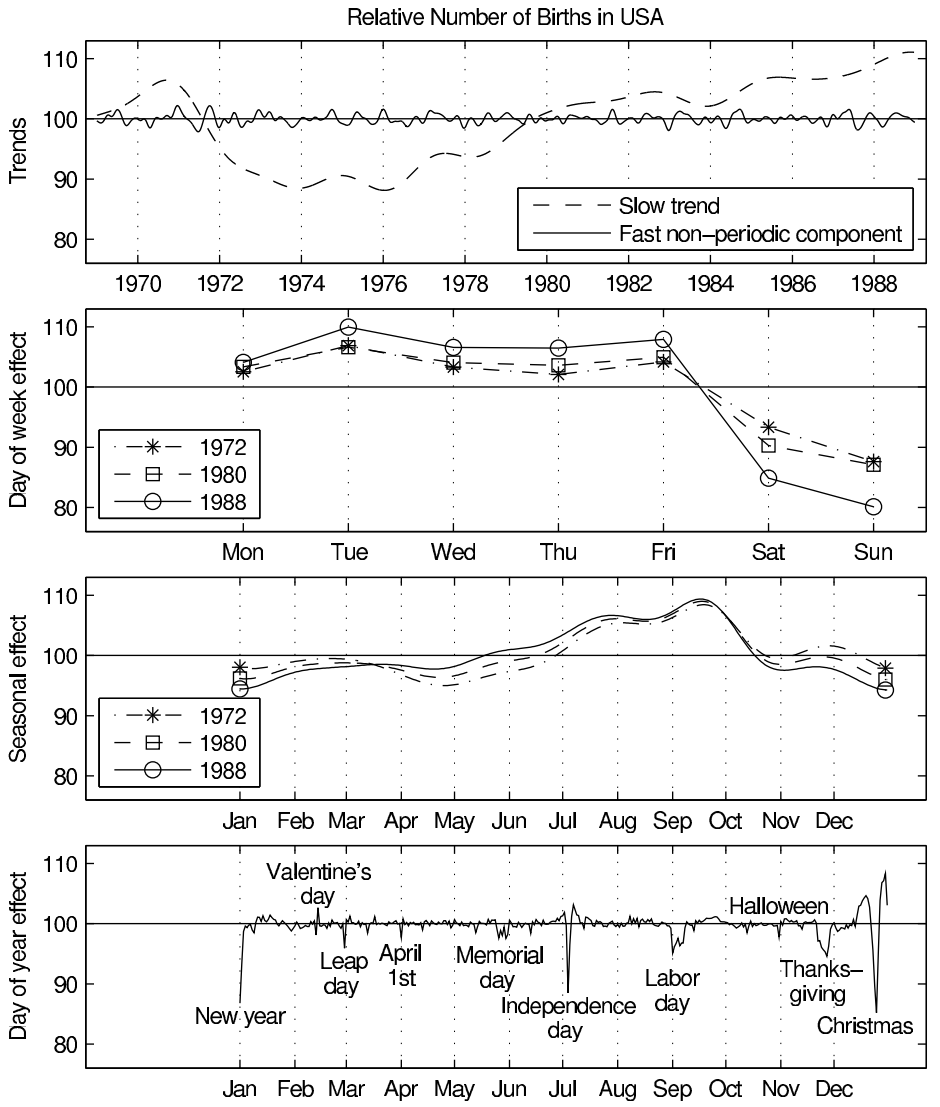


Figure 21.5 *Relative number of births in the United States based on exact data from each day from 1969 through 1988, divided into different components, each with an additive Gaussian process model. Compared to Figure 21.4, this improved model allows individual effects for every day of the year, not merely for a few selected dates.*

We set weakly informative log-$t$ priors for time-scale parameters $l$ (to improve identifiability of the model) and log-uniform prior for all the other hyperparameters. The number of births $y$ was normalized to have mean 0 and standard deviation 1.

Exploiting properties of multivariate Gaussian, leave-one-out cross-validation pointwise predictions can be computed in similar time as posterior predictions. The cross-validated pointwise predictive accuracy is $\text{lppd}_{\text{loo-cv}} = 2074$ for the first model and 2477 for the improved model, showing clear improvement.

Figure 21.5 shows the results for the second model. The trends and day of week effect are indistinguishable from the first model, but the seasonal component is smoother as it does not need to model the increased number of births before or after special days and

before the end of the year, which are now modeled in the day of year component. This new model is not perfect either (for one thing, it would make sense to constrain local positive and negative effects to average approximately to zero so that extra babies are explicitly 'borrowed' from neighboring days), but we believe that the decomposition shown in Figure 21.5 does a good job of identifying the major patterns at different time scales. The trick was to use Gaussian processes to allow different scales of variation for different components of the model. This example also illustrates how we are able to keep adding terms to the additive model without losing control of the estimation.

### 21.3 Latent Gaussian process models

In case of non-Gaussian likelihoods, the Gaussian process prior is set to a latent function $f$ which through a link function determines the likelihood $p(y|f,\phi)$ as in generalized linear models (see Chapter 16). Typically the shape parameter $\phi$ is assumed to be a scalar, but it is also possible to use separate latent Gaussian processes to model location $f$ and shape parameter $\phi$ of the likelihood to allow, for example, the scale to depend on the predictors.

The conditional posterior density of the latent $f$ is $p(f|x,y,\theta,\phi) \propto p(y|f,\phi)p(f|x,\theta)$. For efficient MCMC inference, GP-specific samplers can be used. These samplers exploit the multivariate Gaussian form of the prior for the latent values in the proposal distribution or in the scaling of the latent variables. The most commonly used samplers are the elliptic slice sampler, scaled Metropolis-Hastings, and scaled HMC/NUTS (these are Gaussian-process-specific variations of the samplers discussed in Chapters 11 and 12). Typically the sampling is done alternating the sampling of latent values $f$ and covariance and likelihood parameters $\theta$ and $\phi$. Due to dependency between the latent values and the (hyper)parameters, mixing of the MCMC can be slow, creating difficulties when fitting to larger datasets.

As the prior distribution for latent values is multivariate Gaussian, the posterior distribution of the latent values is also often close to Gaussian; this motivates Gaussian posterior approximations. The simplest approach is to use the normal approximation (Chapter 4)

$$p(f|x,y,\theta,\phi) \approx \mathrm{N}(f|\hat{f},\Sigma),$$

where $\hat{f}$ is the posterior mode and

$$\Sigma^{-1} = K(x,x) + W,$$

where $K(x,x)$ is the prior covariance matrix and $W$ is a diagonal matrix with $W_{ii} = \frac{d^2}{df^2}\log p(y|f_i,\phi)|_{f_i=\hat{f}_i}$. The approximate predictive density

$$p(\tilde{y}_i|\tilde{x}_i,x,y,\theta,\phi) \approx \int p(\tilde{y}_i|\tilde{f}_i,\phi)\mathrm{N}(\tilde{f}_i|\tilde{x}_i,x,y,\theta,\phi)d\tilde{f}_i$$

can be evaluated, for example, with quadrature integration. Log marginal likelihood can be approximated by integrating over f using Laplace's method (Section 13.3)

$$\log p(y|x,\theta,\phi) \approx \log g(y|x,\theta,\phi) \propto \log p(y|\hat{f},\phi) - \frac{1}{2}\hat{f}^{\mathrm{T}}K(x,x)^{-1}\hat{f} - \frac{1}{2}\log|B|, \qquad (21.3)$$

where $|B| = |I + W^{1/2}K(x,x)W^{1/2}|$.

If the likelihood contribution is heavily skewed, as can be the case with the logistic model, expectation propagation (Section 13.8) can be used instead. Variational approximation (Section 13.7) has also been used, but in many cases it is slower than the normal approximation and not as accurate as EP. Using one of these analytic approximations for the latent posterior, the approximate (unnormalized) marginal posterior of the hyperparameters $q(\theta,\phi|x,y)$ and its gradients can be computed analytically, allowing one to efficiently