Chapter 7 - Hierarchical dynamical spatio-temporal models

Odd Kolbjørnsen & Geir Storvik

April 3. 2017

Last time-: Chapter 6:Focus process models

Spatio-temporal covariance functions (sec 6.1, 6.2)

- Stationarity: spatio-temporal, Spatial, temporal
- Spatio-temporal - Kriging
- Seperable correlation function
- Additive correlation functions = independent Additive models
- Multiplicative correlation functions

Last time-: Chapter 6:Focus process models

Stochastic differential/difference equations (sec 6.3)

- Integro-difference equation models
- Using (partial) differential equations (what is the correlation structure?)
- Diffusion-injection models (interpretation of terms)
- Blurring space: Hard in space-time Simple in Fourier domain
- The Matern correlation functions is given by a specific SPDE.
- Discretization

Last time-: Chapter 6:Focus process models

Time series of spatial processes (sec 6.4)

- AR(q) process in time
- Stationary transitions
- Stationary distributions
- Discretization in space as well gives Vector-AR

Today

Hierarchical Dynamical Spatio-Temporal Models

- Data in Process models
- Observation types
- Linear observations
- Kalman filter
- Kalman smoother
- nonlinear/non Gaussian

Bayesian approach: Also include model for parameters

Data models

$$[\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta}_D] = ?$$

Simplifying assumptions:

- $[\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta}_D] = \prod_{t=1}^{T}[\mathbf{Z}_t|\mathbf{Y};\boldsymbol{\theta}_D]$ (same spatial locations for all times)
- $[\mathbf{Z}_t|\mathbf{Y}_t;\boldsymbol{\theta}_D] = \prod_{i=1}^{m_t}[Z(\mathbf{s}_i,t)|\mathbf{Y};\boldsymbol{\theta}_D]$ (one time step alone)
- $[Z(\mathbf{s},t)|\mathbf{Y}_t;\boldsymbol{\theta}_D] = [Z(\mathbf{s};r)|\{Y(\mathbf{x};r):\mathbf{x}\in\mathcal{N}_s, r\in\mathcal{N}_t\},\boldsymbol{\theta}_D]$
  Here
    - $\mathcal{N}_x$: Spatial neighborhood influencing observation $Z(\mathbf{x};r)$
    - $\mathcal{N}_r$: Temporal neighborhood influencing observation $Z(\mathbf{x};r)$

Linear mappings with equal dimensions

$$
\begin{aligned}
Z(\mathbf{s};t) =& Y(\mathbf{s};t) + \varepsilon(\mathbf{s};t), \quad \varepsilon(\mathbf{s};t) \overset{iid}{\sim} N(0,\sigma^2_\varepsilon)\\
\mathcal{N}_s =& \{s\}, \mathcal{N}_t = \{t\}
\end{aligned}
$$

Vector/Matrix formulation

$$
\mathbf{Z}_t = \mathbf{Y}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0,\sigma^2_\varepsilon \mathbf{I})
$$

- Only one parameter σ^2_ε
- Reasonable model in many cases

Linear mappings with equal dimensions

$$
\begin{aligned}
Z(\mathbf{s};t) =& Y(\mathbf{s};t) + \varepsilon(\mathbf{s};t), \quad \varepsilon(\mathbf{s};t) \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)\\
\mathcal{N}_s =& \{s\}, \mathcal{N}_t = \{t\}
\end{aligned}
$$

Vector/Matrix formulation

$$
\mathbf{Z}_t = \mathbf{Y}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2 \mathbf{I})
$$

- Only one parameter σ_ε^2
- Reasonable model in many cases

Extension

$$
\begin{aligned}
Z(\mathbf{s};t) =& a + hY(\mathbf{s};t) + \varepsilon(\mathbf{s};t)\\
\mathbf{Z}_t =& a\mathbf{1} + \text{diag}(\mathbf{h})\mathbf{Y}_t + \boldsymbol{\varepsilon}_t
\end{aligned}
$$

Further extensions

$$
\begin{aligned}
\mathbf{Z}_t =& \mathbf{a}_t + \text{diag}(\mathbf{h}_t)\mathbf{Y}_t + \boldsymbol{\varepsilon}_t\\
\mathbf{Z}_t =& \mathbf{a}_t + \mathbf{H}_t\mathbf{Y}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \mathbf{R}_t)\\
\mathcal{N}_s =& \{\mathbf{x}; H_{t;\mathbf{s},\mathbf{x}} \neq 0\} \qquad \text{Note: Can depend on } t
\end{aligned}
$$

Linear mappings with unequal dimensions

General case of linear observations:

$$\mathbf{Z}_t = \mathbf{a}_t + \mathbf{H}_t\mathbf{Y}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \mathbf{R}_t)$$

Here $\mathbf{H}_t$ is an $m_t \times n$ matrix

Example: $\mathbf{H}_t$ is then an incidence matrix

- Observed $Z_t(\mathbf{s}_{t,1}), ..., Z_t(\mathbf{s}_{t,m_t})$ at time t
- $h_{t;i,j} = I(\mathbf{s}_i = \mathbf{s}_{t,j})$

Multiple sources of data

Different observations $\mathbf{Z}_t^{(1)}$ and $\mathbf{Z}_t^{(2)}$.
Common assumption is conditional independence:

$$[\mathbf{Z}_t^{(1)}, \mathbf{Z}_t^{(2)} | \mathbf{Y}_t; \boldsymbol{\theta}_D] = [\mathbf{Z}_t^{(1)} | \mathbf{Y}_t; \boldsymbol{\theta}_D^{(1)}] \times [\mathbf{Z}_t^{(2)} | \mathbf{Y}_t; \boldsymbol{\theta}_D^{(2)}]$$

Interpretation: The observation process leading to $\mathbf{Z}_t^{(1)}$ and $\mathbf{Z}_t^{(2)}$ are independent. The measurements are related to the same feature
Note:

- You will typically see correlation between $\mathbf{Z}_t^{(1)}$ and $\mathbf{Z}_t^{(2)}$.
- Due to common dependence of $\mathbf{Y}_t$.

Examples

- Data from different satellites
- Satellite data combined with ground measurements
- Satellite data combined with output from numerical models

Change of support

Consider one dimension, $\mathbf{s} = i$.
Assume data

$$
\begin{aligned}
Z_{t,i}^{(1)} =& Y_t(i) + \varepsilon_t^{(1)}(i), & i = 1, ..., n\\
Z_{t,i}^{(2)} =& \frac{1}{2}[Y_t(i) + Y_t(i+1)] + \varepsilon_t^{(2)}(i), & i = 1, 3, 5, .., n
\end{aligned}
$$

Gives

$$
\mathbf{H}_t = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0\\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \cdots & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0\\
0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 & 0\\
\vdots & & & & & & & & \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & \frac{1}{2}
\end{pmatrix}
$$

Two types of area support for data (Change of support)

Continuous space

$$Z_t^{(1)}(\mathbf{s}_j) = Y(\mathbf{s}_j) + \varepsilon_t^{(1)}(\mathbf{s}_j)$$
$$Z_t^{(2)}(\mathbf{s}_k) = \frac{1}{|R|}\int_{\mathbf{v}\in R} Y(\mathbf{s}_k + \mathbf{v})d\mathbf{r} + \varepsilon_t^{(2)}(\mathbf{s}_k)$$

where $\{\varepsilon^{(1))}(\mathbf{s}_j)\}$ and $\{\varepsilon^{(2)}(\mathbf{s}_k)\}$ are independent (i.e. conditional independence)
We have for data type 1 (as we know from earlier):

$$\text{cov}[Y_t(\mathbf{s}), Y_r(\mathbf{s}')] = C_Y(\mathbf{s}-\mathbf{s}'; t-r)$$
$$\text{cov}[Y_t(\mathbf{s}), Z_r^{(1)}(\mathbf{s}')] = C_Y(\mathbf{s}-\mathbf{s}'; t-r)$$
$$\text{cov}[Z_t^{(1)}(\mathbf{s}), Z_r^{(1)}(\mathbf{s}')] = C_Y(\mathbf{s}-\mathbf{s}'; t-r) + \sigma^2_{\varepsilon_t^{(1)}} I(\mathbf{s}=\mathbf{s}', t=r)$$

Correlation to data with change of support

We have for data type 2 (as we know from earlier):

$$
\begin{aligned}
\text{cov}[Y_t(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')] =& \frac{1}{|R|}\int_{\mathbf{v}'\in R} \text{cov}[Y_t(\mathbf{s}), Y_r(\mathbf{s}'+\mathbf{v}')]d\mathbf{v}'\\
=& \frac{1}{|R|}\int_{\mathbf{v}'\in R} C_Y(\mathbf{s}-\mathbf{s}'-\mathbf{v}'; t-r)d\mathbf{v}'\\
=&\\
\text{cov}[Z_t^{(2)}(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')] =& \frac{1}{|R|^2}\int_{\mathbf{v}\in R}\int_{\mathbf{v}'\in R} \text{cov}[Y_t(\mathbf{s}+\mathbf{v}), Y_r(\mathbf{s}'+\mathbf{v}')]d\mathbf{v}\\
&+\sigma^2_{\varepsilon_t^{(2)}} I(\mathbf{s}=\mathbf{s}', t=r)\\
=& \frac{1}{|R|^2}\int_{\mathbf{v}\in R}\int_{\mathbf{v}'\in R} C_Y(\mathbf{s}-\mathbf{s}'+\mathbf{v}-\mathbf{v}'; t-r)d\mathbf{v}d\mathbf{v}'\\
&+\sigma^2_{\varepsilon_t^{(2)}} I(\mathbf{s}=\mathbf{s}', t=r)
\end{aligned}
$$

Change of support cont.

The three by three block matrix of correlations needed for Kriging consist of:

$$\text{cov}[Y_t(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')], \text{cov}[Y_t(\mathbf{s}), Z_r^{(1)}(\mathbf{s}')], \text{cov}[Y_t(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')],$$
$$\text{cov}[Z_t^{(2)}(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')], \text{cov}[Z_t^{(1)}(\mathbf{s}), Z_r^{(1)}(\mathbf{s}')], \text{cov}[Z_t^{(1)}(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')]$$

The last one is the only "new".

$$\begin{aligned}\text{cov}[Z_t^{(1)}(\mathbf{s}), Z_r^{(2)}(\mathbf{s}')] &= \frac{1}{|R|}\int_{\mathbf{v}'\in R} \text{cov}[Y_t(\mathbf{s}), Y_r(\mathbf{s}'+\mathbf{v}')]d\mathbf{v}' \\ &= \frac{1}{|R|}\int_{\mathbf{v}'\in R} C_Y(\mathbf{s}-\mathbf{s}'-\mathbf{v}'; t-r)d\mathbf{v}'\end{aligned}$$

Kalman filtering, Note: slightly confusing notation

State space model

$$\mathbf{Y}_t = \mathbf{M}_t \mathbf{Y}_{t-1} + \boldsymbol{\delta}_t, \qquad \boldsymbol{\delta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \quad \text{NB: Q is a covariance}$$
$$\mathbf{Z}_t = \mathbf{H}_t \mathbf{Y}_t + \varepsilon_t, \qquad \varepsilon_t \sim N(\mathbf{0}, \mathbf{R}_t)$$

Notation:

$$\mathbf{Y}_{t|s} = E[\mathbf{Y}_t | \mathbf{Z}_{1:s}] \qquad \text{NB: } Y_{t|s} \text{is an expectation}$$
$$\mathbf{P}_{t|s} = E[(\mathbf{Y}_t - \mathbf{Y}_{t|s})(\mathbf{Y}_t - \mathbf{Y}_{t|s})' | \mathbf{Z}_{1:s}]$$

Gaussian process model and linear Gaussian observations:
Forecast distribution:

$$p(\mathbf{Y}_t | \mathbf{Z}_{1:(t-1)}) = N(\mathbf{Y}_{t|t-1}, \mathbf{P}_{t|t-1})$$

Filtering distribution:

$$p(\mathbf{Y}_t | \mathbf{Z}_{1:t}) = N(\mathbf{Y}_{t|t}, \mathbf{P}_{t|t})$$

Kalman filtering sec 8.2. Linear Gaussin models

Kalman filter condition to data as they come in, and assimilate them in the distribution. This is done in a sequence of forecast- filter operations.

Given filter distribution at time (t-1) forecast time t using the state space model, get the filter distribution by conditioning to data collected at time t. (Assimilate the data in the distribution) .

Forecast:

$$\mathbf{P}_{t|t-1} = \mathbf{M}_t \mathbf{P}_{t-1|t-1} \mathbf{M}_t' + \mathbf{Q}_t \quad \mathbf{Y}_{t|t-1} = \mathbf{M}_t \mathbf{Y}_{t-1|t-1}$$

Filter:

$$\begin{aligned} \mathbf{S}_t &= \mathbf{H}_t' \mathbf{P}_{t|t-1} \mathbf{H}_t + \mathbf{R}_t & \mathbf{K}_t &= \mathbf{P}_{t|t-1} \mathbf{H}_t' \mathbf{S}_t^{-1} \quad \text{(Kalman gain)} \\ \mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1} & \mathbf{Y}_{t|t} &= \mathbf{Y}_{t|t-1} + \mathbf{K}_t (\mathbf{Z}_t - \mathbf{H}_t \mathbf{Y}_{t|t-1}) \end{aligned}$$

Dimension reduction

If computations are too heavy. Try dimension reduction:

$$
\begin{aligned}
\mathbf{Y}_t =& \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \mathbf{v}_t, && \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{B}_t) \\
\boldsymbol{\alpha}_t =& \mathbf{M}_{\alpha,t}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\gamma}_t && \boldsymbol{\gamma}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \\
\mathbf{Z}_t =& \mathbf{H}_t\mathbf{Y}_t + \varepsilon_t, && \varepsilon_t \sim N(\mathbf{0}, \mathbf{R}_t)
\end{aligned}
$$

where $\dim(\boldsymbol{\alpha}_t) \ll \dim(\mathbf{Y}_t)$.
Can be rewritten as

$$
\begin{aligned}
\boldsymbol{\alpha}_t =& \mathbf{M}_{\alpha,t}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\gamma}_t \\
\mathbf{Z}_t =& \mathbf{H}_t\boldsymbol{\Phi}_t\boldsymbol{\alpha}_t + \mathbf{H}_t\mathbf{v}_t + \varepsilon_t \\
=& \widetilde{\mathbf{H}}_t\boldsymbol{\alpha}_t + \widetilde{\varepsilon}_t, && \widetilde{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{H}_t\mathbf{B}_t\mathbf{H}_t' + \mathbf{R}_t)
\end{aligned}
$$

Kalmanfiltering wrt $\boldsymbol{\alpha}_t$ instead
Choise of $\boldsymbol{\Phi}$:

- Empirical orthogonal functions
- Linear/polynomial functions of spatial coordinates

Kalman filtering vs smoothing

Kalman filter updates the latest time step with all data up until this time, i.e. $p(\mathbf{Y}_T|\mathbf{Z}_{1:T})$ Kalman filter does not compute $p(\mathbf{Y}_t|\mathbf{Z}_{1:T})$, for $t < T$, i.e. the distribution of an earlier time given all data in its past present and future. This can be done with the Kalman Smoother.

$$p(\mathbf{Y}_t|\mathbf{Z}_{1:T}) = N(\mathbf{Y}_{t|T}, \mathbf{P}_{t|T})$$

Recursive backward algorithm (given forward filter and forecast):

$$\begin{aligned}
\mathbf{Y}_{t|T} =& \mathbf{Y}_{t|t} + \mathbf{J}_t(\mathbf{Y}_{t+1|T} - \mathbf{Y}_{t+1|t}) \\
\mathbf{P}_{t|T} =& \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}_t' \\
\mathbf{J}_t =& \mathbf{P}_{t|t}\mathbf{M}_{t+1}'\mathbf{P}_{t+1|t}^{-1}
\end{aligned}$$

Nonlinear mappings/non-Gaussian data

In many problems (non linear observations)

$$\mathbf{Z}(\mathbf{s}, i; t) = h_{i,t}(\mathbf{Y}_t) + \varepsilon(\mathbf{s}_i; t)$$

- $h_{i,t}(\mathbf{Y}_t)$ known form but nonlinear (may depend on unknown parameter).

Alternatively (non Gaussian likelihood)

- $[\mathbf{Z}_t|\mathbf{Y}_t] = \text{Poisson}(\exp(\mathbf{X}_t\boldsymbol{\beta} + \mathbf{Y}_t + \mathbf{v}_t))$

Kalman filtering not possible to apply. Alternatives:

- Linear approximations, (Extended) Kalman filter
- Monte Carlo approximations, MCMC
- Ensemble Kalman filter
- Laplace approximations, INLA