Analysis of non-stationary Data

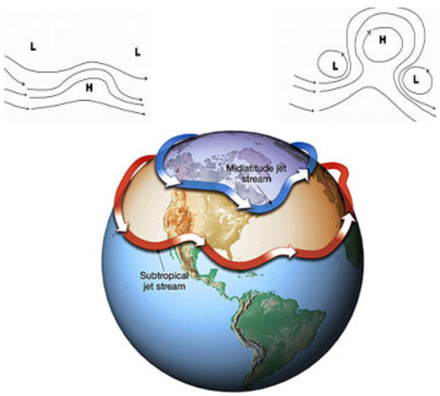
Illia Horenko

Scientific Computing
Institute of Mathematics
Freie Universität Berlin (FU)

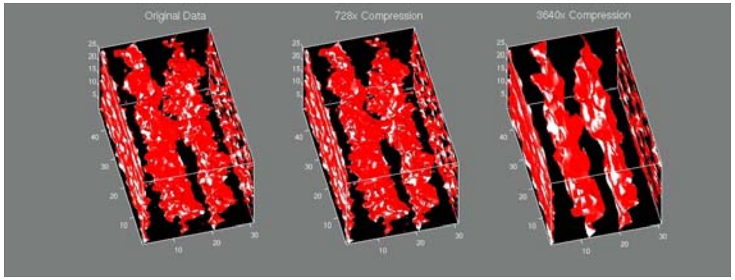
DFG Research Center MATHEON
„Mathematics in key technologies“

- *Motivation*
  - *global stationarity*, *local stationarity*, examples
- *Non-Stationarity for continuous state space processes*:
  - *Fuzzy Clustering with Regression Models (FCRM)*
  - *Finite Element Clustering* for continuos state space processes
- *Non-Stationarity for discrete state space processes:*
  - *Kernel Filtering Methods*
  - *Single Trend Model*
  - *FEM-Clustering* of Markov-chain output
  - Example I: analysis of historical *weather patterns*
  - Example II: analysis of the *historical temperatures* (1947-2007)

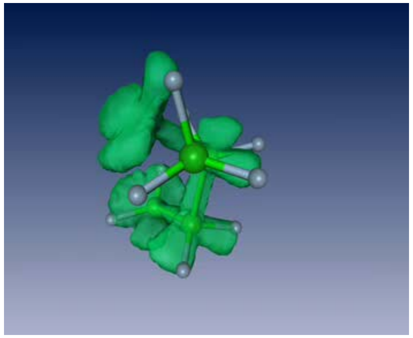
Non-Stationarity



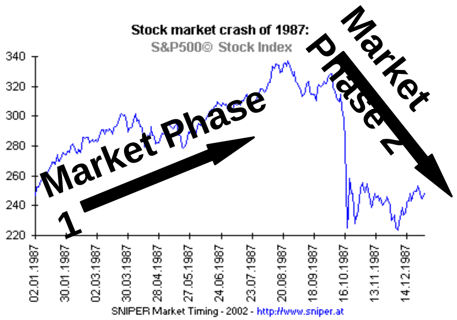
Meteorology/Climate



Fluid Mechanics



Biophysics/Drug Design

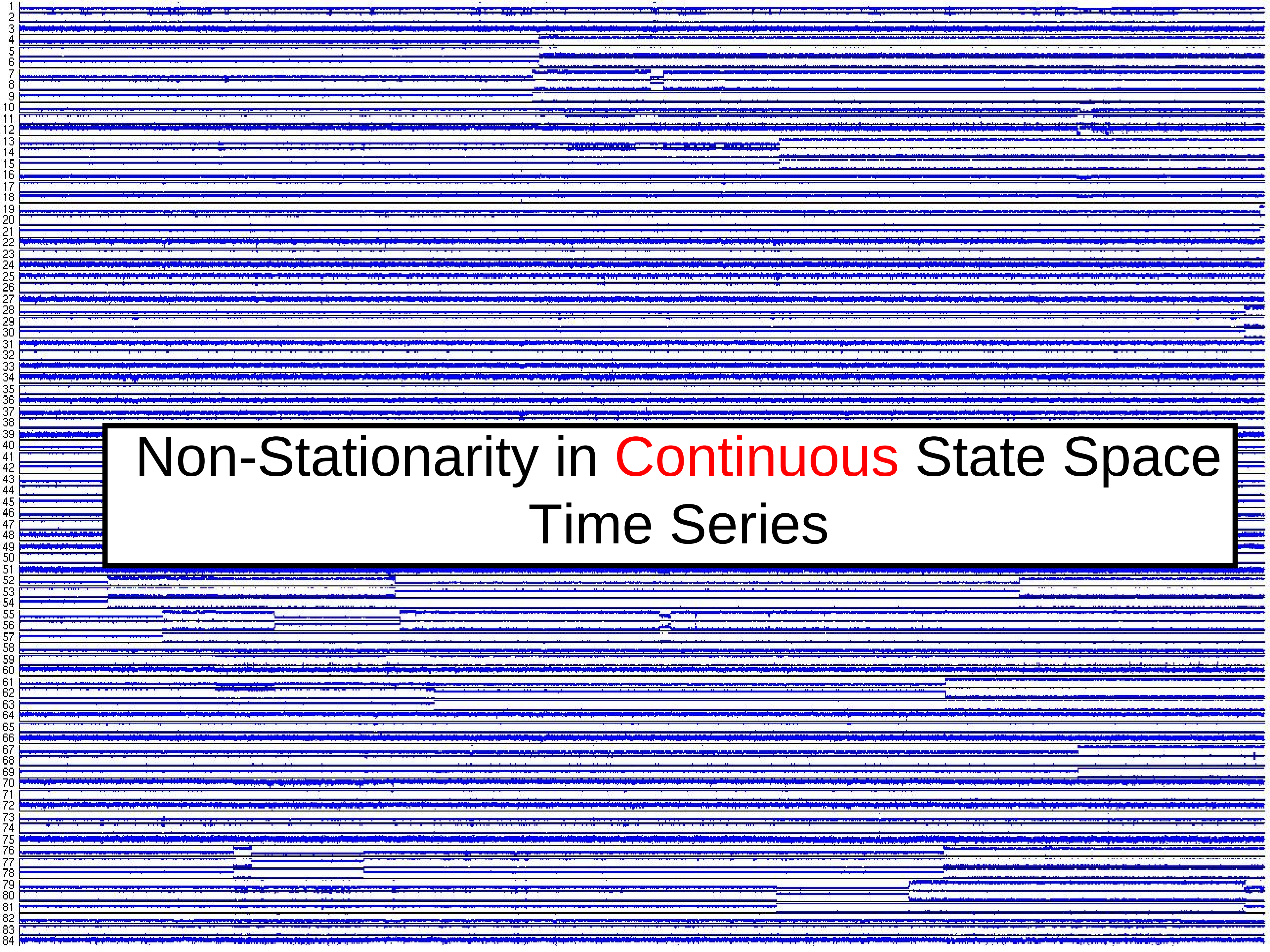


Computational Finance

(Local) weak stationarity: (local) time independence of mean values and covariances

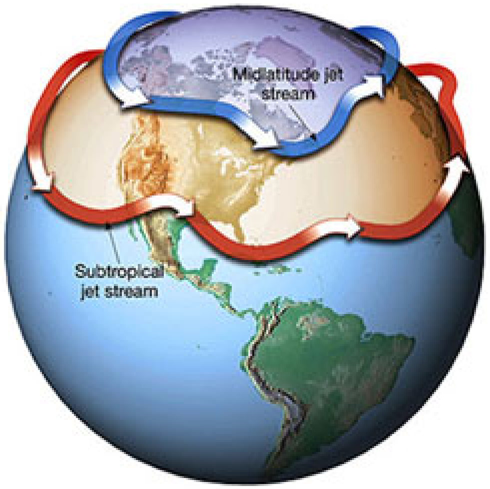
Not fulfilled in many cases

Non-Stationarity in Continuous State Space Time Series

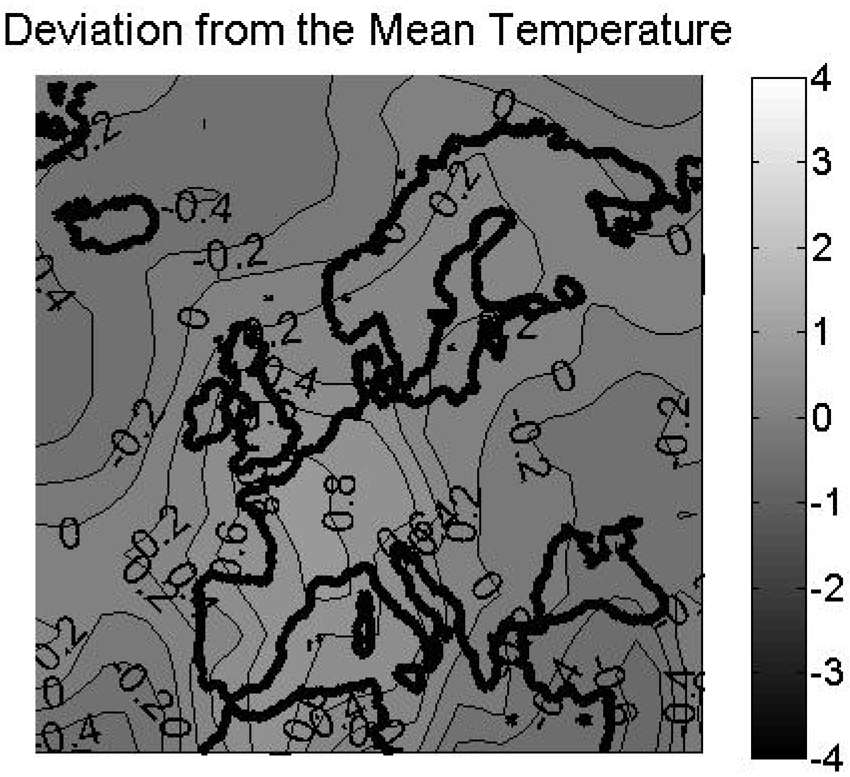


Example: weather data analysis

Global Historical Data



Local Historical Data



Non-stationarity because of the climate change => standard data-analysis methods are ***not applicable***

Memo I: K-Means clustering

Geometrical distance: $\theta_i \in \Psi$ - *time-independent* cluster centers

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

$$t_j, j = 1, \ldots, n \in [0, T]$$

$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i(t_j) \| x_{t_j} - \theta_i \|^2 \rightarrow \min_{\Gamma(t), \Theta}$$

(Bezdek1981, Höppner et.al. 1999)

Iteration number (l):

$$\gamma_i^{(l)}(t_j) = \begin{cases} 1 & i = \arg\min \| x_{t_j} - \theta_i^{(l-1)} \|^2, \\ 0 & \text{otherwise}, \end{cases}$$

$$\theta_i^{(l)} = \frac{\sum_{j=1}^{n} \gamma_i^{(l)}(t_j) x_{t_j}}{\sum_{j=1}^{n} \gamma_i^{(l)}(t_j)}.$$

Assumption: time-independence of cluster centers ➡ *local stationarity*

Non-stationary Extension of K-Means: FCRM

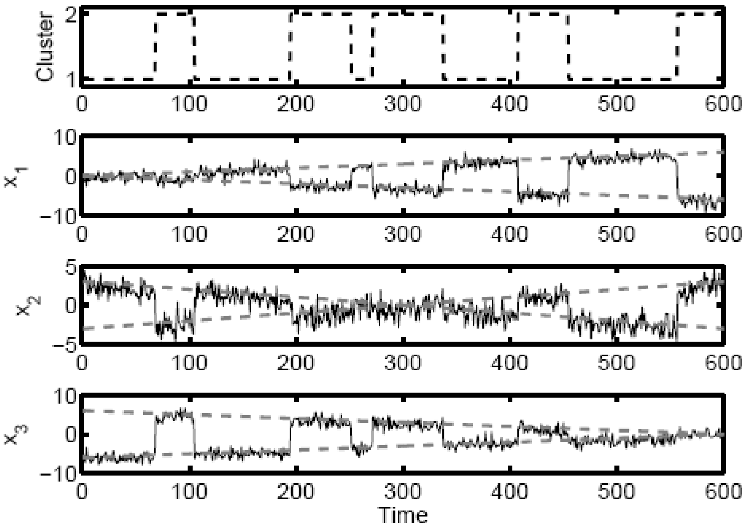
Geometrical distance: *time-dependent* cluster centers as linear combinations of *basis functions* $\phi_k(t), k = 0, \ldots, \mathcal{R}$

$$\theta_i(t) = \sum_{k=0}^{\mathcal{R}} \theta_{ik}\phi_k(t_j) \qquad \text{(Hathaway and Bezdek1993)}$$

$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i^m(t_j) \parallel x_{t_j} - \sum_{k=0}^{\mathcal{R}} \theta_{ik}\phi_k(t_j) \parallel^2 \rightarrow \min_{\Gamma(t),\Theta}$$

Toy Example:

$$\begin{aligned} x_j(t) &= \theta_{i(t)j}(t - \bar{t}_j) + \sigma \mathbf{N}(0,1), \quad i = 1,2, \quad j = 1,2,3 \\ \theta_1 &= \begin{pmatrix} 0.01 & -0.01 & 0.01 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} -0.01 & 0.01 & -0.01 \end{pmatrix} \\ \bar{t} &= \begin{pmatrix} 0 & 300 & 600 \end{pmatrix} \end{aligned}$$



Non-stationary Extension of K-Means: FCRM

Geometrical distance: *time-dependent* cluster centers as linear combinations of *basis functions* $\phi_k(t), k = 0, \ldots, \mathcal{R}$

$$\theta_i(t) = \sum_{k=0}^{\mathcal{R}} \theta_{ik} \phi_k(t_j)$$

(Hathaway and Bezdek1993)

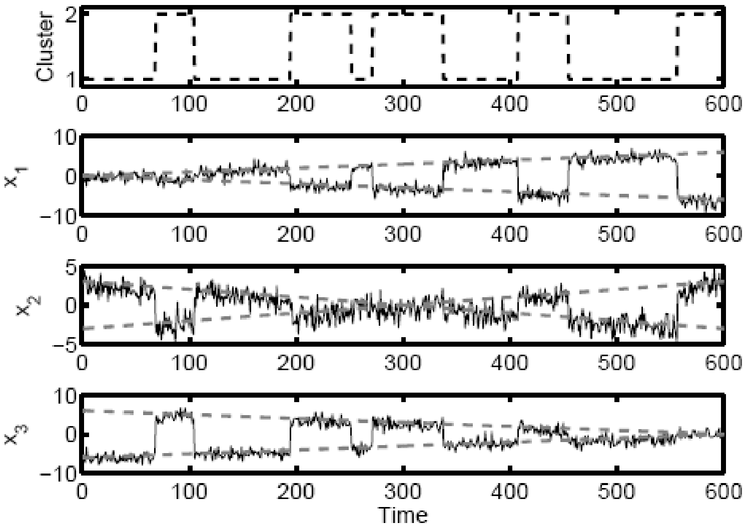
$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i^m(t_j) \parallel x_{t_j} - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \phi_k(t_j) \parallel^2 \ \rightarrow \min_{\Gamma(t),\Theta}$$

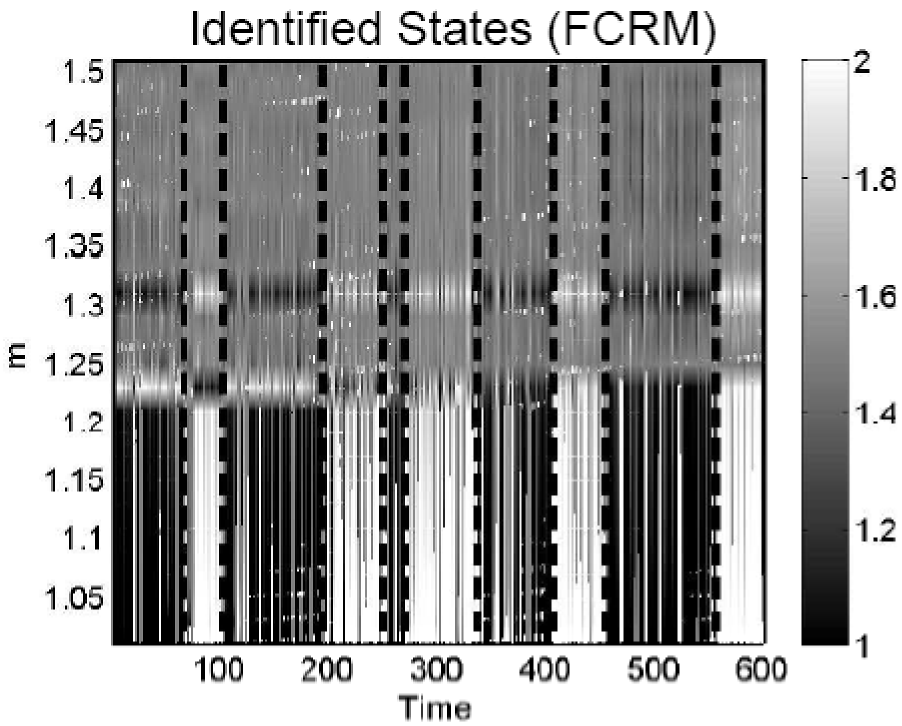
Exercise 1: *Derive an expression for the optimal estimator of the regression parameters θ (x and γ are fixed). What happens to this estimate if R is growing? How to define the optimal R? Suggest and discuss the possible numerical solutions.*

Non-stationary Extension of K-Means: FCRM

Toy Example:

$$
\begin{aligned}
x_j(t) &= \theta_{i(t)j}(t - \bar{t}_j) + \sigma \mathbf{N}(0,1), \quad i = 1,2, \quad j = 1,2,3 \\
\theta_1 &= \begin{pmatrix} 0.01 & -0.01 & 0.01 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} -0.01 & 0.01 & -0.01 \end{pmatrix} \\
\bar{t} &= \begin{pmatrix} 0 & 300 & 600 \end{pmatrix}
\end{aligned}
$$





Proper identification of persistent clusters impossible (try the FEM-Clustering ?)

FEM: Regularized Clustering Functional

Regularized clustering functional:

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

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

$$g\left(x_t, \theta_i\right) = \| \, x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k\left(t\right) \|^2$$

FEM: Regularized Clustering Functional

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

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

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

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

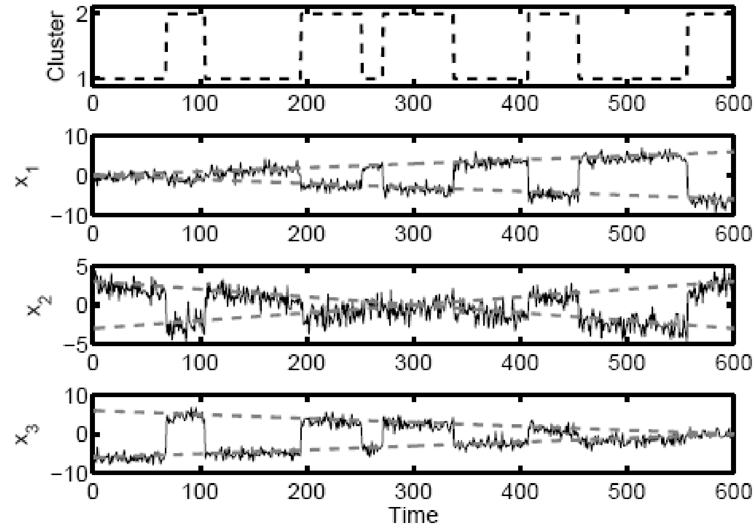
$$a(\theta_i) = \left(\int_{t_1}^{t_2} v_1(t) g(x_t, \theta_i) dt, \ldots, \int_{t_{N-1}}^{t_N} v_N(t) g(x_t, \theta_i) dt \right)$$

$$g(x_t, \theta_i) = \| x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k(t) \|^2$$

FEM: Regularized Clustering Functional

Toy Example:

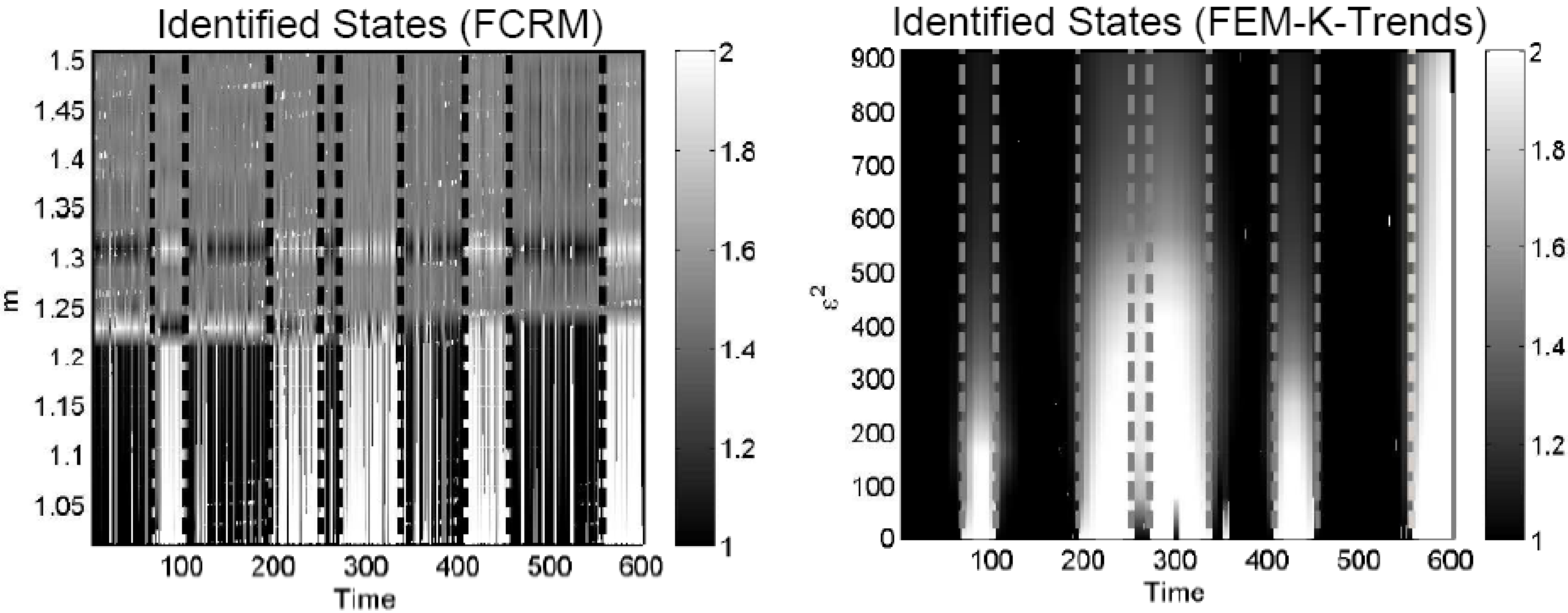
$$\begin{aligned} x_j(t) &= \theta_{i(t)j}(t-\bar{t}_j) + \sigma \mathbf{N}(0,1), \quad i=1,2, \quad j=1,2,3 \\ \theta_1 &= \begin{pmatrix} 0.01 & -0.01 & 0.01 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} -0.01 & 0.01 & -0.01 \end{pmatrix} \\ \bar{t} &= \begin{pmatrix} 0 & 300 & 600 \end{pmatrix} \end{aligned}$$



FEM: Regularized Clustering Functional

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

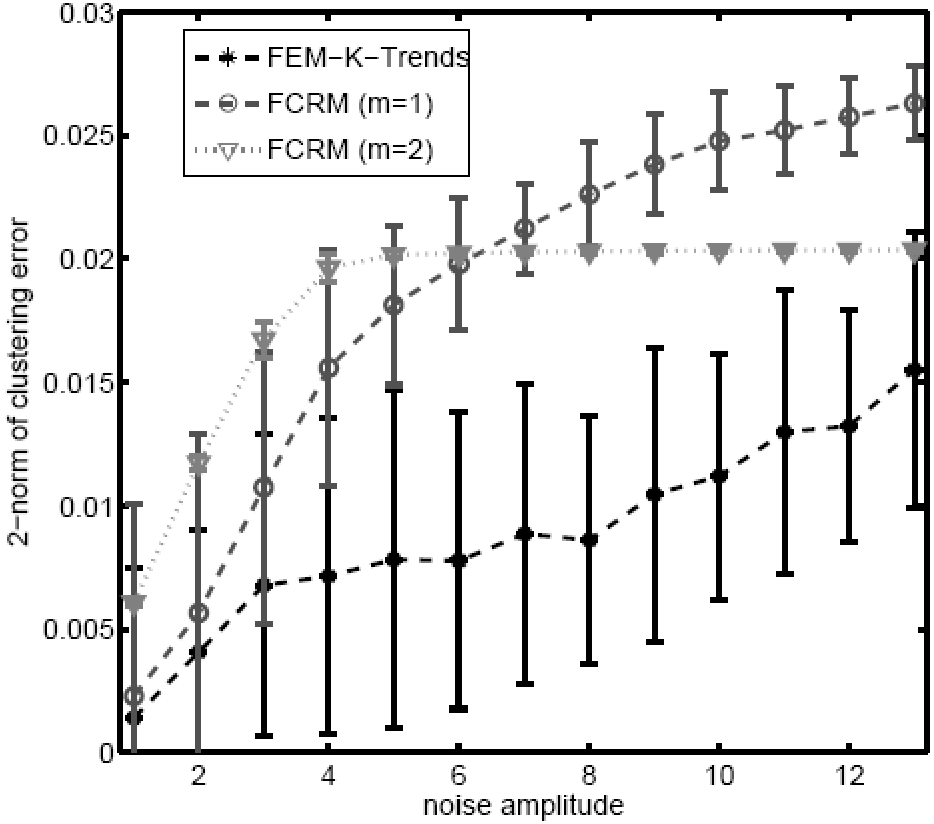
$$g(x_t, \theta_i) = \| x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k(t) \|^2$$



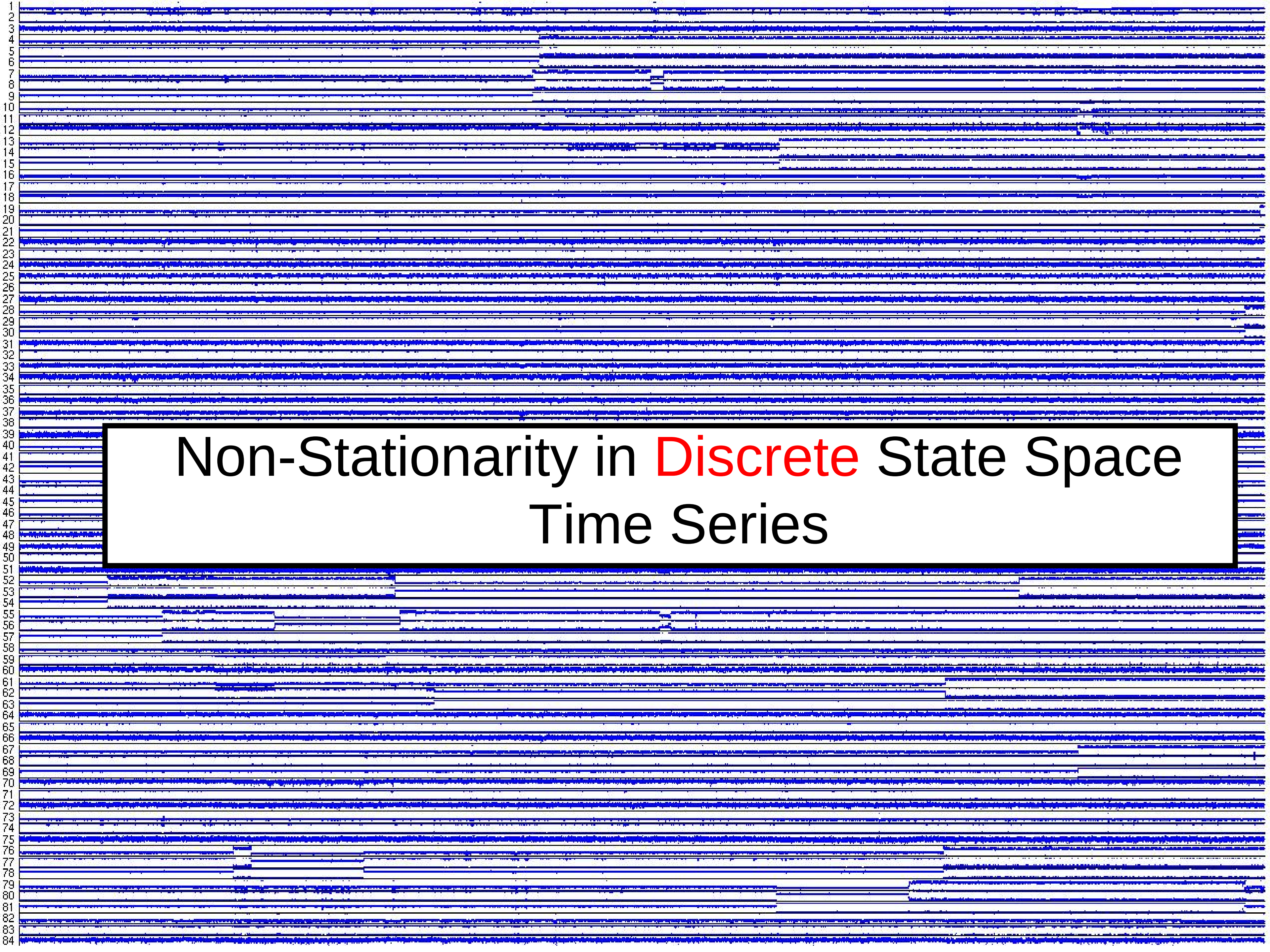
FEM: Regularized Clustering Functional

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

$$g(x_t,\theta_i) \;=\; \| \, x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k(t) \, \|^2$$

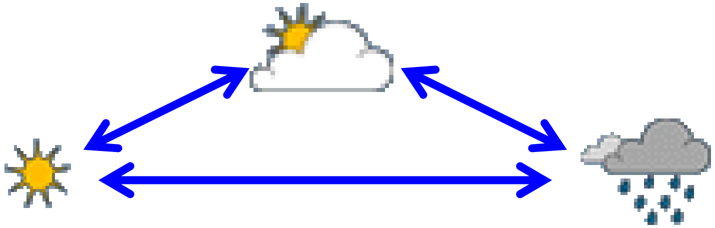


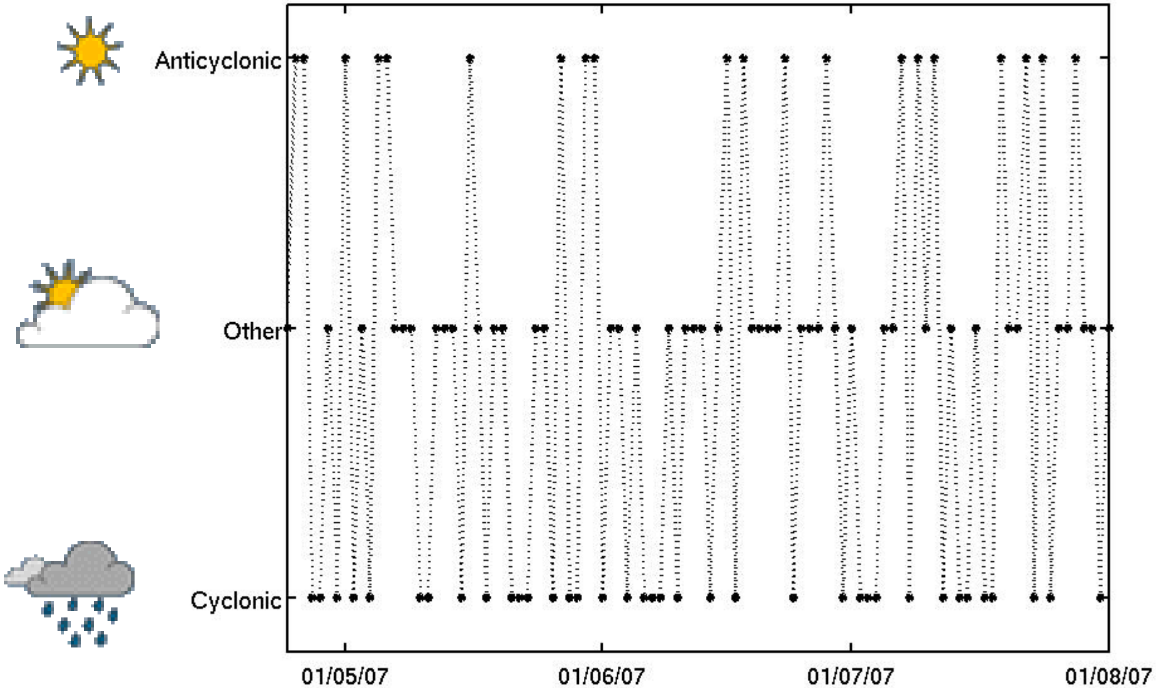
Non-Stationarity in Discrete State Space Time Series



Example:Weather in UK(1945-2007)

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





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: 27 Lamb regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

Regularization: global stationarity assumption

	anticyclonic	other	cyclonic
anticyclonic	0.34	0.40	0.26
other	0.35	0.41	0.24
cyclonic	0.34	0.41	0.25

Locally Stationary Markov Chain

Regularization: local stationarity assumption inside of the window

Gaussian Window: $\gamma(t,t_0) = \frac{1}{c}\exp(-\frac{(t-t_0)^2}{\sigma^2})$

Approximate Log-Likelihood:

$$\mathbf{L}\left(P(t_0)\right) \approx \log \mathrm{P}\left[X_1\right] + \sum_{i,j=1}^{m} \sum_{t\in\{t_{ij}\}} \gamma(t,t_0)\log P_{ij}(t_0)$$

Gaussian Kernel Filtering:

$$P_{ij}(t_0) = \frac{\sum_{t\in\{t_{ij}\}}\gamma(t,t_0)}{\sum_{t\in\{t_i\}}\gamma(t,t_0)}$$

Standard statistical methods applicable to calculate the conf. intervals

Locally Stationary Markov Chain

Regularization: local stationarity assumption inside of the window

Gaussian Window: $\gamma(t, t_0) = \frac{1}{c}\exp(-\frac{(t-t_0)^2}{\sigma^2})$

Approximate Log-Likelihood:

$$\mathbf{L}\left(P(t_0)\right) \approx \log \mathrm{P}\left[X_1\right] + \sum_{i,j=1}^{m} \sum_{t \in \{t_{ij}\}} \gamma(t, t_0) \log P_{ij}(t_0)$$

Gaussian Kernel Filtering:

$$P_{ij}(t_0) = \frac{\sum_{t \in \{t_{ij}\}} \gamma(t, t_0)}{\sum_{t \in \{t_i\}} \gamma(t, t_0)}$$

Exercise 2*: proof that this formula is true*

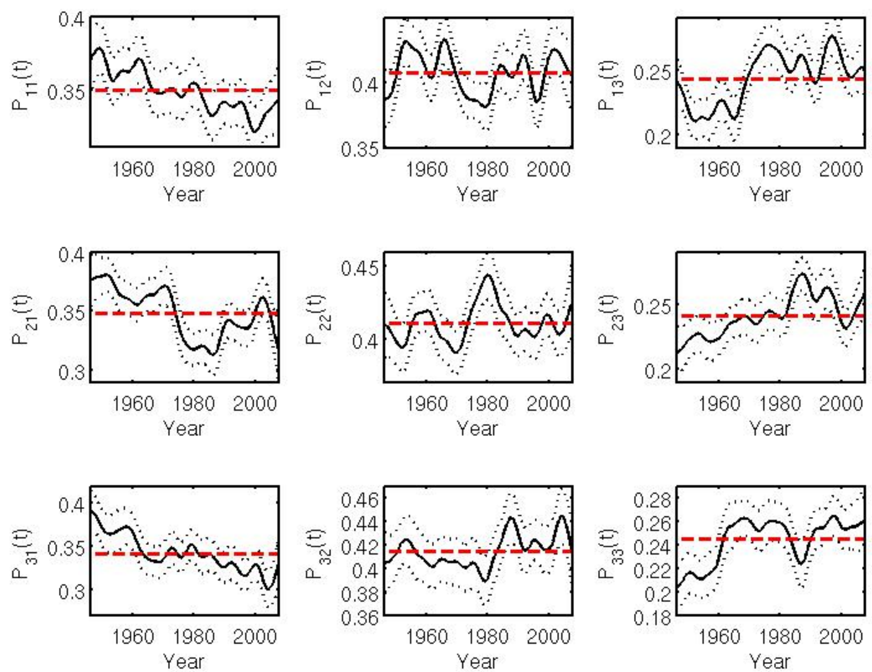
Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 27 Lamb regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

Gaussian Window vs. Homogenous Estimator

anticyclonic *other* *cyclonic*

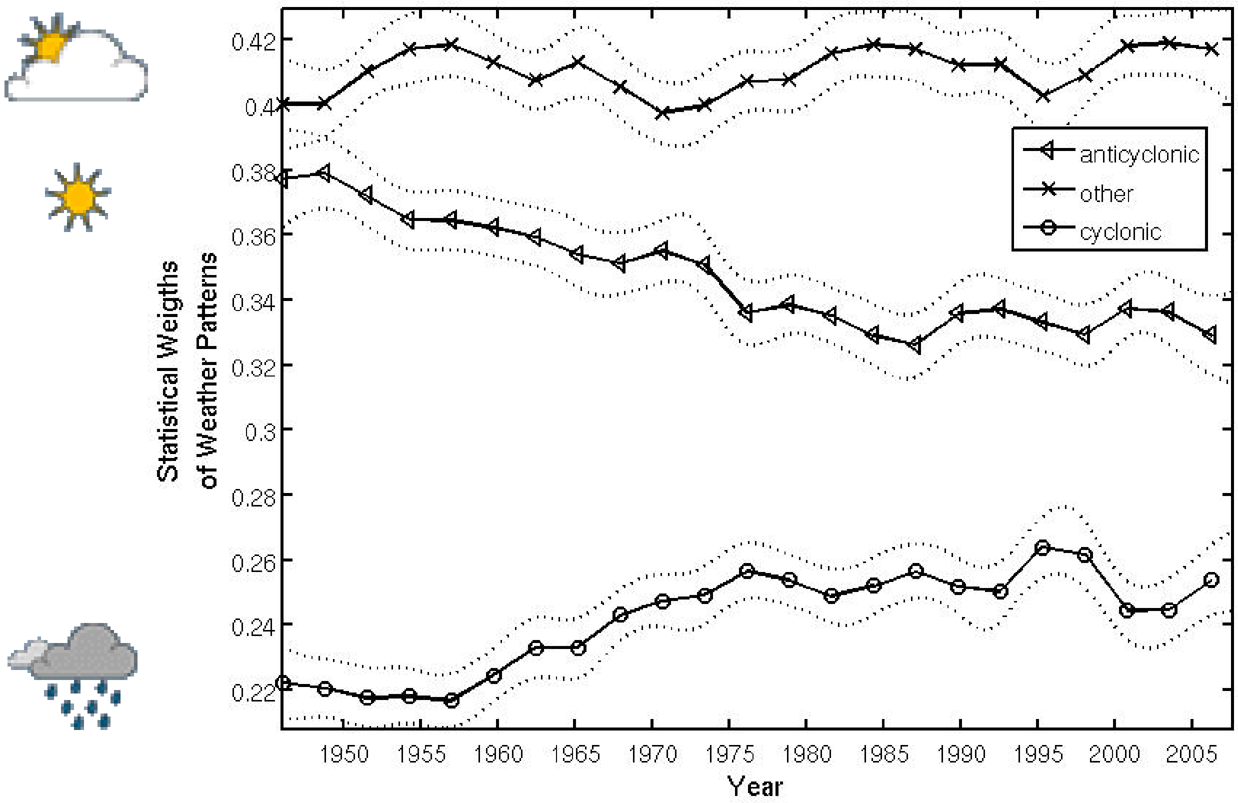
anticyclonic
other
cyclonic



Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 27 Lamb regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

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



Parametric Estimation

(H. 08, to appear in J. of Atmos. Sci.)

Markovian Trend Model:

$$P(t) \;=\; P^{(0)} + P^{(1)}\phi\,(t)\,, \quad \phi : [1,T] \rightarrow (-\infty,+\infty)$$

Log-Likelihood:

$$\sum_{j=1}^{m} \sum_{t\in\{t_{ij}\}} \log\left(P_{ij}^{(0)} + P_{ij}^{(1)}\phi\,(t)\right) \;\rightarrow\; \max_{P^{(0)},P^{(1)}},$$

$$\sum_{j=1}^{m} P_{ij}^{(0)} \;=\; 1,$$

$$\sum_{j=1}^{m} P_{ij}^{(1)} \;=\; 0,$$

$$P_{ij}^{(0)} + P_{ij}^{(1)} \sup_{t\in[1,T]} \phi(t) \;\geq\; 0, \quad \text{for all } j,$$

$$P_{ij}^{(0)} + P_{ij}^{(1)} \inf_{t\in[1,T]} \phi(t) \;\geq\; 0, \quad \text{for all } j.$$

Numerics: Nelder-Mead Optimization Algorithm

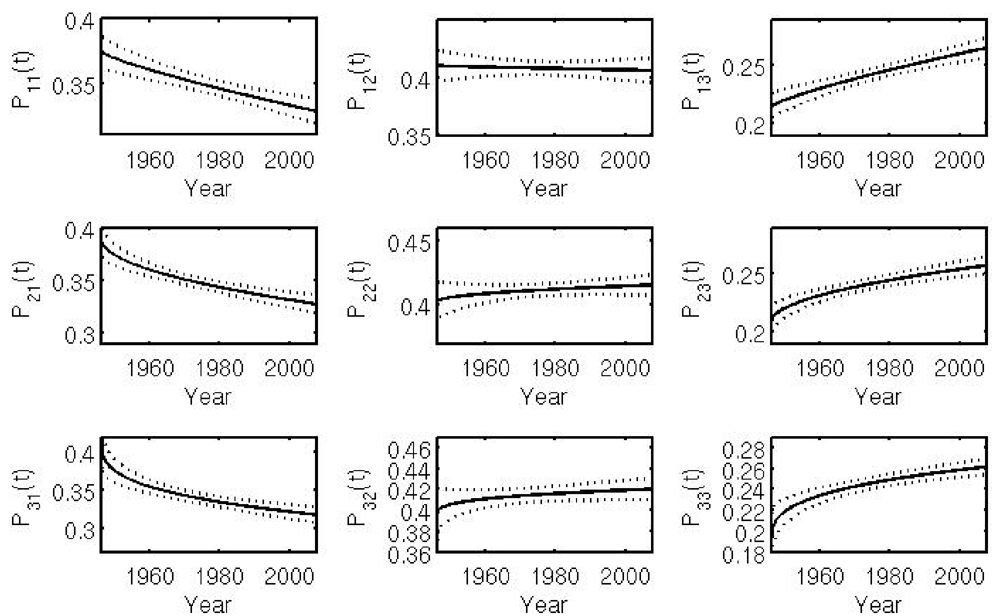
Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 27 Lamb regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

Polynomial Trend Model

anticyclonic *other* *cyclonic*

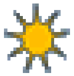
anticyclonic
other
cyclonic

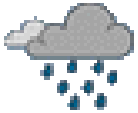


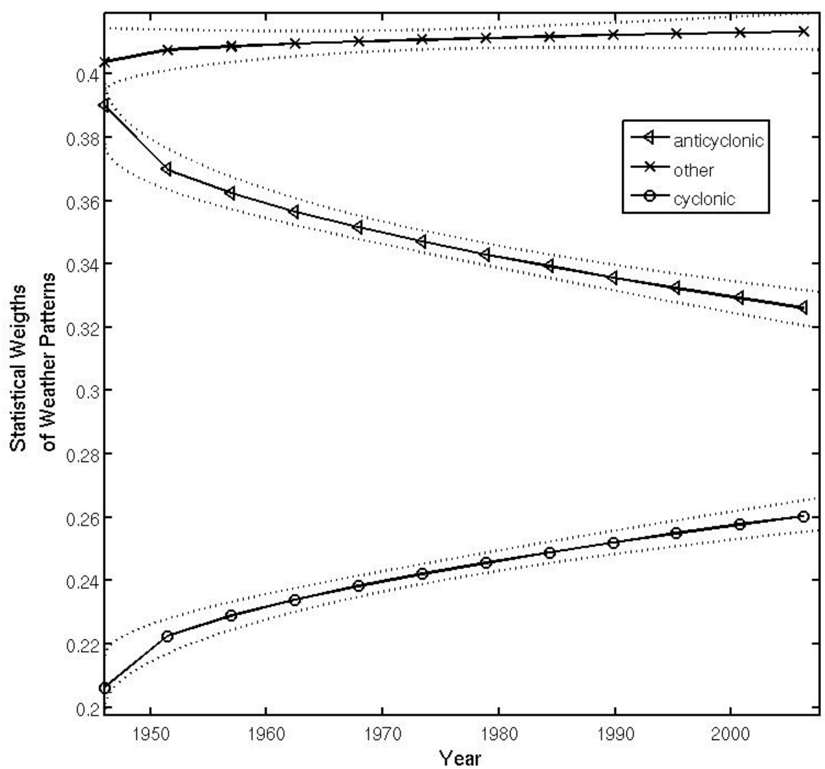
Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 27 Lamb regimes
(Data from the Univ. of East Anglia)
3 atmospherical states considered

Polynomial Trend Model







FEM: Regularized Clustering Functional

Regularized clustering functional:

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

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

FEM: Regularized Clustering Functional

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

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

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

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

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

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

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

veritas iustitia libertas

FEM: Regularized Clustering Functional

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

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

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

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

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

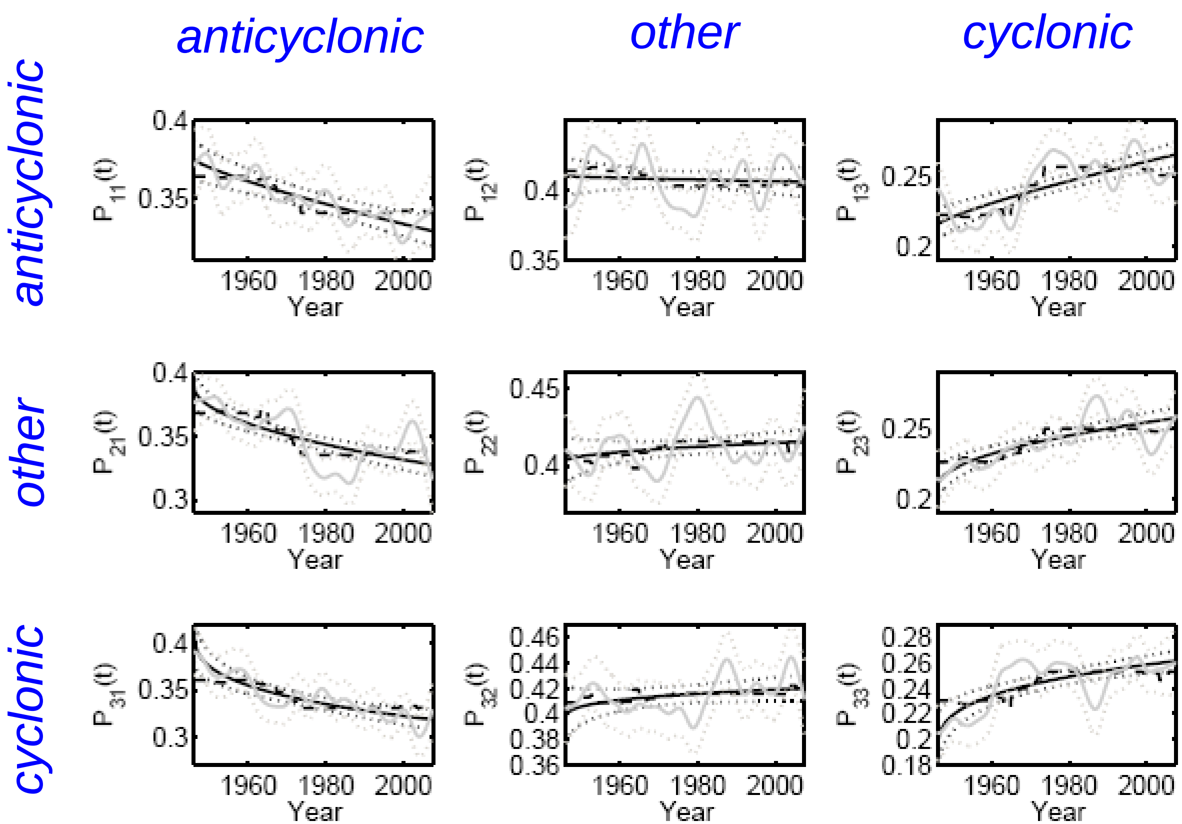
$$\sum_{i=1}^{\mathbf{K}} \tilde{\gamma}_i^{(k+1)} \quad = \quad 1, \quad \forall k = 1, \ldots, N,$$

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

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



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

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

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

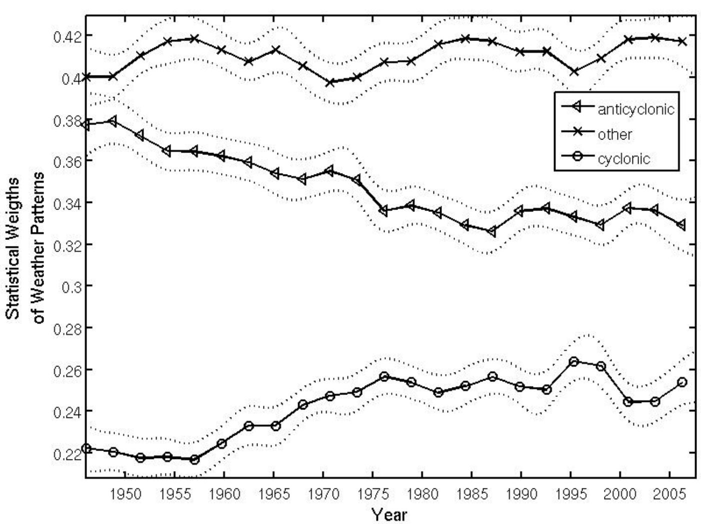
Circulation Patterns for UK(1945-2007)

Historical Circulation Data: 28 Lamb regimes

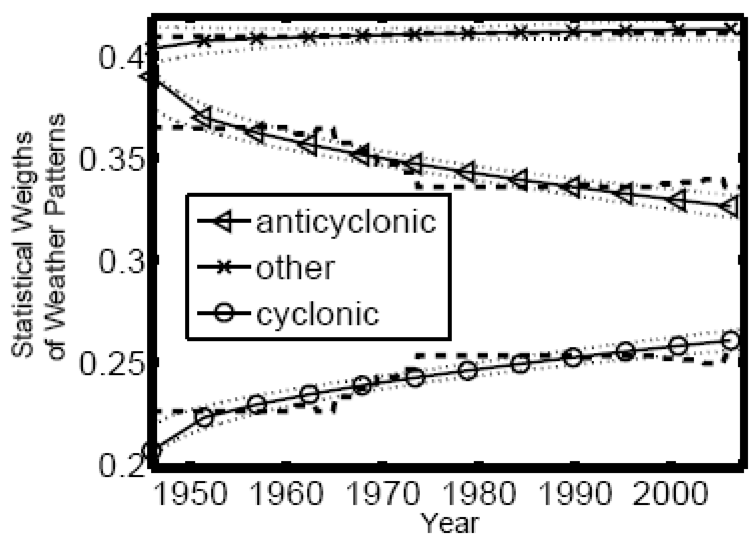
(Data from the Univ. of East Anglia)

3 atmospherical states considered

Gaussian Kernel Estimator



FEM-Clustering vs. Single Trend



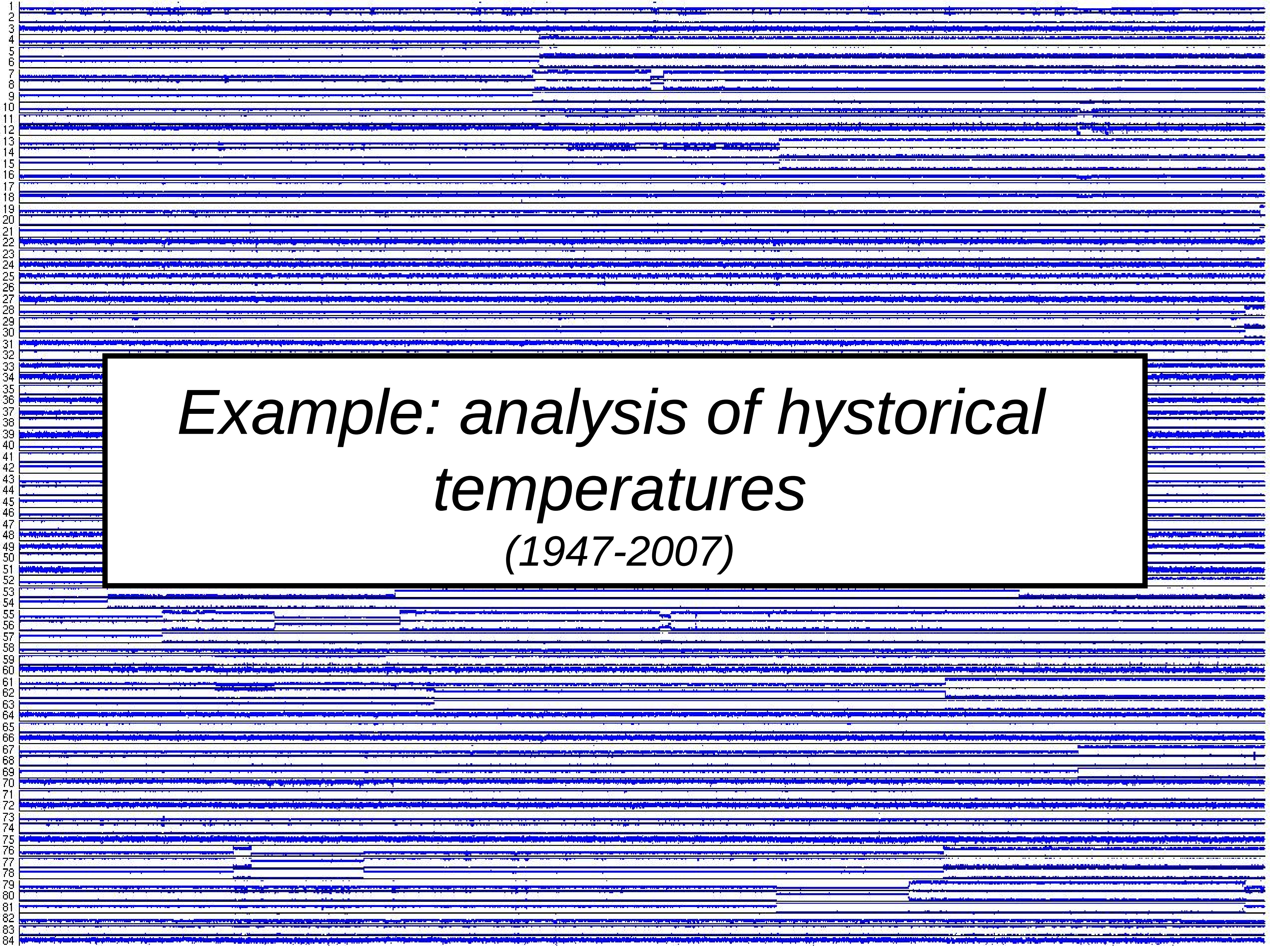
$$g(x_t, P^i) = -\log P^i_{X_t X_{t+1}}$$

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

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

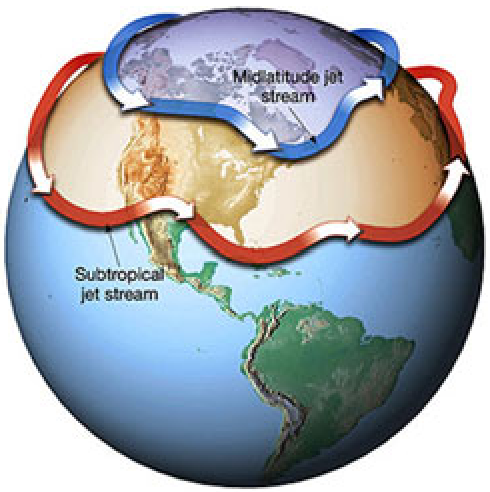
Example: analysis of hystorical temperatures

(1947-2007)



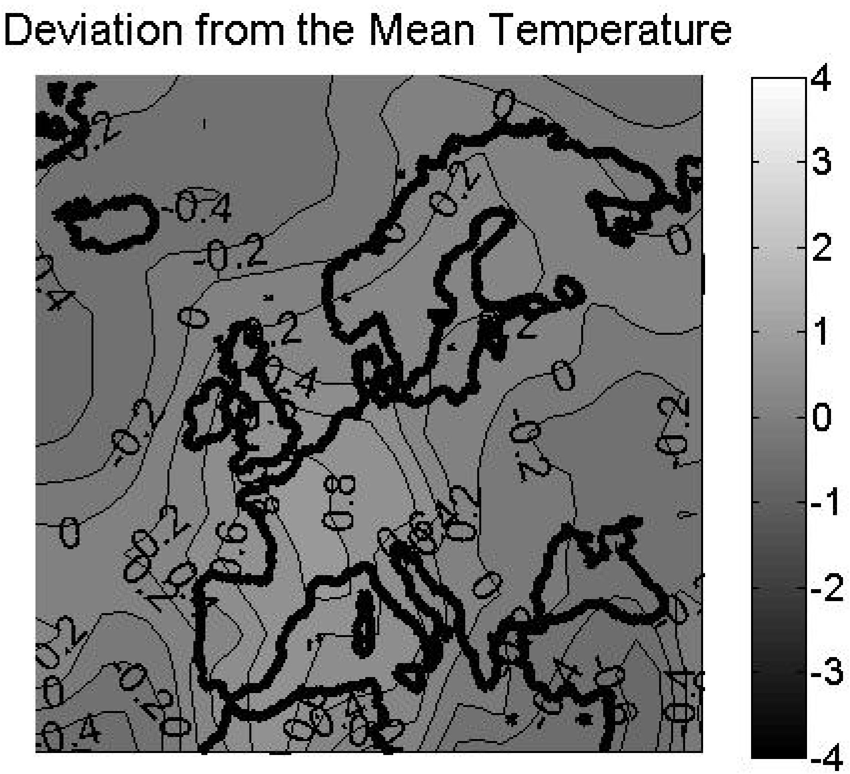
Temperature Data Analysis

Global Historical Temperatures



Data from Boulder Center, USA

Historical Temperatures in Europe



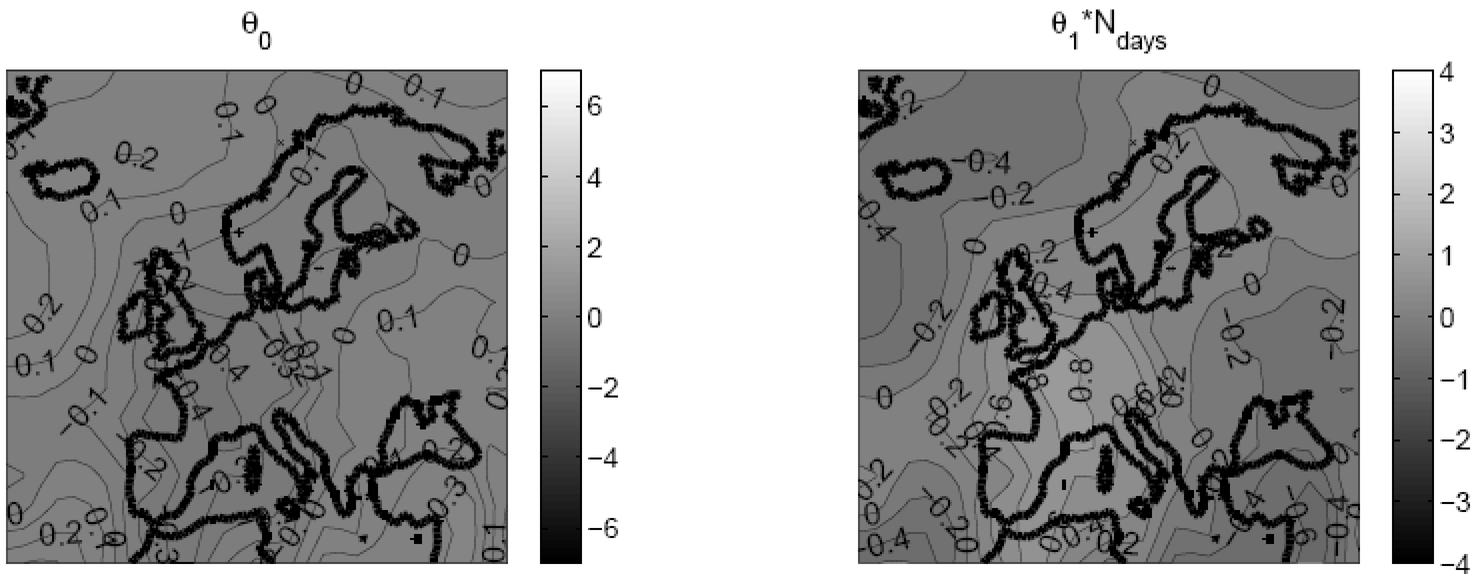
Data from Ch. Franzke, BAS Cambridge

Temperature Data on the 2D-Grid (1947-2007)

European Data Analysis

Temperature Data in Europe: 29x20 grid

Seasonal Cycle Eliminated



$$\mathbf{L}\left(\gamma_i(t), \mathbf{T}_i, \mu_i\right) \longrightarrow \min$$: K=1

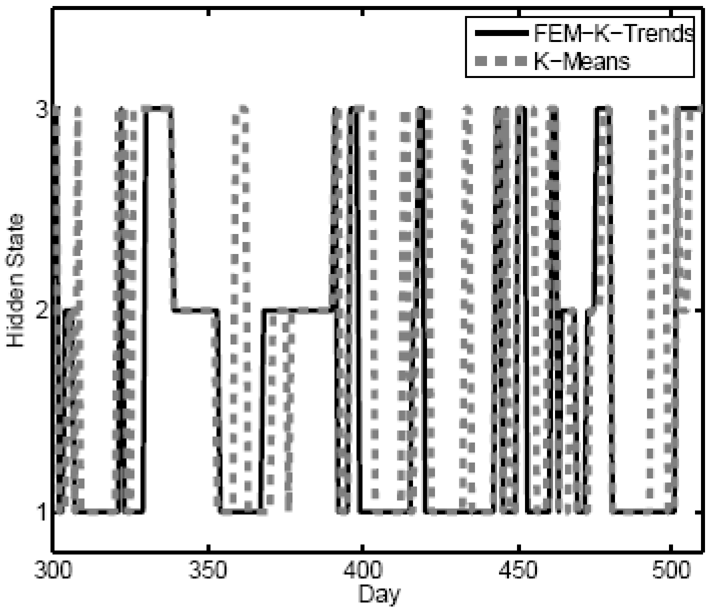
$$g\left(x_t, \theta_i\right) \;=\; \| \, x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k\left(t\right) \|^2$$

Linear Regression

European Data Analysis

Temperature Data in Europe: 29x20 grid

Seasonal Cycle Eliminated

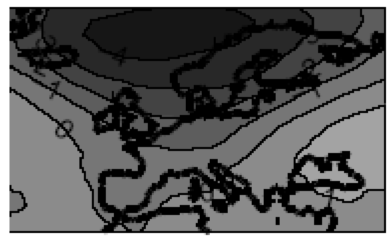


European Data Analysis

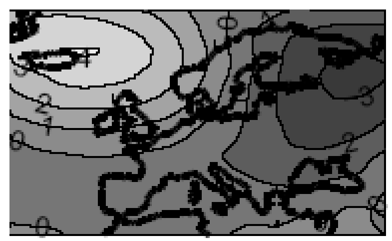
Temperature Data in Europe: 29x20 grid

Δ T(01–Jan–1947)

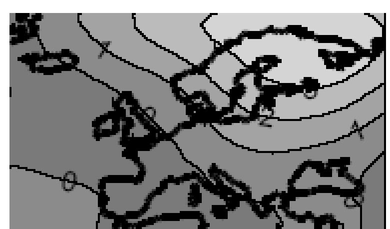
[***Cluster 1***]



[***Cluster 2***]



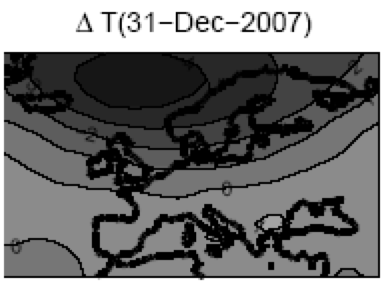
[***Cluster 3***]



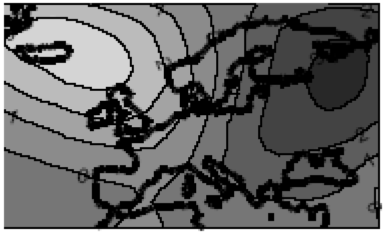
European Data Analysis

Temperature Data in Europe: 29x20 grid

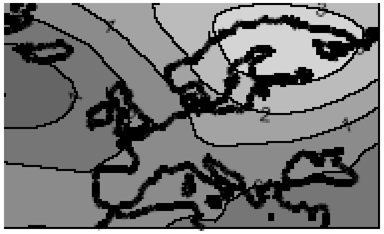
[***Cluster 1***]



[***Cluster 2***]



[***Cluster 3***]

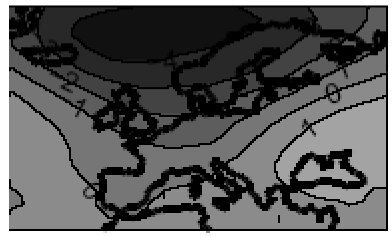


European Data Analysis

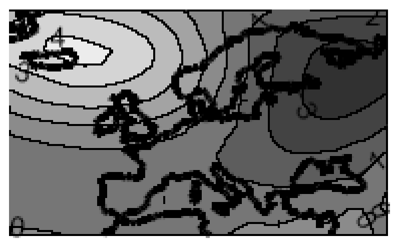
Temperature Data in Europe: 29x20 grid

Δ T(01−Jan−1947)

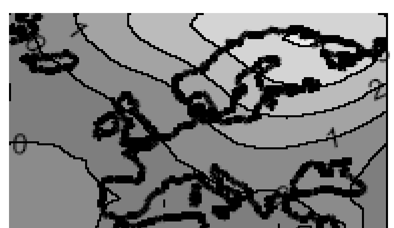
[Cluster 1]()



[Cluster 2]()



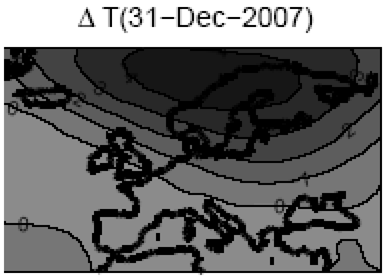
[Cluster 3]()



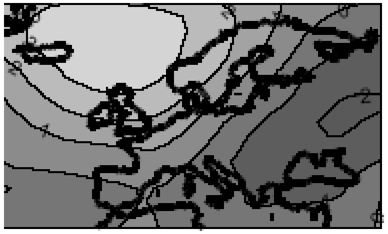
European Data Analysis

Temperature Data in Europe: 29x20 grid

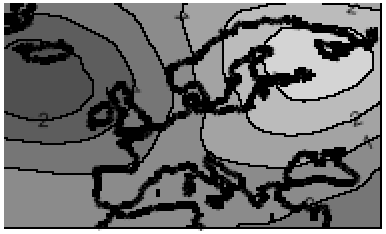
[Cluster 1]



[Cluster 2]

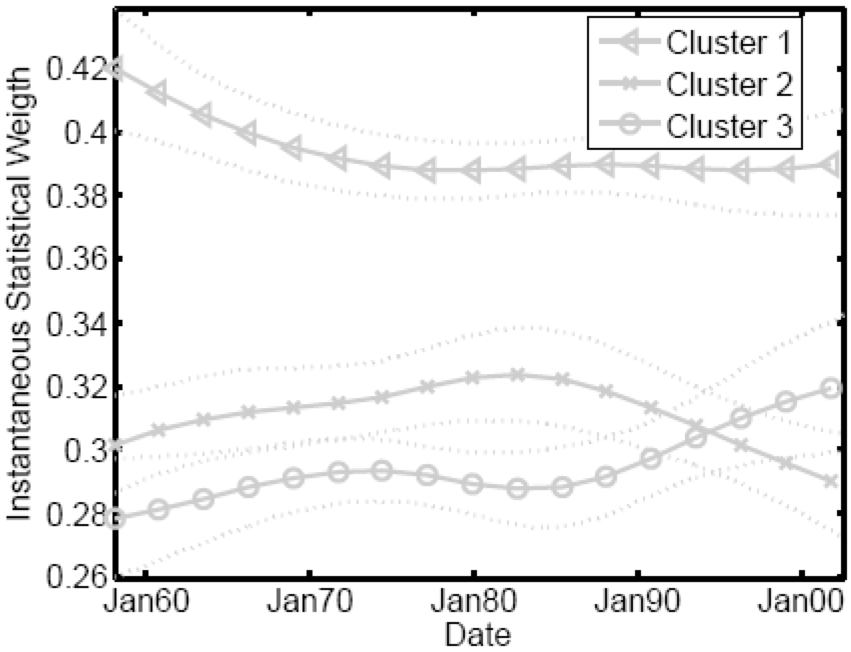


[Cluster 3]



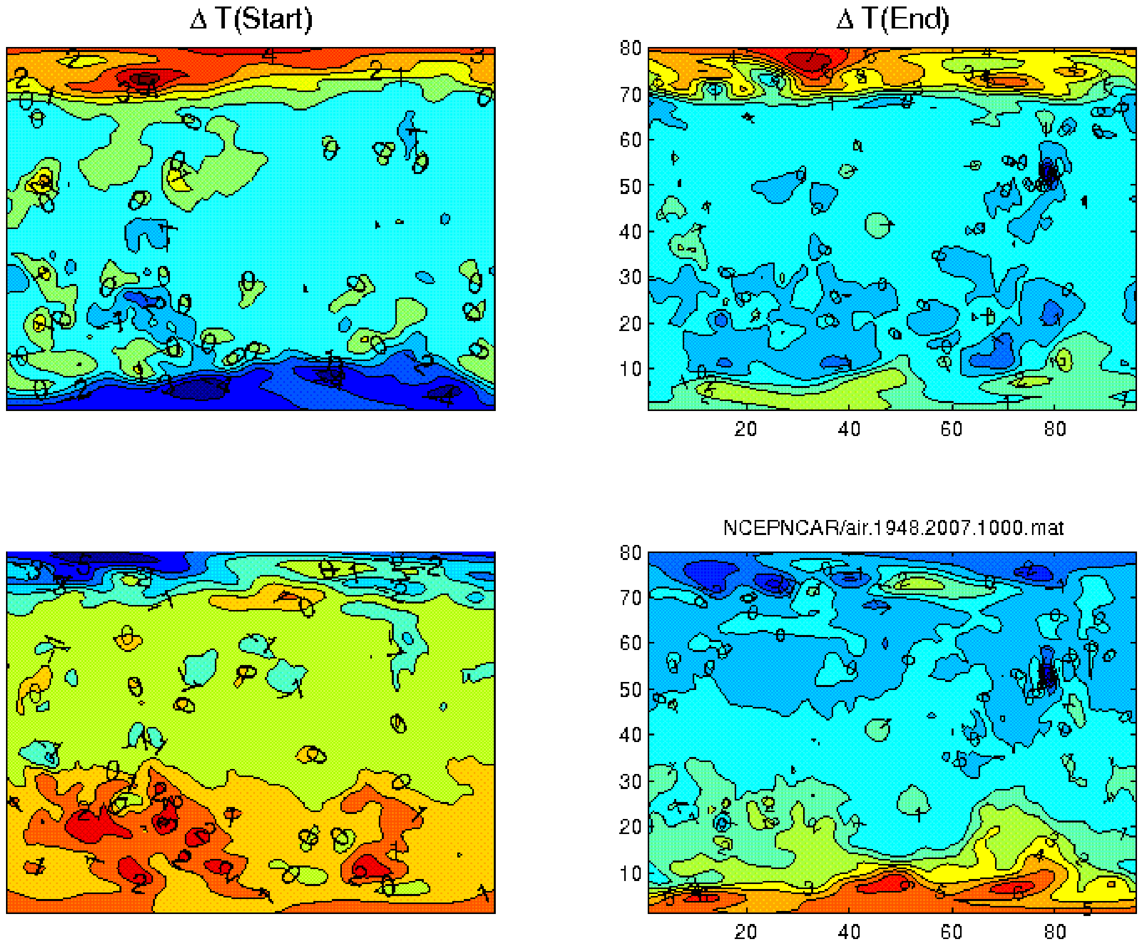
European Data Analysis

Temperature Data in Europe: 29x20 grid



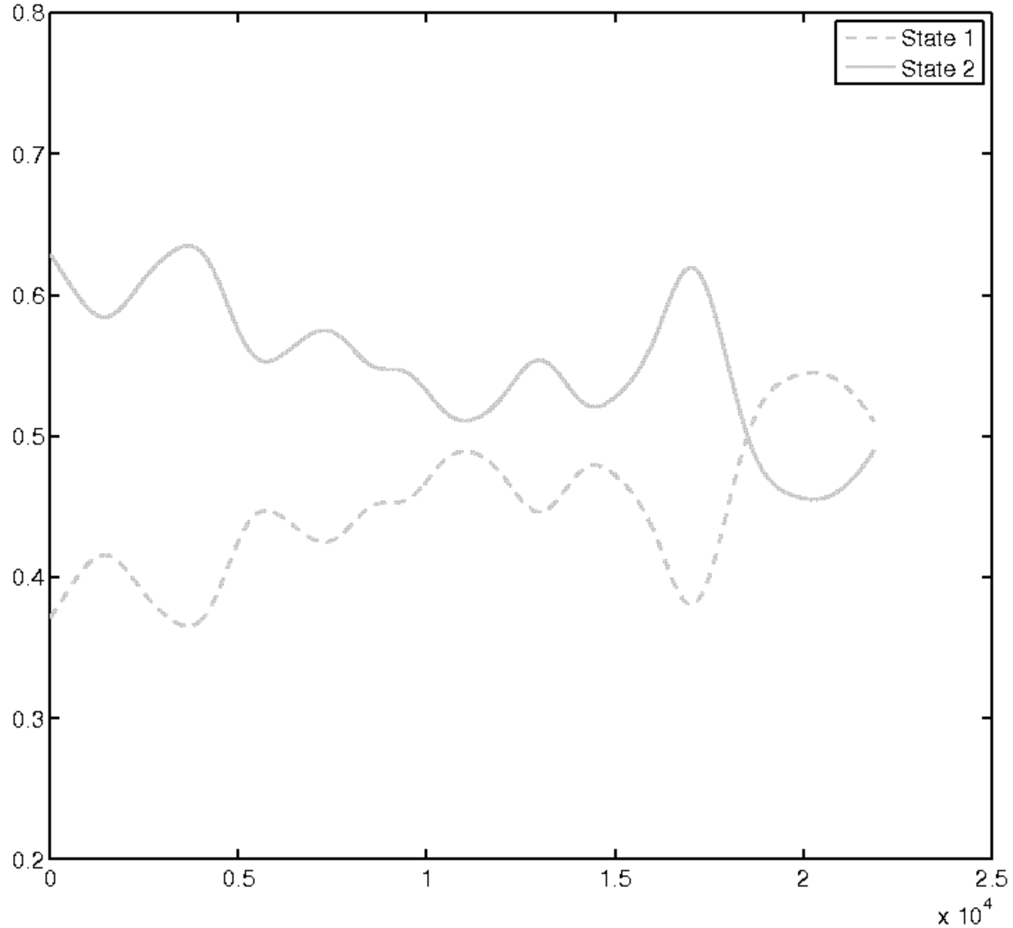
Global Data Analysis

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)

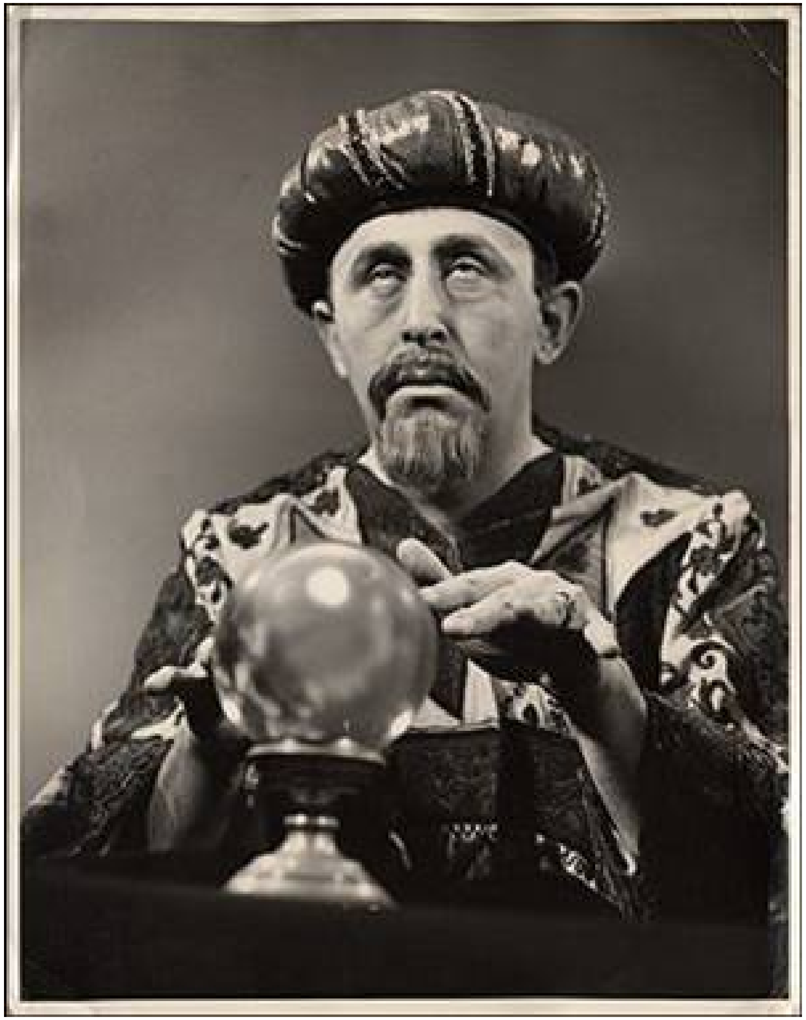


Take-Home-Messages

1. Different methods for analysis of non-stationary data were presented

2. One can use similar approaches in both continuous and discrete data analysis

3. Issue of non-stationarity is important in analysis of historical time series



Thank you for attention!