Aalto University
School of Science

Explicit Link Between Periodic Covariance Functions and State Space Models

Arno Solin and Simo Särkkä

Department of Biomedical Engineering and Computational Science
Aalto University, Finland
arno.solin@aalto.fi

April 22, 2014

- Gaussian processes and state space models
- Periodic covariance functions
- Quasi-periodic covariance functions
- Example studies

Gaussian Process Regression in 1D

- **The kernel approach:**

  $$f(t) \sim \mathcal{GP}(0, k(t,t'))$$

  $$y_k = f(t_k) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_\mathrm{n}^2),$$

  where the observed data is $\{(t_k, y_k), k = 1, 2, \ldots, n\}$.
- Prior assumptions of the process encoded into the covariance function $k(t,t')$.
- Can be solved in closed-form, but the naive solution scales as $\mathcal{O}(n^3)$.

Gaussian Process Regression in 1D

- **The kernel approach:**

$$f(t) \sim \mathcal{GP}(0, k(t, t'))$$
$$y_k = f(t_k) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_\mathrm{n}^2),$$

  where the observed data is $\{(t_k, y_k), k = 1, 2, \ldots, n\}$.

- Prior assumptions of the process encoded into the covariance function $k(t, t')$.
- Can be solved in closed-form, but the naive solution scales as $\mathcal{O}(n^3)$.

- **State space approach:**

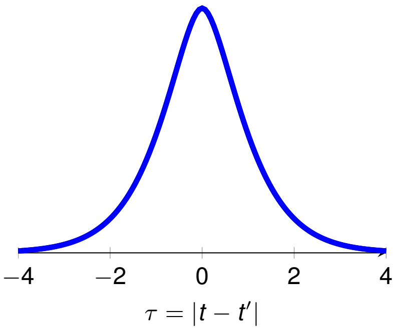
$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t)$$
$$y_k = \mathbf{H}\mathbf{f}(t_k) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_\mathrm{n}^2),$$

  where $\mathbf{w}(t)$ is a white noise process with spectral density $\mathbf{Q}_\mathrm{c}$.

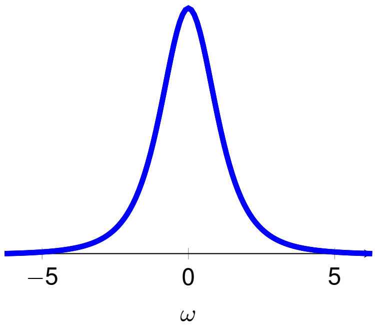
- Model defined by $\mathbf{F}$, $\mathbf{L}$, $\mathbf{Q}_\mathrm{c}$, the stationary covariance $\mathbf{P}_\infty$, and the observation model $\mathbf{H}$.
- Solved using Kalman filtering and Rauch–Tung–Striebel smoothing in $\mathcal{O}(n)$ time complexity.

Covariance and Spectral Density

$$k(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \exp(-\mathrm{i}\,\omega\tau)\, \mathrm{d}\omega.$$

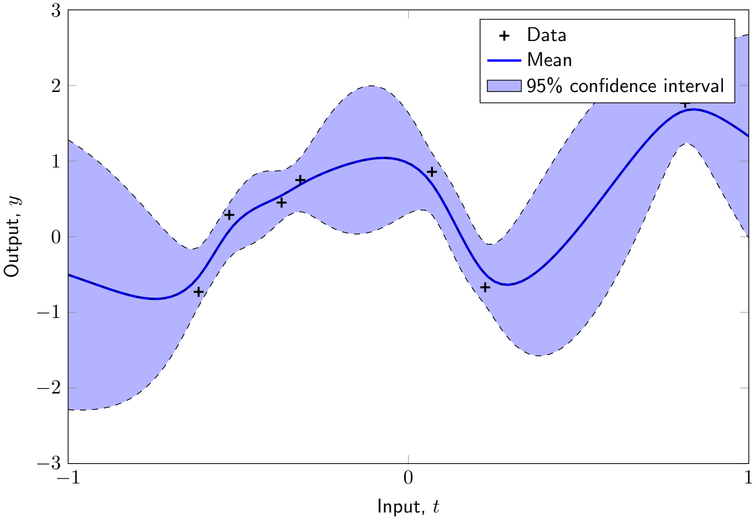


Covariance function $k(\tau)$

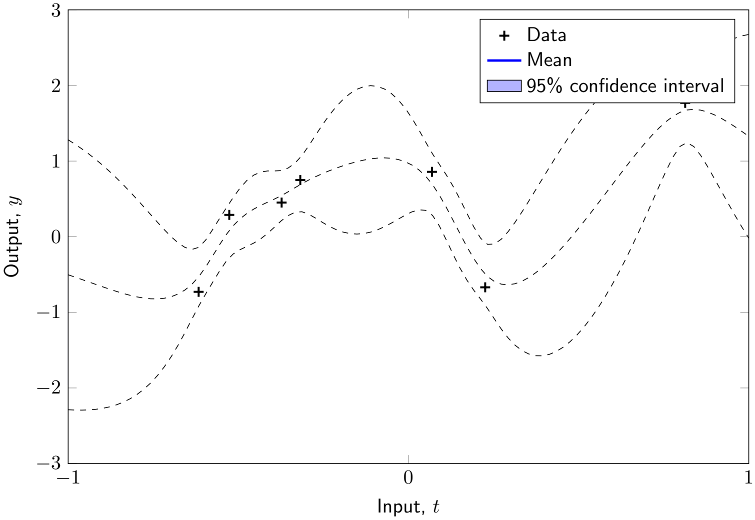


Spectral density $S(\omega)$

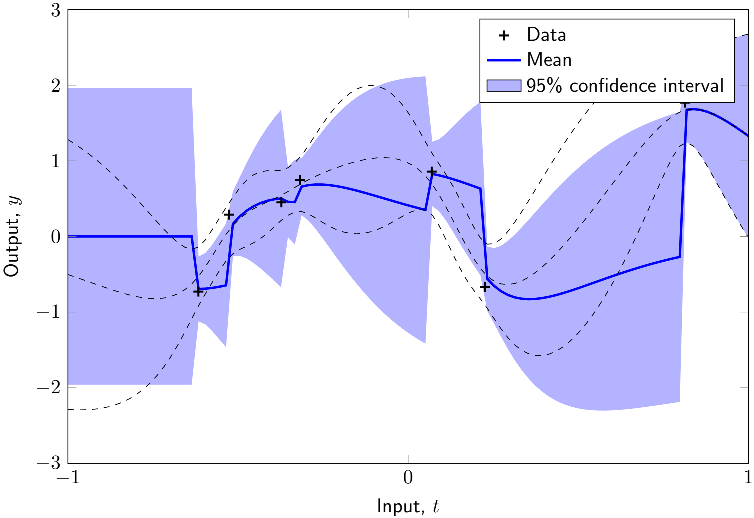
GP Regression (naive solution)



GP Regression (filtering)



GP Regression (smoothing)



Periodic Covariance Functions

Canonical Periodic Covariance Function

Start off with the squared exponential:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right)$$

Canonical Periodic Covariance Function

Start off with the squared exponential:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right)$$

Polar coordinates:

$$\mathbf{x}(t) = \begin{pmatrix} \cos(\omega_0\, t) \\ \sin(\omega_0\, t) \end{pmatrix}$$

Canonical Periodic Covariance Function

Start off with the squared exponential:

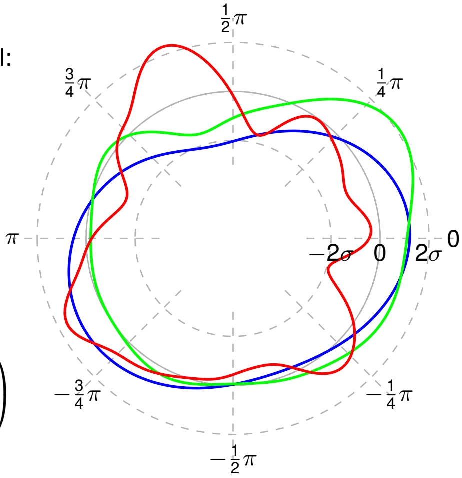
$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right)$$

Polar coordinates:

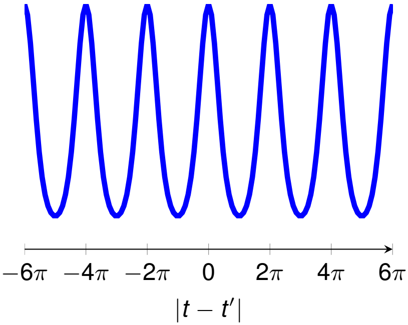
$$\mathbf{x}(t) = \begin{pmatrix} \cos(\omega_0\, t) \\ \sin(\omega_0\, t) \end{pmatrix}$$

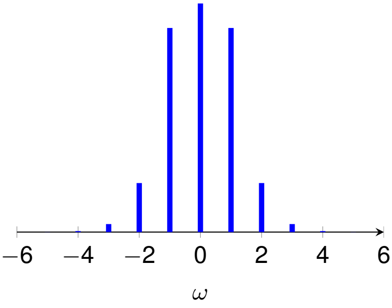
The *canonical* periodic covariance:

$$k_{\mathrm{p}}(t, t') = \sigma^2 \exp\left(-\frac{2\sin^2\left(\omega_0 \frac{t-t'}{2}\right)}{\ell^2}\right)$$



Canonical Periodic Covariance Function





Covariance function | **Spectral density**

State Space Formulation (1/2)

- Fourier series representation ($\tau = |t - t'|$)

$$k_{\mathrm{p}}(\tau) = \sum_{j=0}^{\infty} q_j^2 \cos(j\, \omega_0\, \tau)$$

State Space Formulation (1/2)

- Fourier series representation ($\tau = |t - t'|$)

$$k_{\mathrm{p}}(\tau) = \sum_{j=0}^{\infty} q_j^2 \cos(j\,\omega_0\,\tau)$$

- The model can be constructed as solutions to second-order ODEs:

$$\begin{pmatrix} \dot{x}_j(t) \\ \dot{y}_j(t) \end{pmatrix} = \begin{pmatrix} 0 & -\omega_0\,j \\ \omega_0\,j & 0 \end{pmatrix} \begin{pmatrix} x_j(t) \\ y_j(t) \end{pmatrix}$$

State Space Formulation (2/2)

- The state space model can be given as a superposition the following kind of models:

$$\mathbf{F}_j^{\mathrm{p}} = \begin{pmatrix} 0 & -\omega_0\, j \\ \omega_0\, j & 0 \end{pmatrix}, \quad \mathbf{H}_j^{\mathrm{p}} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{P}_{\infty,j}^{\mathrm{p}} = q_j^2 \mathbf{I}_2$$

The diffusion part is zero (*i.e.* the model is deterministic).

State Space Formulation (2/2)

- The state space model can be given as a superposition the following kind of models:

$$\mathbf{F}_j^{\mathrm{p}} = \begin{pmatrix} 0 & -\omega_0\, j \\ \omega_0\, j & 0 \end{pmatrix}, \quad \mathbf{H}_j^{\mathrm{p}} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{P}_{\infty,j}^{\mathrm{p}} = q_j^2 \mathbf{I}_2$$

  The diffusion part is zero (*i.e.* the model is deterministic).

- The spectral (variance) coefficients are

$$q_j^2 = \frac{2\,\mathrm{I}_j(\ell^{-2})}{\exp(\ell^{-2})}, \text{ for } j = 1, 2, \ldots,$$

  and $q_0^2 = \mathrm{I}_0(\ell^{-2}) / \exp(\ell^{-2})$, where $\mathrm{I}_\alpha(z)$ is the modified Bessel function.

State Space Formulation (2/2)

- The state space model can be given as a superposition the following kind of models:

$$\mathbf{F}_j^{\mathrm{p}} = \begin{pmatrix} 0 & -\omega_0\, j \\ \omega_0\, j & 0 \end{pmatrix}, \quad \mathbf{H}_j^{\mathrm{p}} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{P}_{\infty,j}^{\mathrm{p}} = q_j^2 \mathbf{I}_2$$

  The diffusion part is zero (*i.e.* the model is deterministic).

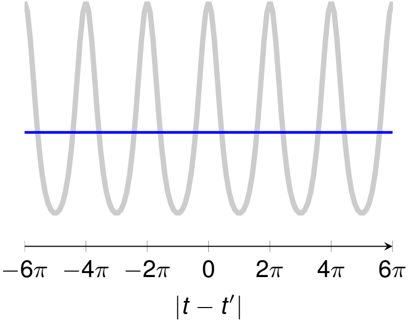
- The spectral (variance) coefficients are

$$q_j^2 = \frac{2\,\mathrm{I}_j(\ell^{-2})}{\exp(\ell^{-2})}, \text{ for } j = 1, 2, \ldots,$$

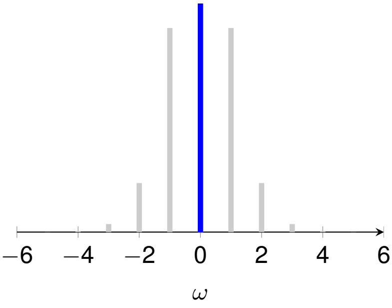
  and $q_0^2 = \mathrm{I}_0(\ell^{-2}) / \exp(\ell^{-2})$, where $\mathrm{I}_\alpha(z)$ is the modified Bessel function.

- Taking the J first terms in the series gives an approximation, and this approximation converges uniformly to the actual covariance as $J \to \infty$.

Approximative Covariance Function ($J = 0$)

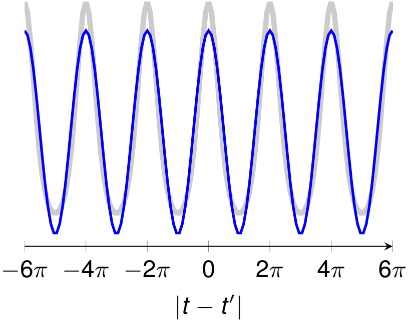


Covariance function

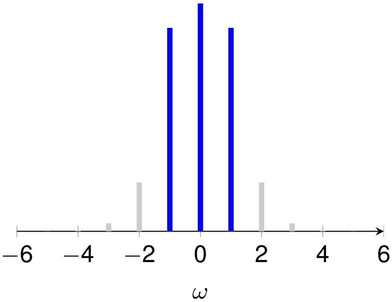


Spectral density

Approximative Covariance Function ($J = 1$)

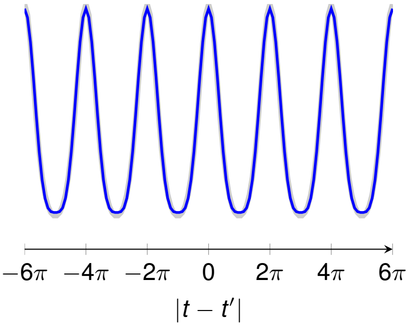


Covariance function

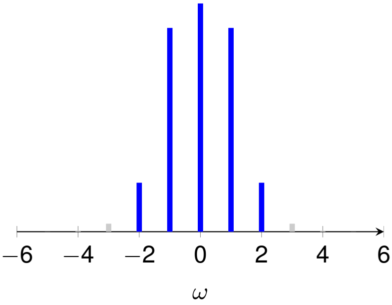


Spectral density

Approximative Covariance Function ($J = 2$)

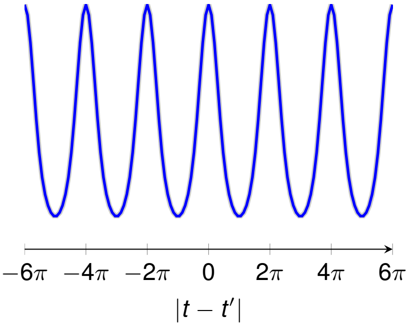


Covariance function

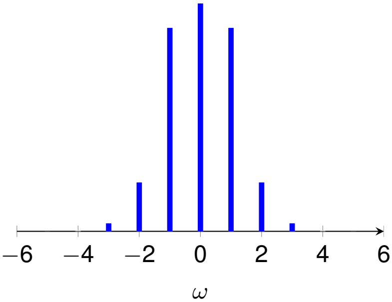


Spectral density

Approximative Covariance Function ($J = 3$)



Covariance function



Spectral density

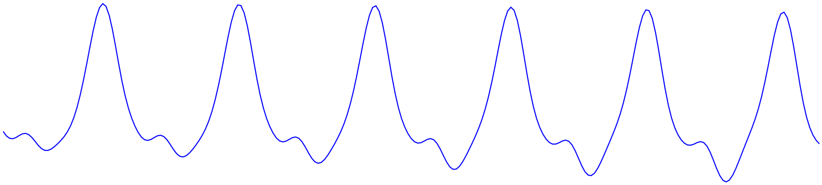
Quasi-Periodic Covariance Functions

Quasi-Periodic Covariance

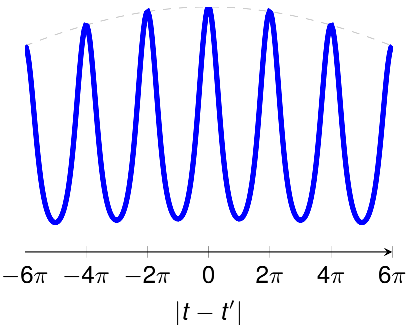
- Allow the shape of the periodic effect to change over time.
- Take the product of a periodic covariance function $k_\mathrm{p}(t, t')$ with a covariance function $k_\mathrm{q}(t, t')$ with rather long characteristic length-scale,

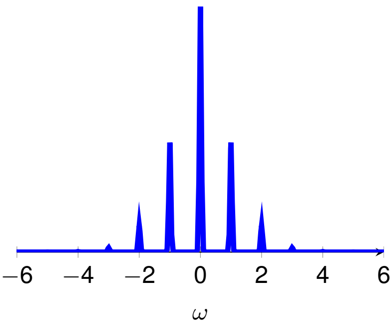
$$k(t, t') = k_\mathrm{p}(t, t')\, k_\mathrm{q}(t, t'),$$

allowing the covariance to decay away from exact periodicity.



Quasi-Periodic Covariance Function





Covariance function

Spectral density

Quasi-Periodic Covariances in State Space Form

- State space representation of both the ‘quasi’ and ‘periodic’ part.
- The quasi-periodic state space model needs to be set up so that the feedback matrices commute ($\mathbf{F}_\mathrm{p}\mathbf{F}_\mathrm{q} = \mathbf{F}_\mathrm{q}\mathbf{F}_\mathrm{p}$.).
- This can be accomplished by using the special features of the Kronecker product.

Quasi-Periodic Covariances in State Space Form

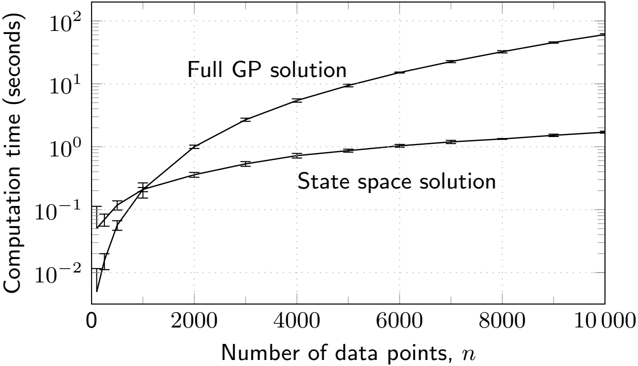
- State space representation of both the 'quasi' and 'periodic' part.
- The quasi-periodic state space model needs to be set up so that the feedback matrices commute ($\mathbf{F}_\mathrm{p}\mathbf{F}_\mathrm{q} = \mathbf{F}_\mathrm{q}\mathbf{F}_\mathrm{p}$.).
- This can be accomplished by using the special features of the Kronecker product.
- The joint model corresponding to the quasi-periodic product of the two covariance functions can then be given in a block-form:

$$
\begin{aligned}
\mathbf{F}_j &= \mathbf{F}^\mathrm{q} \otimes \mathbf{I}_2 + \mathbf{I}_q \otimes \mathbf{F}_j^\mathrm{p}, \\
\mathbf{L}_j &= \mathbf{L}^\mathrm{q} \otimes \mathbf{L}_j^\mathrm{p}, \\
\mathbf{Q}_{\mathrm{c},j} &= \mathbf{Q}_\mathrm{c}^\mathrm{q} \otimes q_j^2\, \mathbf{I}_2, \\
\mathbf{P}_{\infty,j} &= \mathbf{P}_\infty^\mathrm{q} \otimes \mathbf{P}_{\infty,j}^\mathrm{p}, \\
\mathbf{H}_j &= \mathbf{H}^\mathrm{q} \otimes \mathbf{H}_j^\mathrm{p},
\end{aligned}
$$

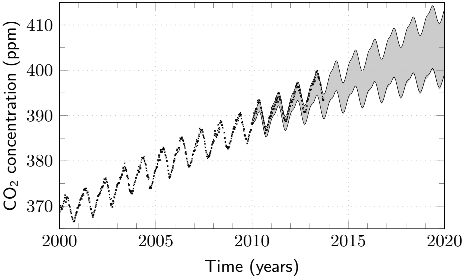
where '$\otimes$' denotes the Kronecker product of two matrices.

Example Studies

Example: Computational Complexity



Example: Mauna Loa CO_2 Concentration

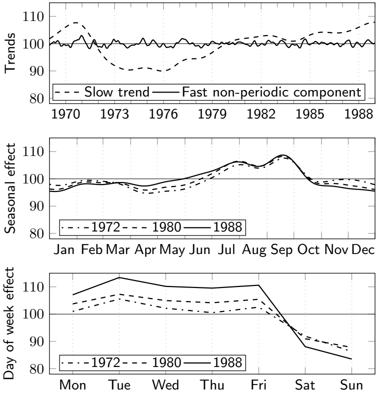


- CO_2 concentration observations ($n = 2227$, years 1958–2000 not shown in figure).
- The GP covariance function is as follows:

$$\begin{aligned} k(t,t') &= k_{\mathrm{SE}}(t,t') \\ &\quad + k_{\mathrm{p}}(t,t')\, k_{\nu=3/2}(t,t') \\ &\quad + k_{\nu=3/2}(t,t') \end{aligned}$$

- Converted to state space and hyperparameters optimized with respect to marginal likelihood.

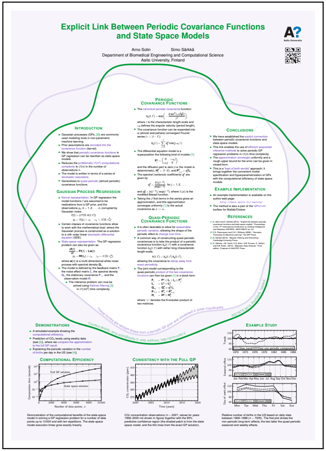
Example: Daily Births in 1969–1988



- Relative number of births in the US based on daily data between 1969–1988 ($n = 7305$).
- The model is as follows:
  - Matern ($\nu = 7/2$) for the slow trend
  - Matern ($\nu = 3/2$) for faster variation
  - Quasi-periodic (yearly) with Matern ($\nu = 3/2$) damping
  - Quasi-periodic (weekly) with Matern ($\nu = 3/2$) damping
- Converted to state space and hyperparameters optimized with respect to marginal likelihood

Conclusion

- We have established the explicit connection between periodic covariance functions and state space models.
- This link enables the use of efficient sequential inference methods to solve periodic GP regression problems in $\mathcal{O}(n)$ time complexity.
- The approximation converges uniformly and a rough upper bound for the error can be given in closed-form.
- This is a 'best of both worlds' approach; it brings together the convenient model specification and hyperparametrization of GPs with the computational efficiency of state space models.



- Codes for the examples available at:
  `http://arno.solin.fi`
- The methods also implemented into the GPstuff toolbox for Matlab/Octave.