Numerics of Stochastic Processes VII (09.01.2009)

Illia Horenko

Research Group "***Computational Time Series Analysis*** "
Institute of Mathematics
Freie Universität Berlin (FU)

DFG Research Center ***MATHEON***
„Mathematics in key technologies"

Model Distance Functional

Let $x(t) : \mathbf{R}^1 \rightarrow \Psi \subset \mathbf{R}^n$ be the *observed process* $t \in [0, T]$

Define $\mathbf{K}$ local models by a *model distance functional*:

$$g(x, \theta_i) \quad : \quad \Psi \times \Omega \rightarrow [0, \bar{g}], \quad 0 < \bar{g} < +\infty,$$

$$\theta_1, \ldots, \theta_{\mathbf{K}} \in \Omega \subset \mathbf{R}^d$$

Examples

- *Geometrical clustering*: $\theta_i \in \Psi$ - cluster centers

$$g(x, \theta_i) \quad = \quad \| x - \theta_i \|^2,$$

- *Gaussian clustering*: $\theta_i = (\mu_i, \Sigma_i)$ - Gaussian parameters

$$g(x, \theta_i) \quad = \quad \| x - \mu_i \|^2_{\Sigma_i^{-1}}$$

Averaged Clustering Functional

Find $\Gamma(t) = (\gamma_1(t), \ldots, \gamma_{\mathbf{K}}(t))$ such that for each t :

$$\mathbf{L}(\Theta, \Gamma(t)) = \int_0^T \sum_{i=1}^{\mathbf{K}} \gamma_i(t) g\left(x_t, \theta_i\right) \rightarrow \min_{\Gamma(t), \Theta},$$

subjected to constraints:

$$\begin{aligned} \sum_{i=1}^{\mathbf{K}} \gamma_i(t) &= 1, \quad \forall t \in [0, T] \\ \gamma_i(t) &\geq 0, \quad \forall t \in [0, T], i = 1, \ldots, \mathbf{K} \end{aligned}$$

Numerical Method: Subspace Iteration (splitting scheme)

No global convergence (non-convex optimization, simulated annealing)

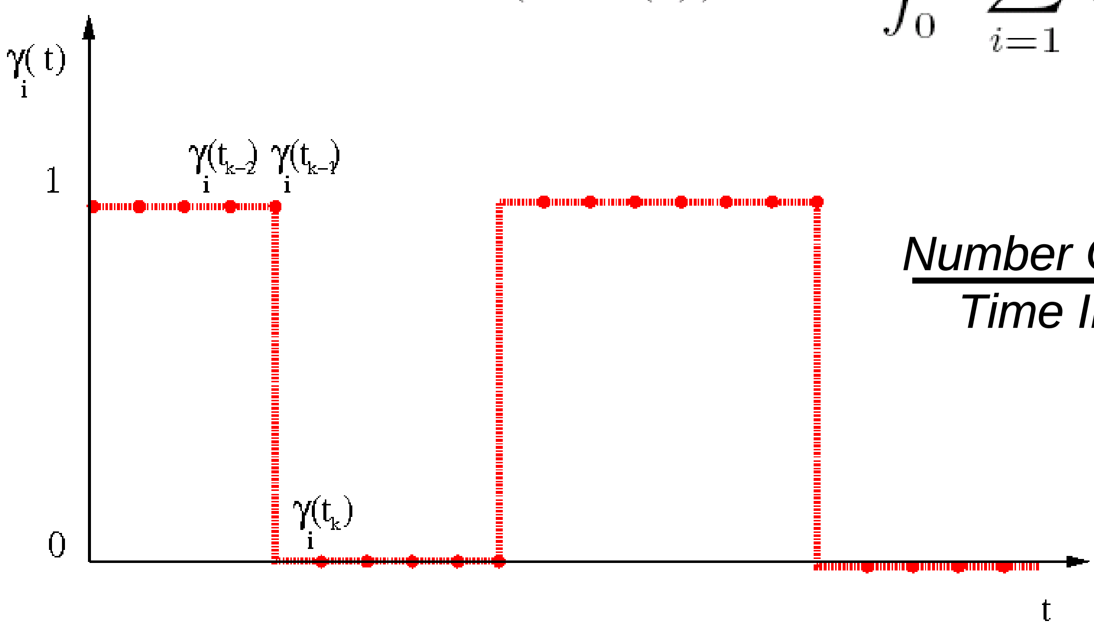
Incorporation of Temporal Information

Problem 2: *identification of "persistent states"*

Let $\gamma_i(\cdot)$ on $t \in [0,T], i = 1,\ldots,\mathbf{K}$ be differentiable and

$\partial_t \gamma_i \in \mathcal{L}_2(0,T)$, i. e. :

$$\mathbf{L}(\Theta,\Gamma(t)) = \int_0^T \sum_{i=1}^{\mathbf{K}} \gamma_i(t) g(x_t,\theta_i) \to \min_{\Gamma(t),\Theta},$$



$$\frac{\textit{Number Of Jumps}}{\textit{Time Interval}} = $$

$$= \sum_{k=1} \frac{\left(\gamma_i(t_{k+1}) - \gamma_i(t_k)\right)^2}{\Delta t}$$

Tikhonov-type Regularization

Problem 2: ***incorporation of temporal information***

Let $\gamma_i(\cdot)$ on $t \in [0,T], i = 1,\ldots,\mathbf{K}$ be differentiable and

$\partial_t \gamma_i \in \mathcal{L}_2(0,T)$, i. e. :

$$\mathbf{L}(\Theta,\Gamma(t)) = \int_0^T \sum_{i=1}^{\mathbf{K}} \gamma_i(t) g(x_t,\theta_i) \rightarrow \min_{\Gamma(t),\Theta},$$

subjected to

$$|\gamma_i|_{\mathcal{H}^1(0,T)} = \| \partial_t \gamma_i(\cdot) \|_{\mathcal{L}_2(0,T)} = \int_0^T (\partial_t \gamma_i(t))^2 \, dt \leq C_\epsilon^i < +\infty,$$

Regularized clustering functional:

$$\mathbf{L}^\epsilon(\Theta,\Gamma(t),\epsilon^2) = \mathbf{L}(\Theta,\Gamma(t)) + \epsilon^2 \sum_{i=1}^{\mathbf{K}} \int_0^T (\partial_t \gamma_i(t))^2 \, dt \rightarrow \min_{\Gamma(t),\Theta}$$

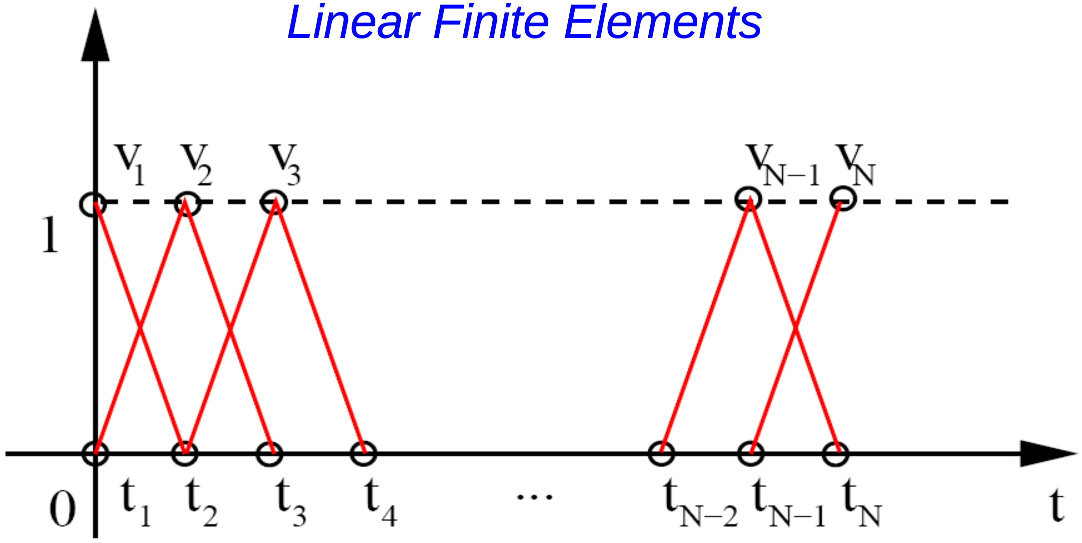
(H. 08, to appear in SISC)

FEM: Time Discretization

Let $\{0 = t_1, t_2, \ldots, t_{N-1}, t_N = T\}$. *We define a set of of continuous functions on* $[0, T]$ *, or finite elements set* $\{v_1(t), v_2(t), \ldots, v_N(t)\} \in \mathcal{L}_2(0, T)$ *with local support.*

Example:

Linear Finite Elements



FEM: Regularized Clustering Functional

Regularized clustering functional: *(H. 08, to appear in SISC)*

$$\mathbf{L}^{\epsilon}(\Theta,\Gamma(t),\epsilon^2) \quad = \quad \mathbf{L}(\Theta,\Gamma(t)) + \epsilon^2 \sum_{i=1}^{\mathbf{K}} \int_0^T \left(\partial_t \gamma_i\left(t\right)\right)^2 dt \longrightarrow \min_{\Gamma(t),\Theta},$$

FEM: Regularized Clustering Functional

Regularized clustering functional: (H. 08, to appear in SISC)

$$\mathbf{L}^{\epsilon}(\Theta,\Gamma(t),\epsilon^2) \quad = \quad \mathbf{L}(\Theta,\Gamma(t)) + \epsilon^2 \sum_{i=1}^{\mathbf{K}} \int_0^T \left(\partial_t \gamma_i\left(t\right)\right)^2 dt \longrightarrow \min_{\Gamma(t),\Theta},$$

Galerkin-Ansatz :

$$\begin{aligned} \gamma_i\left(t\right) &= \tilde{\gamma}_i\left(t\right) + \delta_N \\ &= \sum_{k=1}^{N} \tilde{\gamma}_i^{(k)} v_k\left(t\right) + \delta_N \end{aligned}$$

where $\tilde{\gamma}_i^{(k)} = \int_0^T \gamma_i\left(t\right) v_k\left(t\right) dt$, and

$$\mathbf{L}^{\epsilon} \quad = \quad \tilde{\mathbf{L}}^{\epsilon} + \mathcal{O}\left(\delta_N\right) \longrightarrow \min_{\tilde{\gamma}_i(t),\Theta},$$

$$\tilde{\mathbf{L}}^{\epsilon} = \sum_{i=1}^{\mathbf{K}} \int_0^T \left[\tilde{\gamma}_i\left(t\right) g(x_t,\theta_i) + \epsilon^2 \left(\partial_t \tilde{\gamma}_i\left(t\right)\right)^2\right] dt.$$

FEM-Discretized Clustering Functional

(H. 08, to appear in SISC)

$$\tilde{\mathbf{L}}^{\epsilon} = \sum_{i=1}^{\mathbf{K}} \left[a^{\mathbf{T}}(\theta_i)\bar{\gamma}_i + \epsilon^2 \bar{\gamma}_i^{\mathbf{T}} \mathbf{H} \bar{\gamma}_i \right] \rightarrow \min_{\bar{\gamma}_i, \Theta}$$

subjected to

$$\sum_{i=1}^{\mathbf{K}} \tilde{\gamma}_i^{(k+1)} = 1, \quad \forall k = 1, \ldots, N,$$

$$\tilde{\gamma}_i^{(k+1)} \geq 0, \quad \forall k = 1, \ldots, N; i = 1, \ldots, \mathbf{K}.$$

Iterative Subspace Minimization: ***sparse QP*** can be used

where $a(\theta_i) = \left(\int_{t_1}^{t_2} v_1(t) g(x_t, \theta_i) dt, \ldots, \int_{t_{N-1}}^{t_N} v_N(t) g(x_t, \theta_i) dt \right)$

is a vector of *FEM-discretized model distances* and $\mathbf{H}$ is

a *mass-matrix* of the *FEM-basis*

Algorithm: monotony conditions

Theorem *Let for a given observed time series* $x(t) : \mathbf{R}^1 \to \Psi \subset \mathbf{R}^n$*, the model distance functional is chosen such that it satisfies (2),* Ψ *and* Ω *are compact,* $g(x_t, \cdot)$ *is continuously differentiable function of* θ *and*

$$\frac{\partial}{\partial \Theta} \tilde{\mathbf{L}}^{\epsilon} \left(\Theta^*, \bar{\gamma}\right) \quad = \quad 0,$$

has a unique solution $\Theta^* = \left(\theta_1^*, \ldots, \theta_{\mathbf{K}}^*\right), \theta_i* \in \Omega$ *for any fixed* $\bar{\gamma}$ *satisfying (18-19) and* $\frac{\partial^2}{\partial \Theta^2} \tilde{\mathbf{L}}^{\epsilon} \left(\Theta^*, \bar{\gamma}\right)$ *is positive definite. Then for any* $\epsilon^2 \geq 0$ *and any finite continuous non-negative finite elements set* $\{v_1(t), v_2(t), \ldots, v_N(t)\} \in \mathcal{L}_2(0, T)$ *such that the respective* mass-matrix $\mathcal{H}$ *is positive definite, the above algorithm is* monotonous, *i. e., for any* $j \geq 1$

$$\tilde{\mathbf{L}}^{\epsilon} \left(\Theta^{[j+1]}, \bar{\gamma}_i^{[j+1]}\right) \quad \leq \quad \tilde{\mathbf{L}}^{\epsilon} \left(\Theta^{[j]}, \bar{\gamma}_i^{[j]}\right).$$

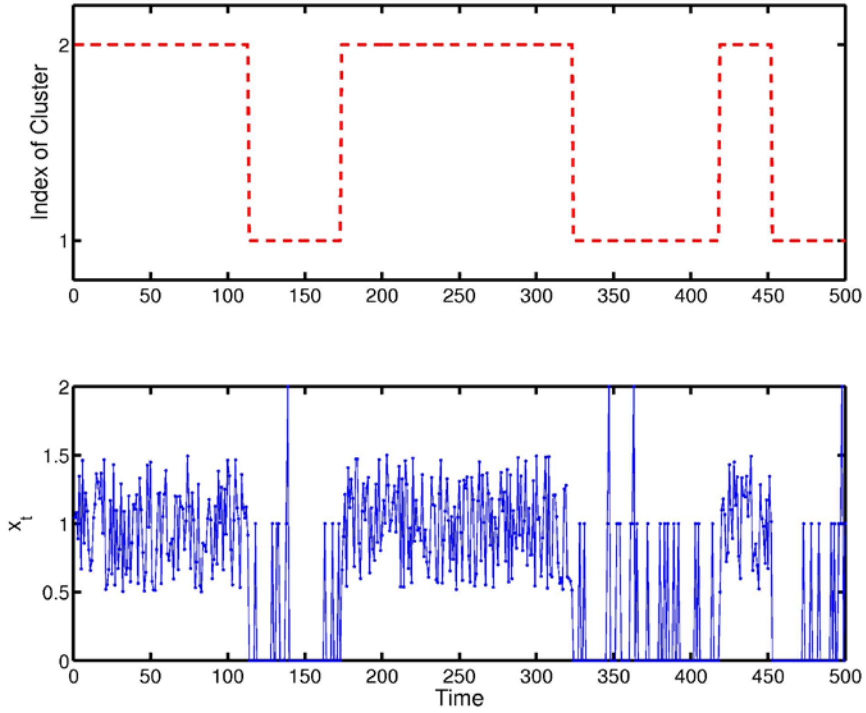
Convergence to a local optimum only!

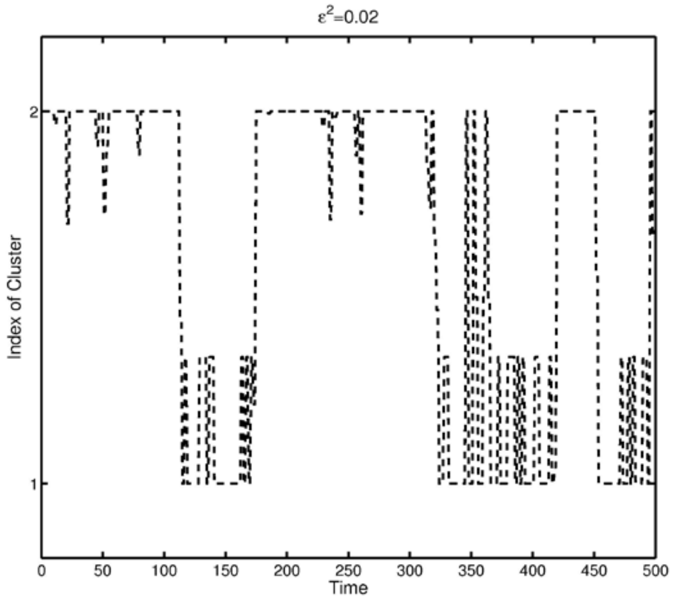
(coupling to some *global opimizer* necessary)

Toy Example I

$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$

$$g\left(x, \theta_i\right) \quad = \quad \| x - \theta_i \|^2,$$

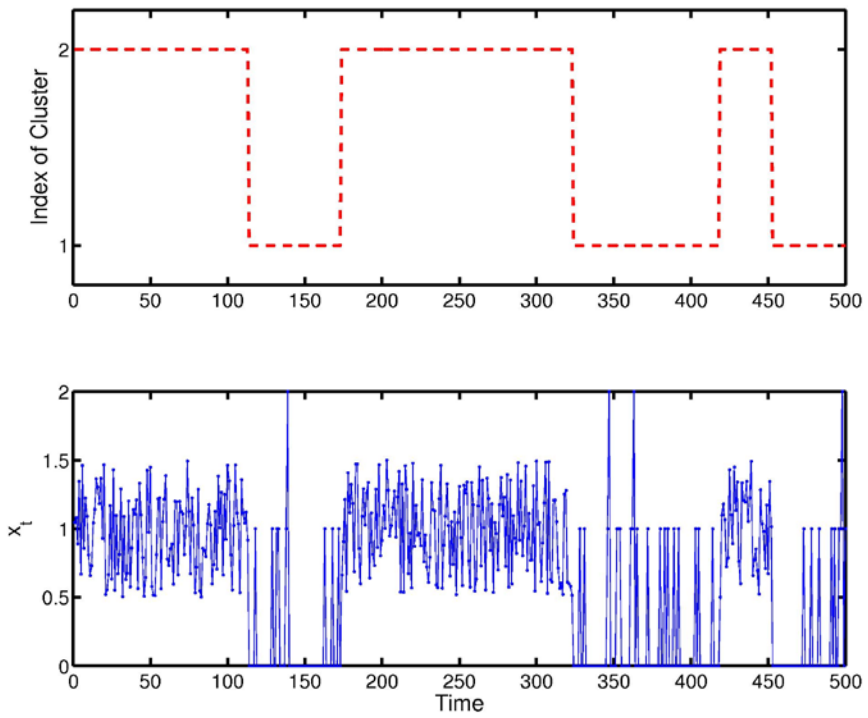


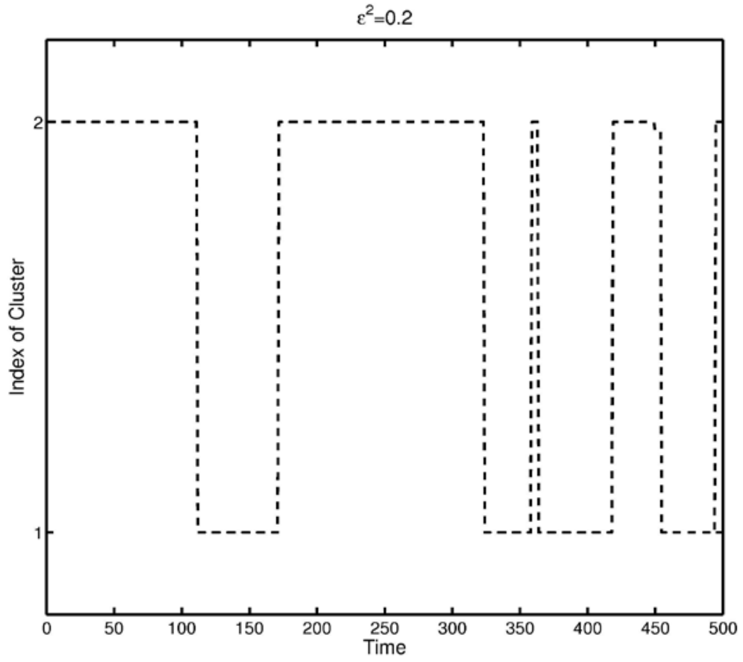


Toy Example I

$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$

$$g\left(x, \theta_i\right) = \| x - \theta_i \|^2,$$

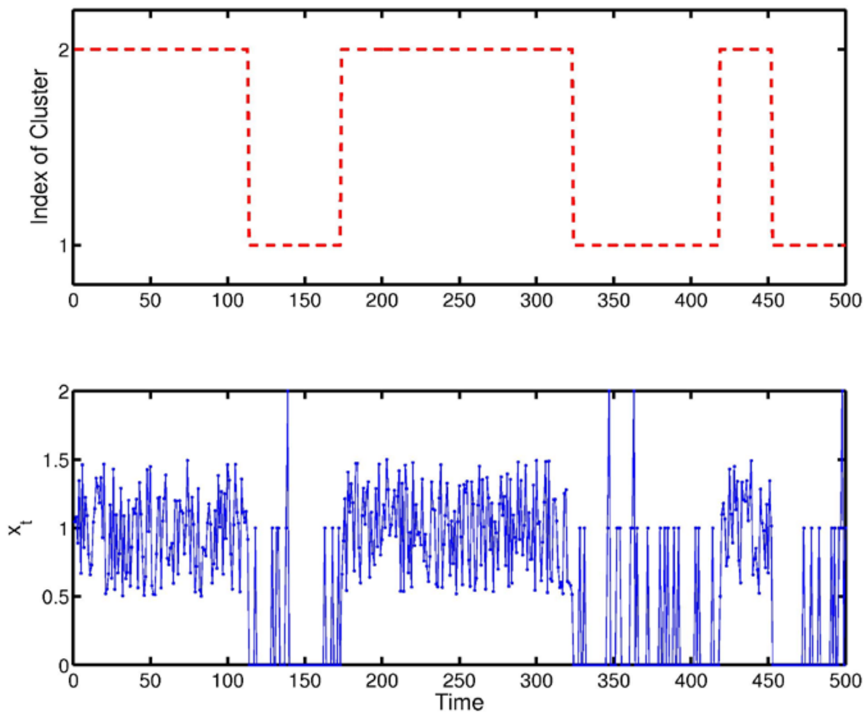


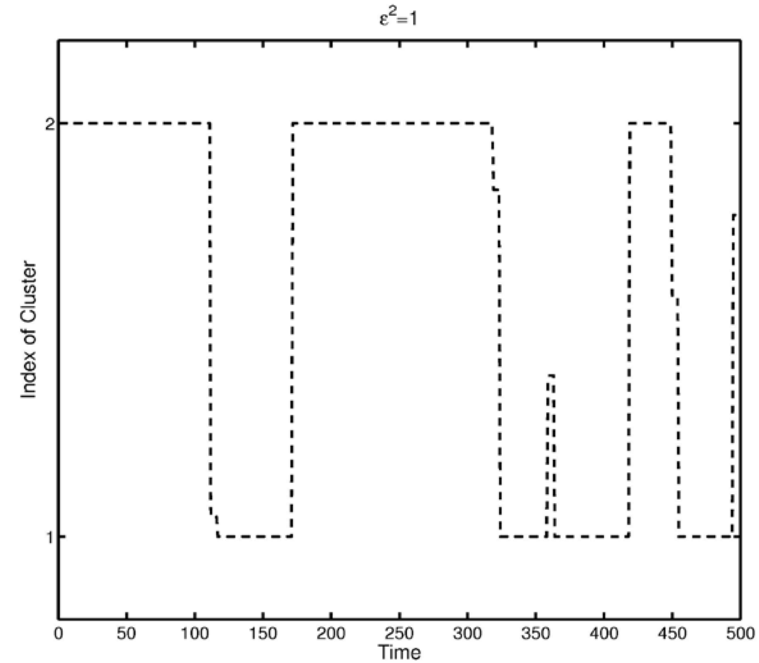


Toy Example I

$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$

$$g\left(x, \theta_i\right) \quad = \quad \| \, x - \theta_i \, \|^2,$$

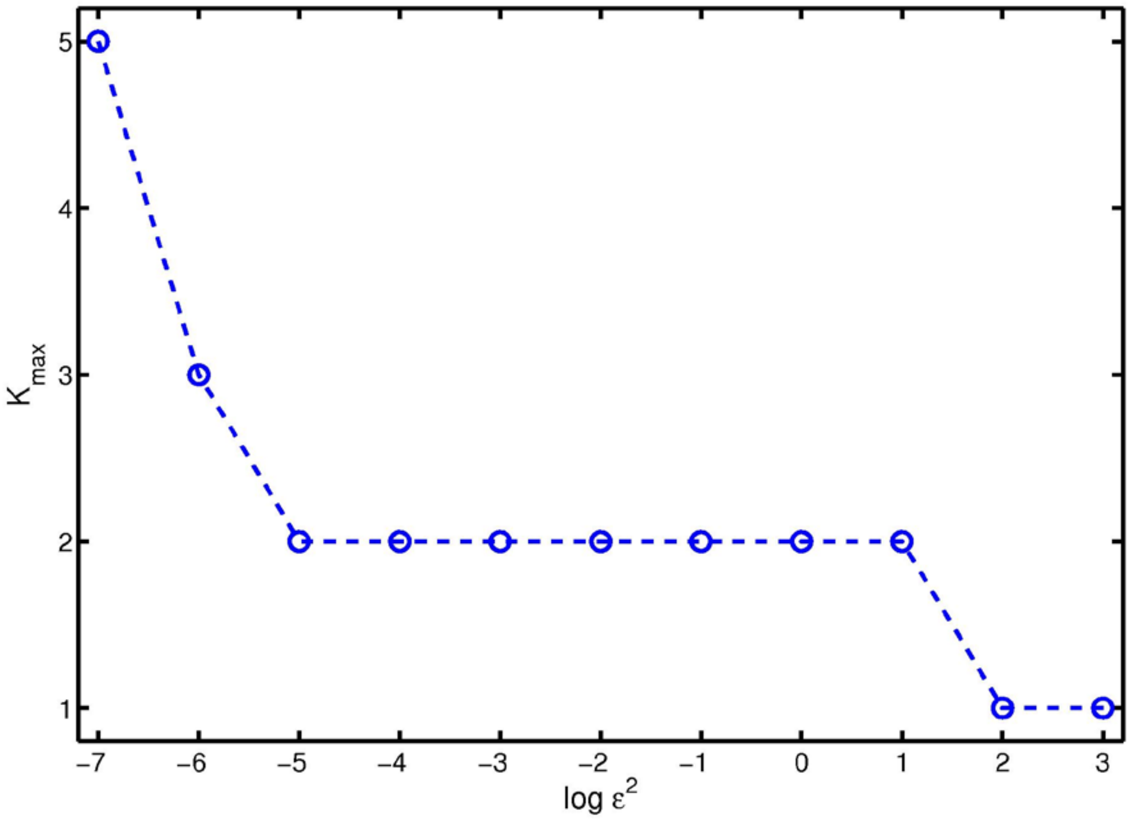




Toy Example I

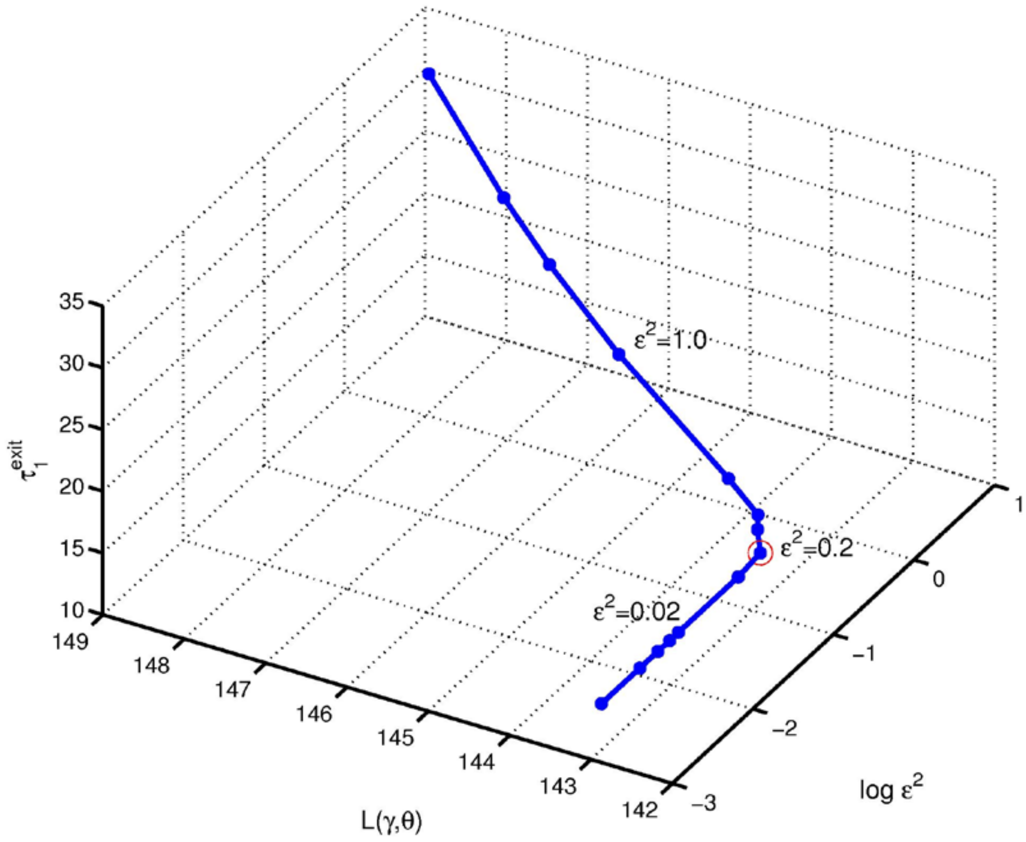
How to determine the optimal K: probabilistic model assumptions a posteriory

$$g(x, \theta_i) = \| x - \theta_i \|^2,$$



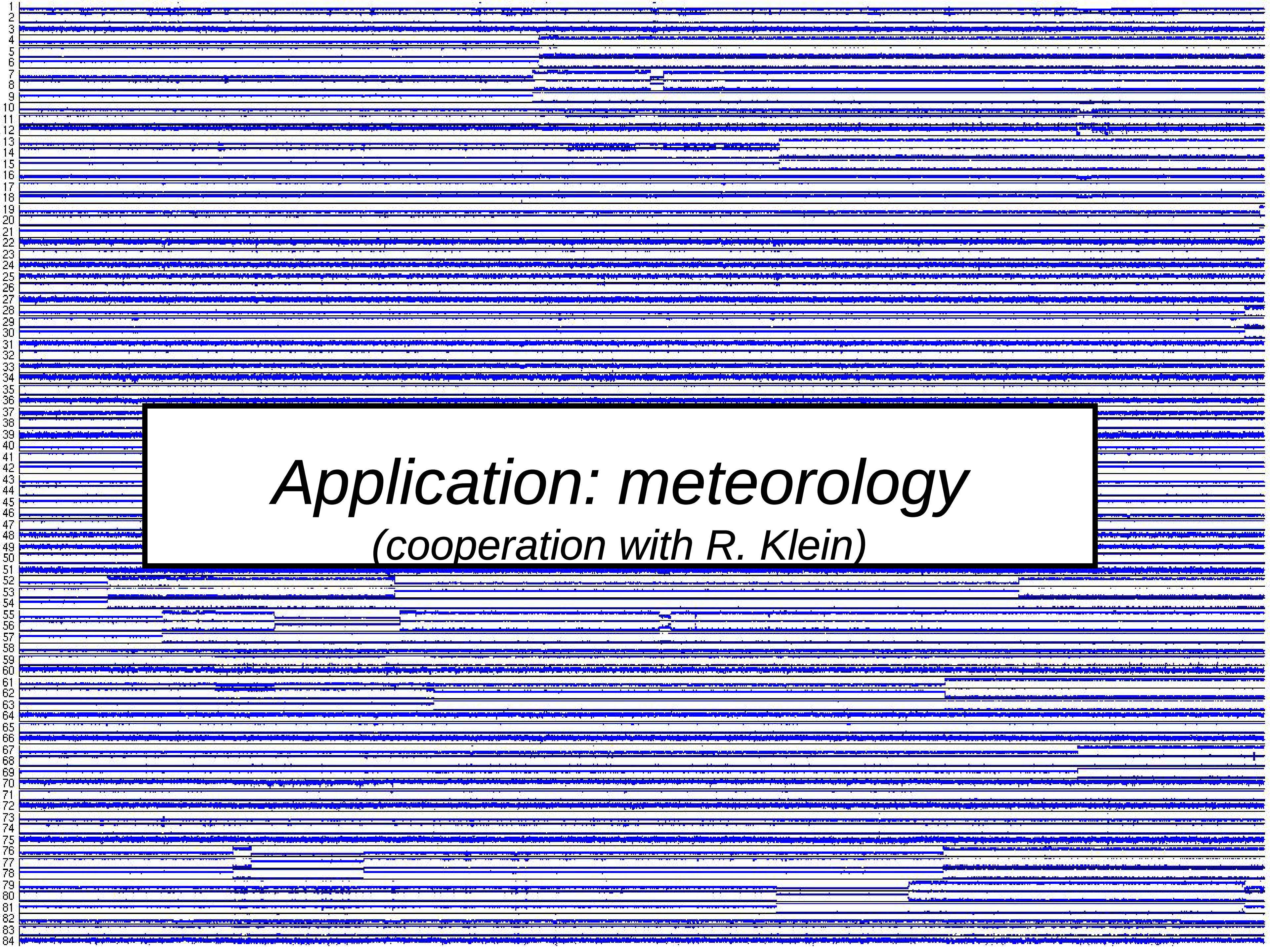
Toy Example I

How to determine the optimal ε: standard L-Curve approach from Tikhonov-regularized linear least-squares problems (Cullum(79), Hansen(99))



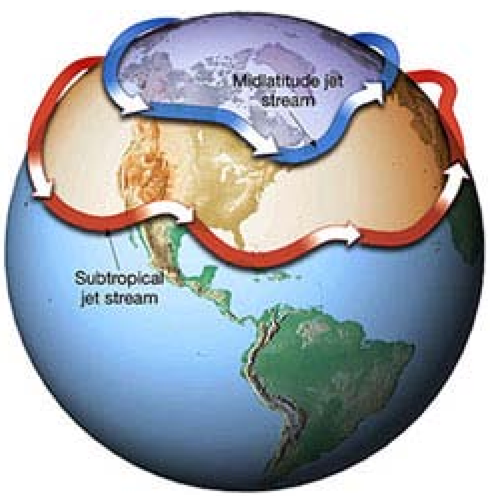
Application: meteorology

(cooperation with R. Klein)



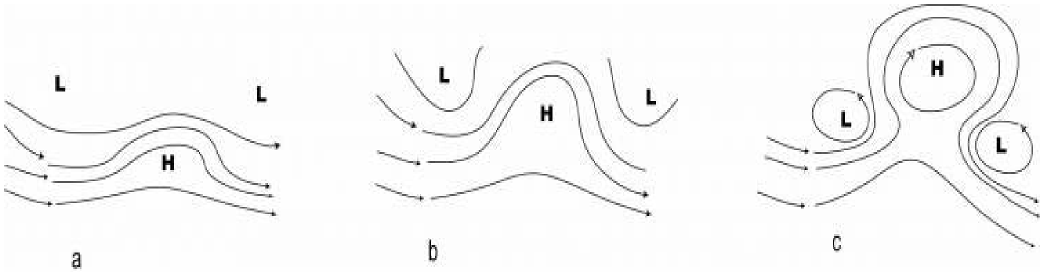
“Hidden“ Weather Phases: motivation

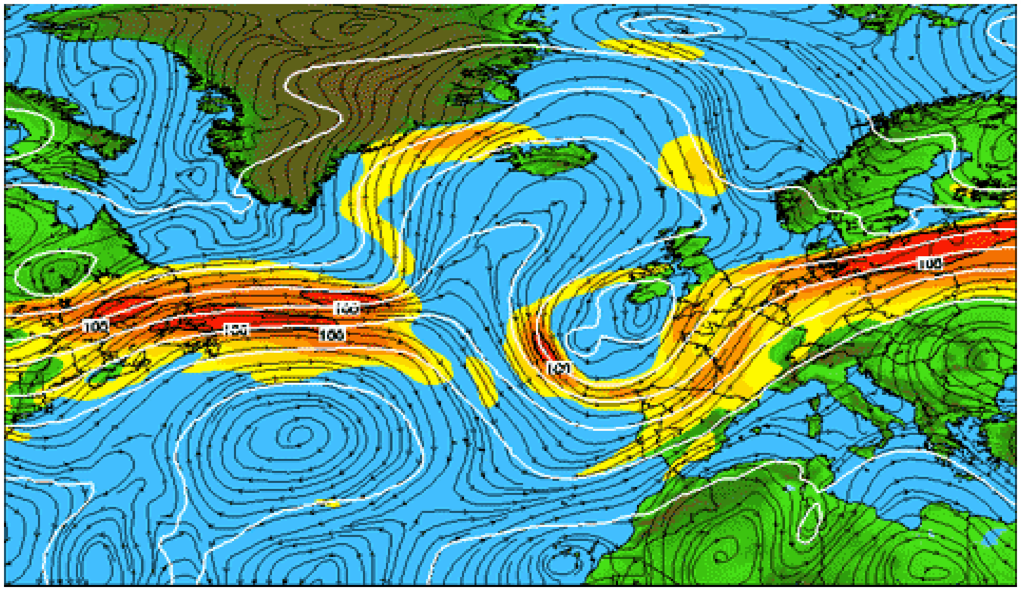
Wind Jets transport moisture from US to Europe



Weather Data in Europe: 29x20 grid (44 years)
(Data from H.Osterle, PIK)
Up to 4 hidden states are statistically separable

Jet-Blocking

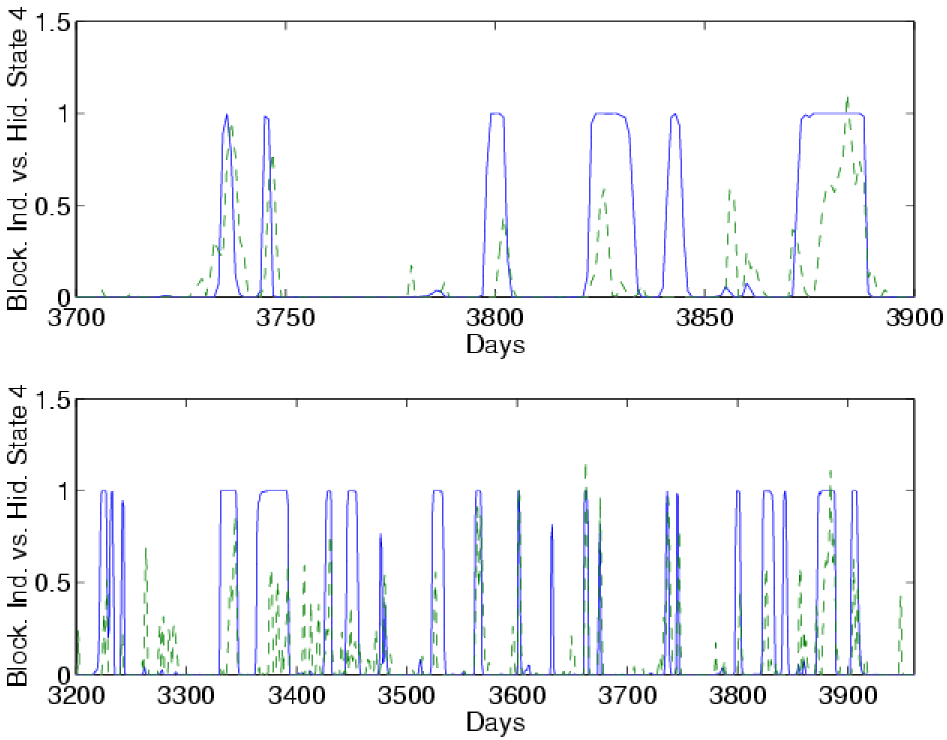




$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$: K=4

Hidden State 4: Jet Blocking Situation

Up to 4 hidden states are statistically separable



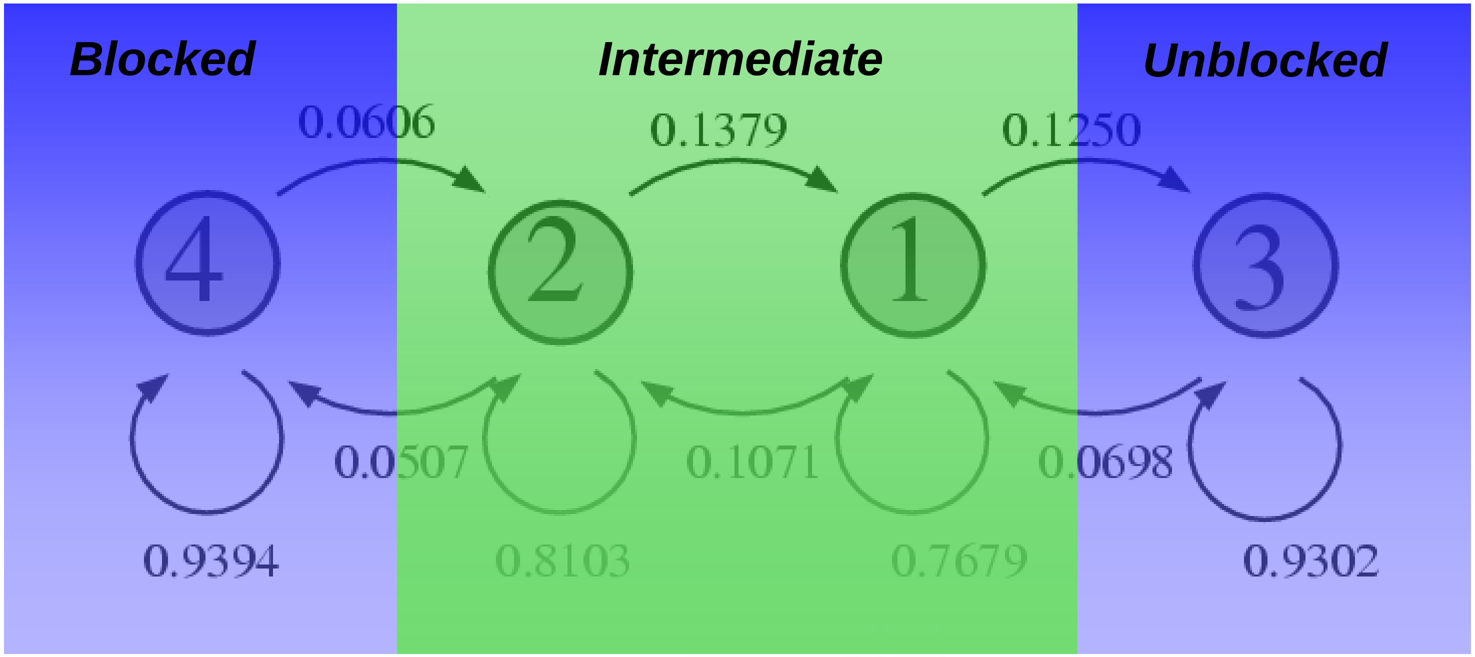
Comparison with Lejenas-Okland blocking index

$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$

$$g\left(x, \theta_i\right) \quad = \quad \| \, x - \mathcal{T}_i \mathcal{T}_i^{\mathbf{T}} x \, \|^2$$

$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$ *: K=4*

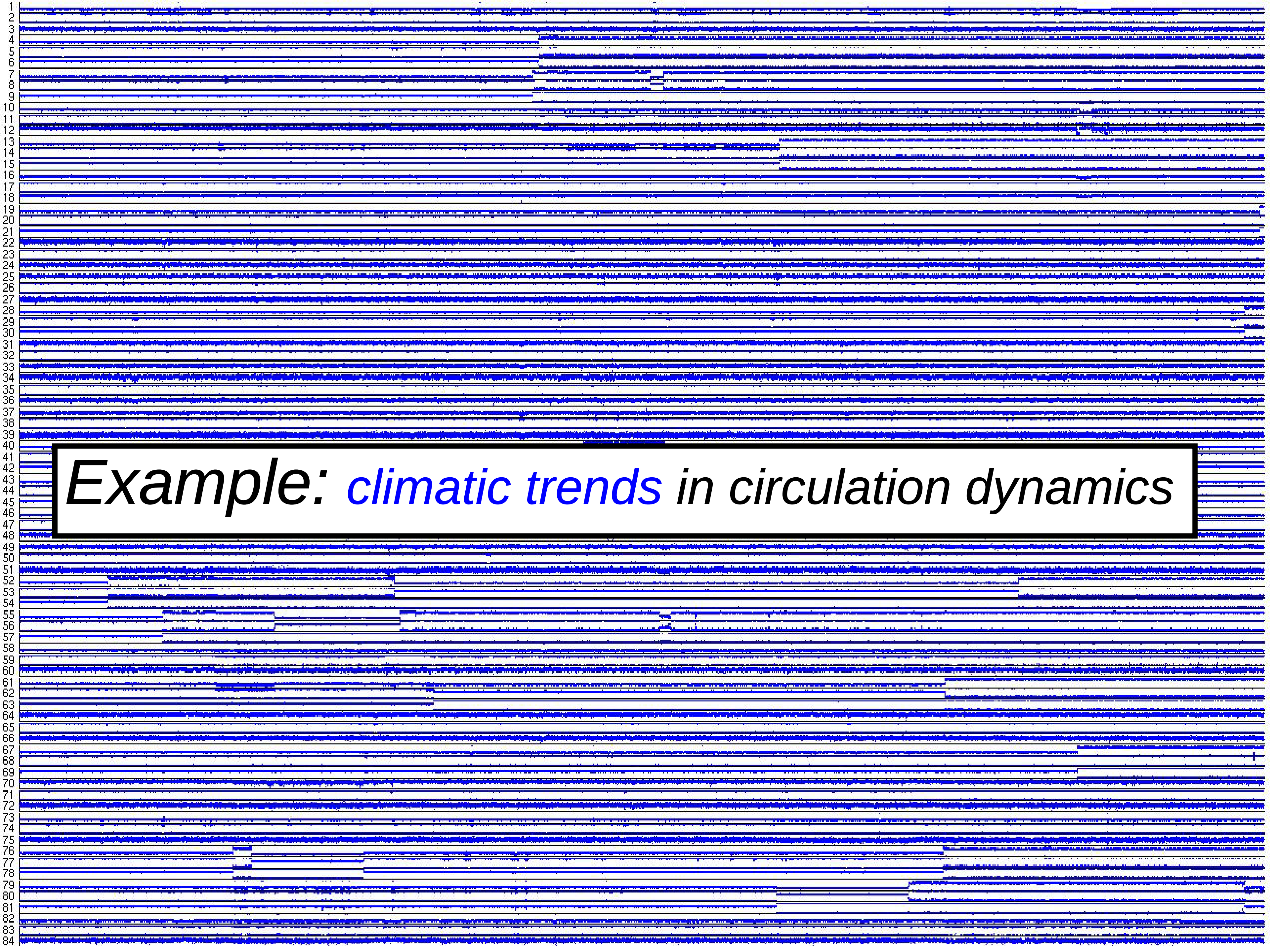
SPP 1276 “MetStröm“ Project



Looking through Markovian  ***: predictions***

$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$

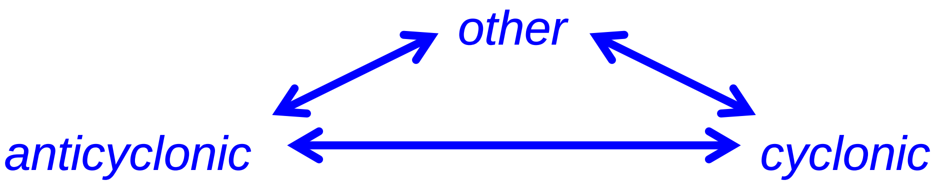
$$g\left(x, \theta_i\right) \quad = \quad \| \, x - \mathcal{T}_i \mathcal{T}_i^{\mathbf{T}} x \, \|^2$$


Example: climatic trends in circulation dynamics

Circulation Patterns for UK(1945-2007)

Historical Circulation Data: weather regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

Lamb, Geophys. Mem. (1972),
Jones/Hulme/Briffa, Int.J.of Climat. (1977),



$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \rightarrow \min$$

$$x_t : X_t \rightarrow X_{t+1}$$

$$g(x_t, P^i) \quad = \quad -\log P^i_{X_t X_{t+1}}$$

Clustering of Markovian Transitions with FEM-Clustering

Memo II: Markov Chain Log-Likelihood

Observed Time Series: $\{X_1, \ldots, X_T\},\ X_t \in s_1, \ldots, s_m$

Markov-Property:

$$\mathrm{P}\left[X_t = s_j | X_1, X_2, \ldots, X_{t-1} = s_i\right] = \mathrm{P}\left[X_t = s_j | X_{t-1} = s_i\right] = P_{ij}(t)$$

Log-Likelihood:

$$\begin{aligned} \sum_{j=1}^{m} P_{ij}(t) &= 1, \quad \text{for all } t, i \\ P_{ij}(t) &\geq 0, \quad \text{for all } t, i, j \end{aligned}$$

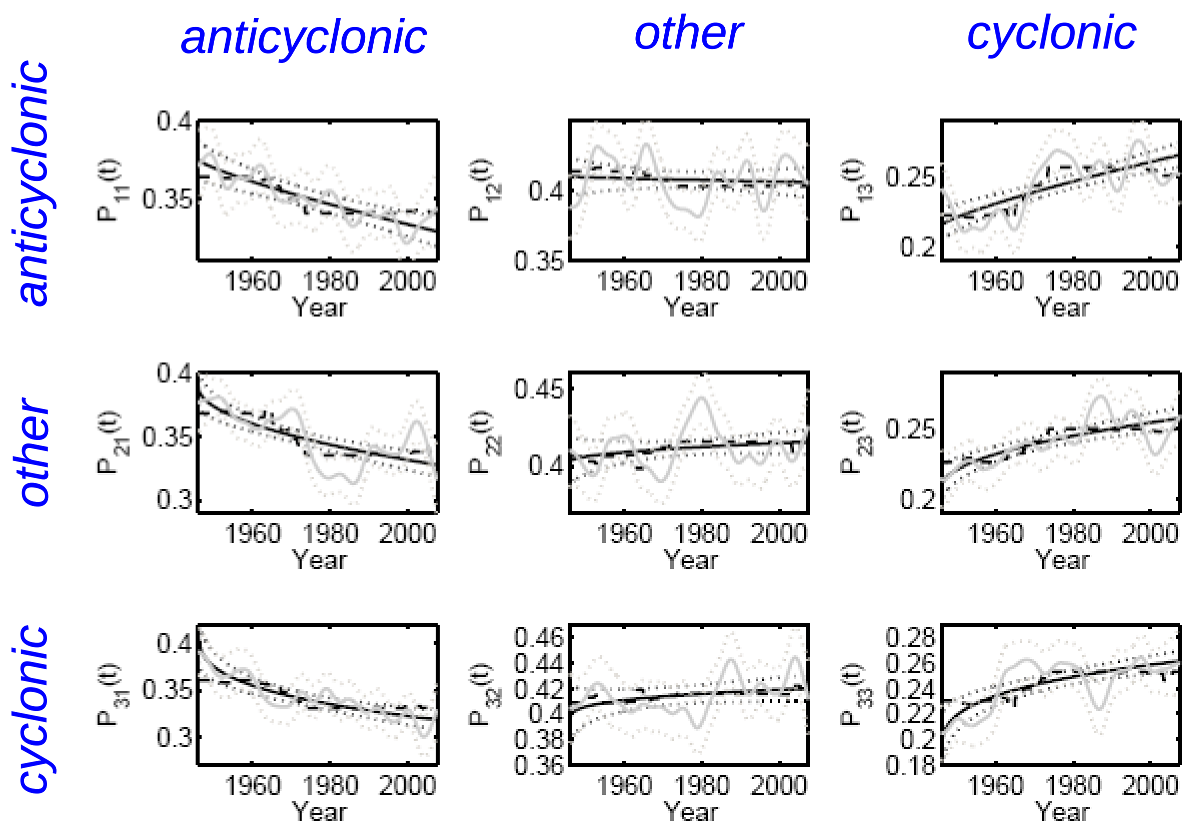
$$\begin{aligned} \mathbf{L}\left(P(t)\right) &= \log \mathrm{P}\left[X_1, \ldots, X_T\right] \\ &= \log \mathrm{P}\left[X_1\right] + \sum_{i}^{m} \sum_{j=1}^{m} \sum_{t \in \{t_{ij}\}} \log P_{ij}(t) \rightarrow \max_{P(t)} \end{aligned}$$

Maximization problem is ill-posed => regularization necessary

Circulation Patterns for UK(1945-2007)

Historical Circulation Data: weather regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

H., submitted to Journal of Atmos. Sci. (2008)



$$\mathbf{L}(\gamma_i(t), \mathbf{T}_i, \mu_i) \longrightarrow \min$$

$$x_t : X_t \rightarrow X_{t+1}$$

$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

FEM: Regularized Clustering Functional

Regularized clustering functional: *(H. 08, to appear in J. of Atmos. Sci.)*

$$\mathbf{L}^{\epsilon}(\Theta, \Gamma(t), \epsilon^2) = \mathbf{L}(\Theta, \Gamma(t)) + \epsilon^2 \sum_{i=1}^{\mathbf{K}} \int_0^T (\partial_t \gamma_i(t))^2 \, dt \rightarrow \min_{\Gamma(t), \Theta},$$

$$\mathbf{L}(\Theta, \Gamma(t)) = \int_0^T \sum_{i=1}^{\mathbf{K}} \gamma_i(t) g(x_t, \theta_i) \rightarrow \min_{\Gamma(t), \Theta},$$

$$x_t : X_t \rightarrow X_{t+1}$$

$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

$$\sum_{i=1}^{\mathbf{K}} \gamma_i(t) = 1, \quad \forall t \in [0, T]$$

$$\gamma_i(t) \geq 0, \quad \forall t \in [0, T], i = 1, \ldots, \mathbf{K}$$

veritas iustitia libertas

FEM: Regularized Clustering Functional

Regularized clustering functional: *(H. 08, to appear in J. of Atmos. Sci.)*

$$\mathbf{L}^{\epsilon}(\Theta,\Gamma(t),\epsilon^2) = \mathbf{L}(\Theta,\Gamma(t)) + \epsilon^2 \sum_{i=1}^{\mathbf{K}} \int_0^T \left(\partial_t \gamma_i(t)\right)^2 dt \longrightarrow \min_{\Gamma(t),\Theta},$$

$$x_t : X_t \rightarrow X_{t+1}$$

$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

$$\tilde{\mathbf{L}}^{\epsilon} = \sum_{i=1}^{\mathbf{K}} \left[a^{\mathbf{T}}(\theta_i)\bar{\gamma}_i + \epsilon^2 \bar{\gamma}_i^{\mathbf{T}} \mathbf{H} \bar{\gamma}_i\right] \rightarrow \min_{\bar{\gamma}_i,\Theta}$$

$$\sum_{i=1}^{\mathbf{K}} \tilde{\gamma}_i^{(k+1)} = 1, \quad \forall k = 1,\ldots,N,$$

$$\tilde{\gamma}_i^{(k+1)} \geq 0, \quad \forall k = 1,\ldots,N; i = 1,\ldots,\mathbf{K}.$$

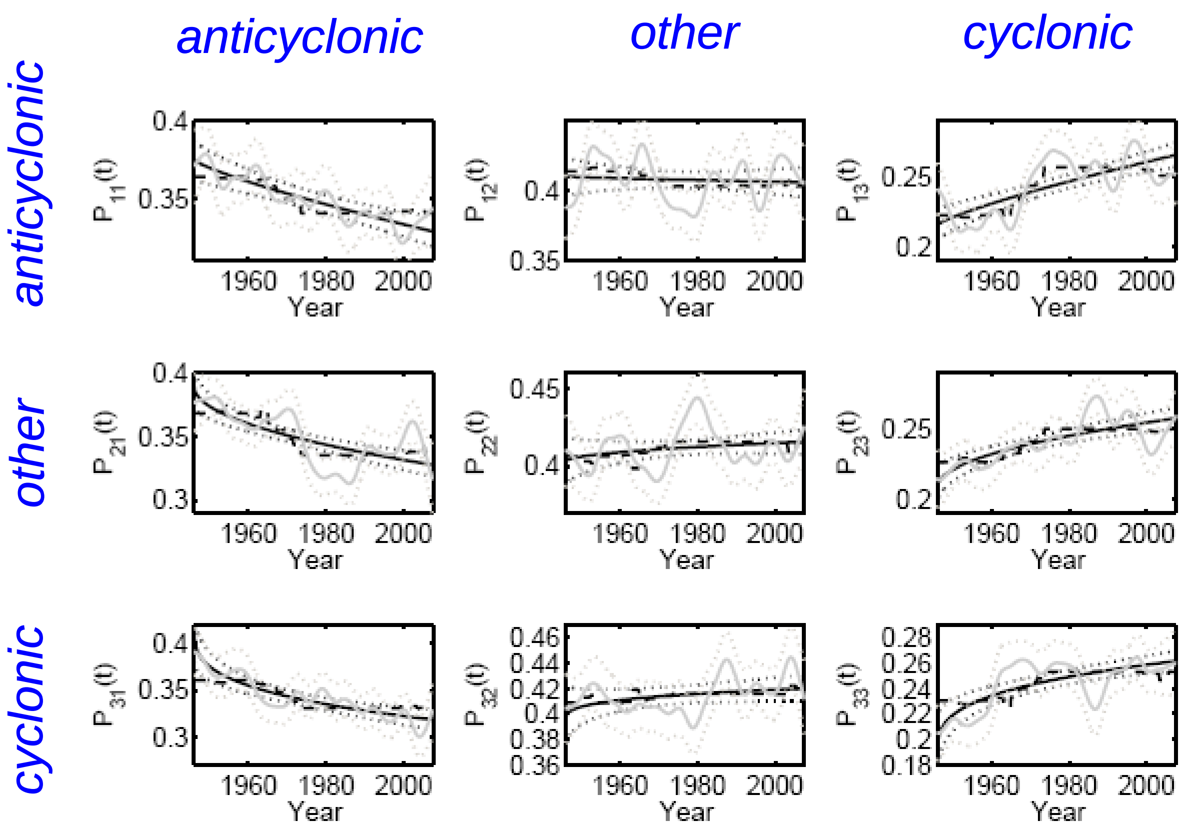
Iterative Subspace Minimization: ***sparse QP*** *can be used*

Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 28 Lamb regimes

(Data from the Univ. of East Anglia)

3 atmospherical states considered



$$\mathbf{L}(\gamma_i(t), \mathbf{T}_i, \mu_i) \rightarrow \min$$

$$x_t : X_t \rightarrow X_{t+1}$$

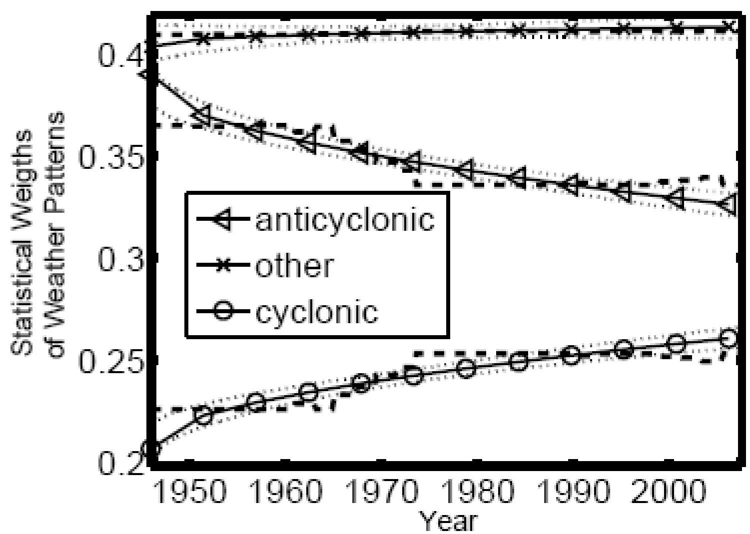
$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

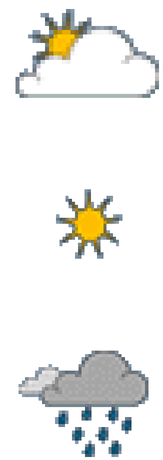
Circulation Patterns for UK(1945-2007)

Historical Circulation Data: weather regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

H., submitted to Journal of Atmos. Sci. (2008)

$$\pi(t)\,P(t) = \pi(t)$$





$$\mathbf{L}(\gamma_i(t), \mathbf{T}_i, \mu_i) \longrightarrow \min$$

$$x_t : X_t \rightarrow X_{t+1}$$

$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

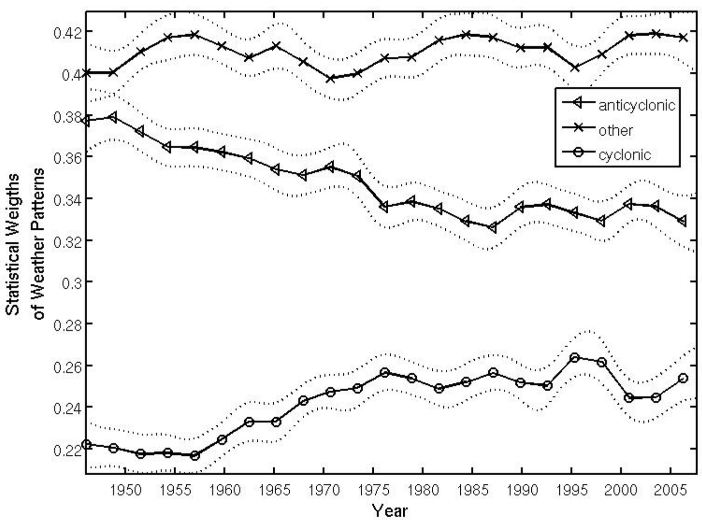
Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 28 Lamb regimes

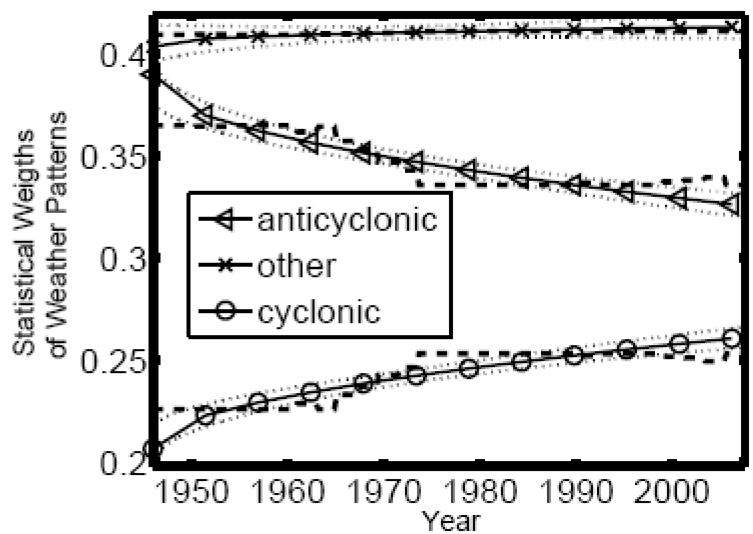
(Data from the Univ. of East Anglia)

3 atmospherical states considered

Gaussian Kernel Estimator



FEM-Clustering vs. Single Trend

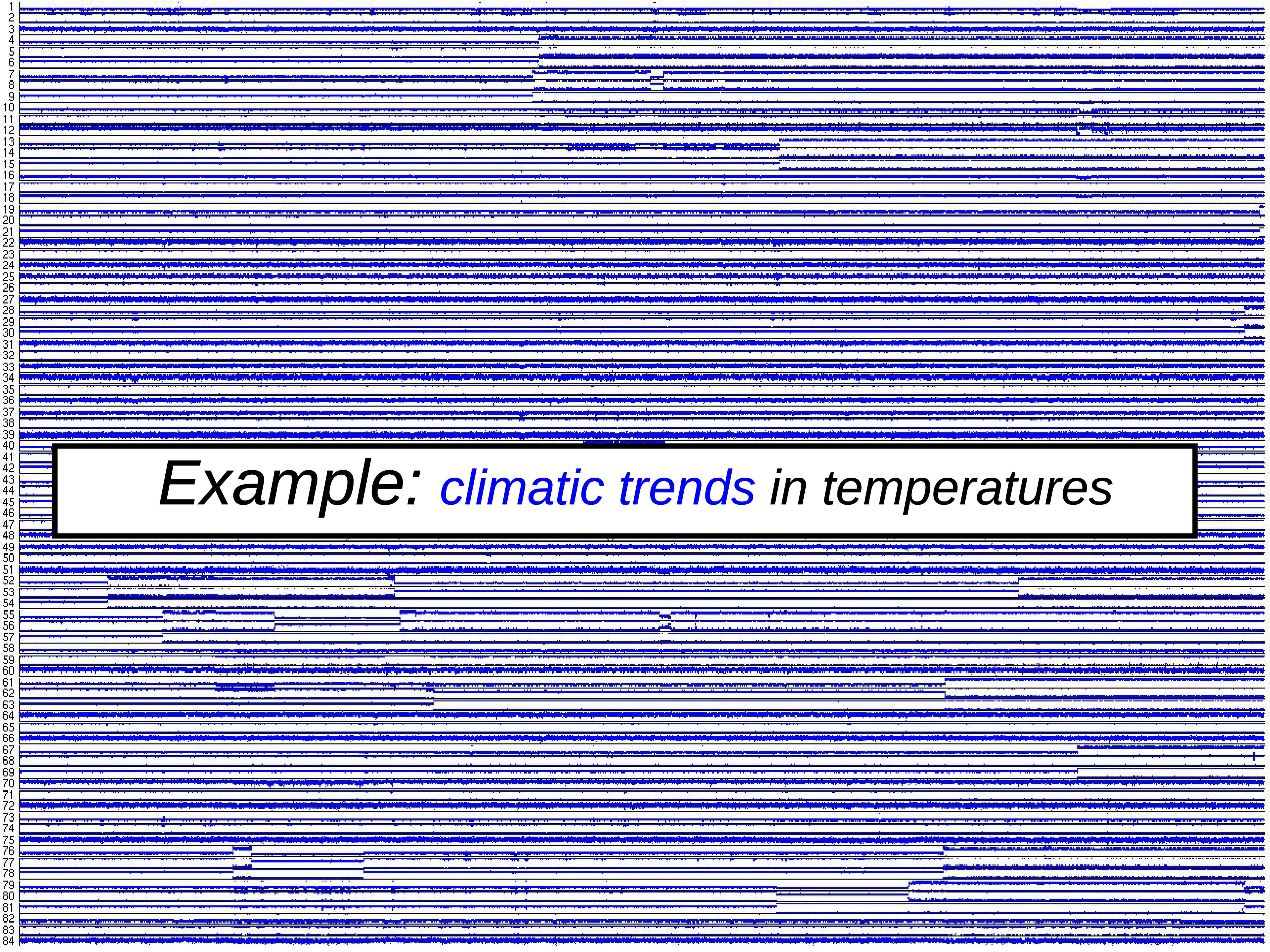


$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

$$x_t : X_t \rightarrow X_{t+1}$$

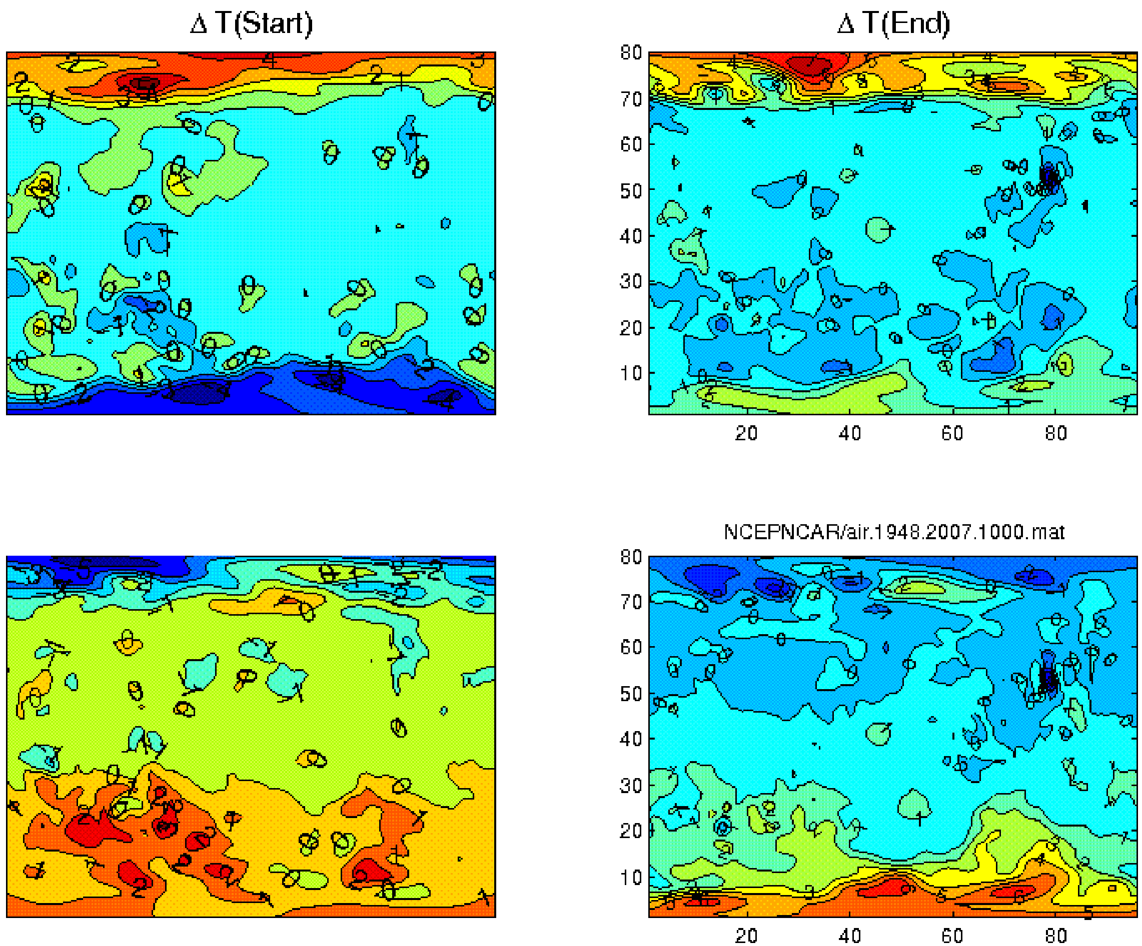
$$\mathbf{L}(\gamma_i(t), \mathbf{\Gamma}_i, \mu_i) \longrightarrow \min$$

Example: climatic trends in temperatures



Global Data Analysis

Global Historical Temperature Data
(80x120 grid, daily values 1947-2007)
(Data from the NCAR, Boulder, US)



Global Data Analysis

Global Historical Temperature Data
(80x120 grid, daily values 1947-2007)
(Data from the NCAR, Boulder, US)