Computational Time Series Analysis
(IPAM, Part I, stationary methods)

Illia Horenko

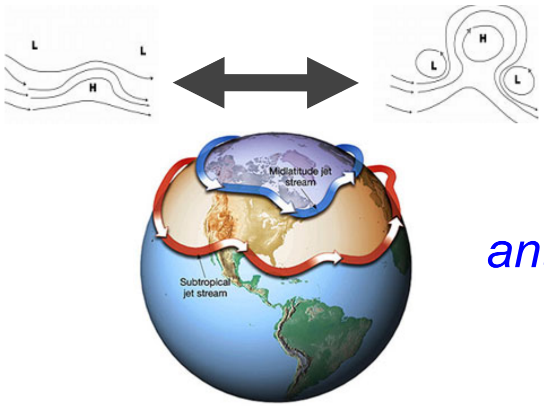
Research Group "*Computational Time Series Analysis* "
Institute of Mathematics
Freie Universität Berlin
Germany

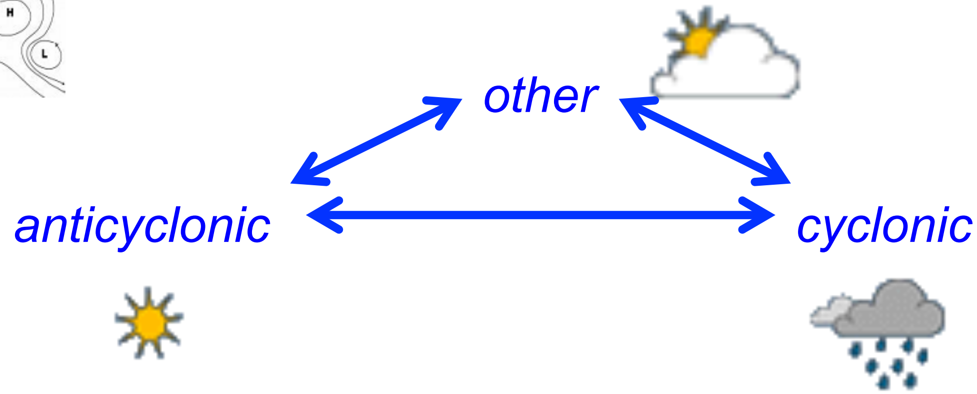
and

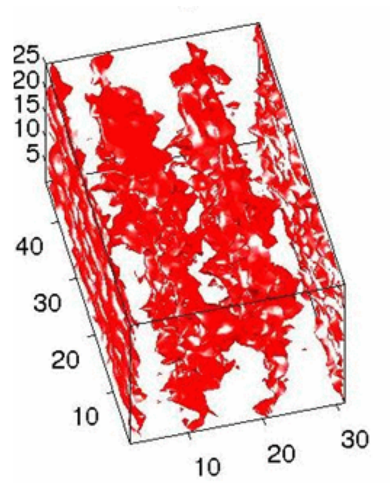
Institute of Computational Science
Universita della Swizzera Italiana, Lugano
Switzerland

Motivation (I): Data

Inverse problem: data-based identification of model parameters







time series of vectors (continous, finite dim.)

time series of regimes (discrete, finite dim.)

time series of functions (continuous, infin. dim.)

Challenges:

VERY LARGE (~Tb size) , non-Markovian (memory), non-stationary (trends), multidimensional, ext. influence

Motivation (II): Models

Inverse problem: data-based identification of model parameters

Dynamical
(f.e., deterministic, stochastic, ...)

Geometrical
(f.e, essential manifolds, ...)

Models

Mixed
(f.e., MTV, PIP/POP, ...)

Challenges:

existence and uniqueness, robustness, conditioning

Plan for Today (stationary case)

- *"eagle eye" perspective on* ***stochastic processes*** *from the viewpoint of* ***deterministic dynamical systems***

- *geometric model inference**: EOF/SSA***

- *multivariate dynamical model inference:* ***VARX***

- *handling the ill-posed problem*

- *motivation for tomorrow : examples where stationarity assumption does not work*

Plan for Today (stationary case)

- *“eagle eye“ perspective on* ***stochastic processes*** *from the viewpoint of* ***deterministic dynamical systems***

- *geometric model inference****: EOF/SSA***

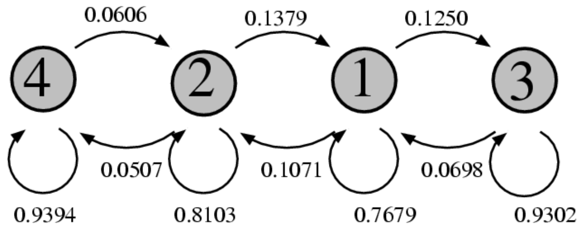
- *multivariate dynamical model inference:* ***VARX***

- *handling the ill-posed problem*

- *motivation for tomorrow : examples where stationarity assumption does not work*

Classification of Stochastic Process

Discrete State Space,
Discrete Time:
Markov Chain



Discrete State Space,
Continuous Time:
Markov Process

Stochastic Processes

Continuous State Space,
Discrete Time:
Autoregressive Process

$$X_{t+\tau} = \sum_{k=0}^{p} \alpha_k X_{t-k\tau} + \sigma \epsilon_t$$

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$dX_t = f\left(X_t, t\right) dt + \sigma dW_t$$

Discrete Markov Process…

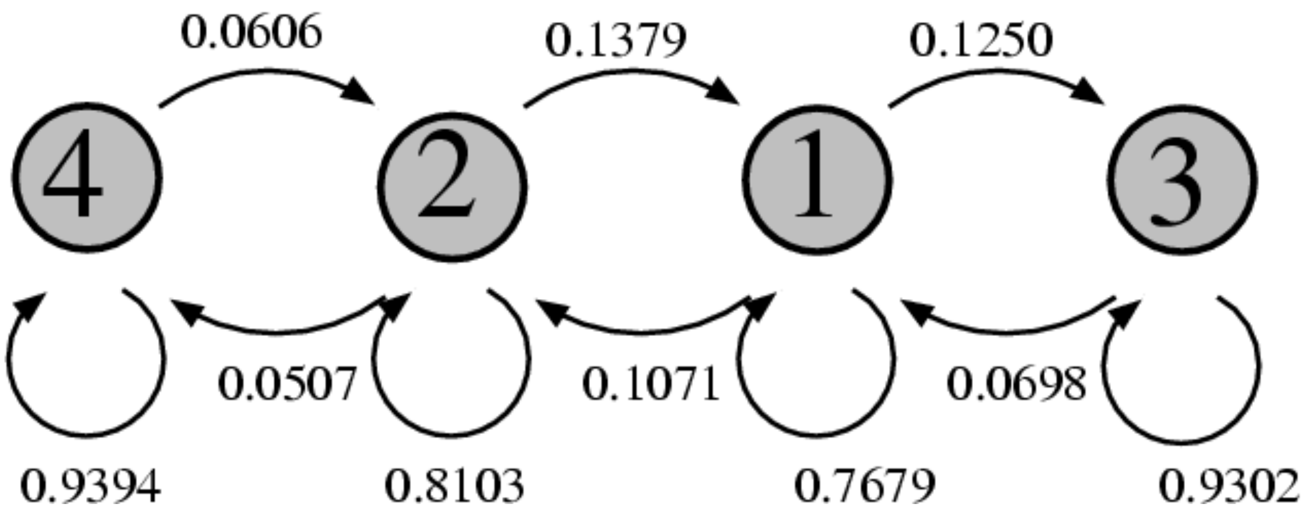
Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, X_{t-\tau} = s_i] = \mathbb{P}[X_t = s_j | X_{t-\tau} = s_i] = P_{ij}(t-\tau, \tau)$$

Example:



Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, X_{t-\tau} = s_i] = \mathbb{P}[X_t = s_j | X_{t-\tau} = s_i] = P_{ij}\,(t-\tau, \tau)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, X_{t-\tau} = s_i] = \mathbb{P}[X_t = s_j | X_{t-\tau} = s_i] = P_{ij}(t-\tau, \tau)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

$$= \sum_{k=1} P_{ki}(t)\pi_k(t)$$

Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, X_{t-\tau} = s_i] = \mathbb{P}[X_t = s_j | X_{t-\tau} = s_i] = P_{ij}(t-\tau, \tau)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

$$= \sum_{k=1} P_{ki}(t)\pi_k(t)$$

$$\boxed{\pi(t+\tau) = \pi(t)P(t,\tau)}$$

This Equation is Deterministic

Continuous Markov Process…

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

Continuous Markov Process…

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

$$\frac{\pi(t+\tau) - \pi(t)}{\tau} = \frac{\pi(t)P(t,\tau) - \pi(t)}{\tau}$$

Continuous Markov Process…

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

$$\frac{\pi(t+\tau)-\pi(t)}{\tau} = \frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\lim_{\tau\to 0}\frac{\pi(t+\tau)-\pi(t)}{\tau} = \lim_{\tau\to 0}\frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

Continuous Markov Process…

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

$$\frac{\pi(t+\tau)-\pi(t)}{\tau} = \frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\lim_{\tau\to 0}\frac{\pi(t+\tau)-\pi(t)}{\tau} = \lim_{\tau\to 0}\frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\mathcal{G}(t) = \lim_{\tau\to 0}\frac{P(t,\tau)-\mathcal{I}}{\tau}$$

infenitisimal

generator

Continuous Markov Process…

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

$$\frac{\pi(t+\tau) - \pi(t)}{\tau} = \frac{\pi(t)P(t,\tau) - \pi(t)}{\tau}$$

$$\lim_{\tau \to 0} \frac{\pi(t+\tau) - \pi(t)}{\tau} = \lim_{\tau \to 0} \frac{\pi(t)P(t,\tau) - \pi(t)}{\tau}$$

$$\mathcal{G}(t) = \lim_{\tau \to 0} \frac{P(t,\tau) - \mathcal{I}}{\tau}$$

infenitisimal

generator

$$\dot{\pi} = \pi \mathcal{G}(t)$$

This Equation is Deterministic: ODE

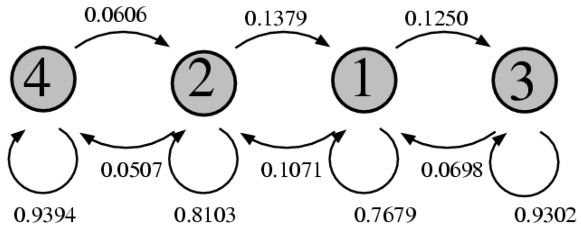
Continuous State Space

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

$\tau \rightarrow 0$

$\dot{\pi} = \pi\mathcal{G}(t)$

Discrete State Space,
Discrete Time:
Markov Chain



Discrete State Space,
Continuous Time:
Markov Process

Stochastic Processes

Continuous State Space,
Discrete Time:
AR(1)

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$

$$dX_t = f(X_t, t)\,dt + \sigma dW_t$$

Deterministic Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R}$

Process:

$$\begin{aligned} \dot{x} &= f(x,t), \\ x(0) &\sim \rho(x,0) \end{aligned}$$

Flow operator:

$$x(t) = \Phi_t(x_0)$$

Properties:

$$\begin{aligned} \Phi_t \cdot \Phi_s(x) &= \Phi_{t+s}(x) \\ \Phi_0(x) &= x \end{aligned}$$

Deterministic Markov Process in **R**

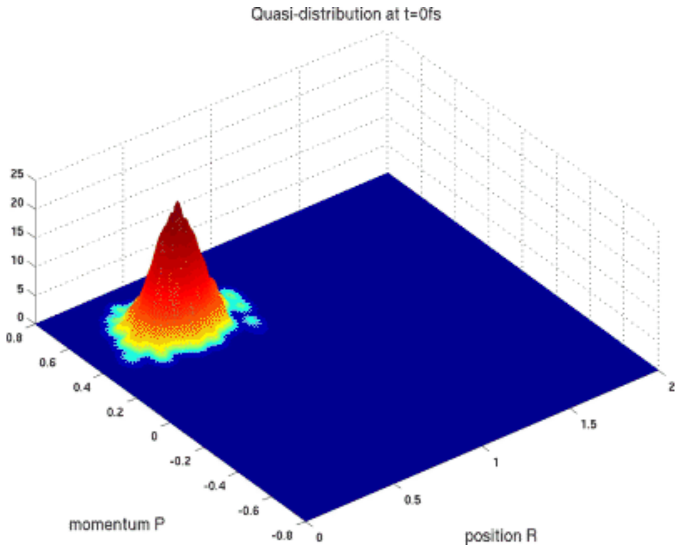
Realizations of the process: $X_t \in \mathbb{R}$

Process:
$$\begin{aligned} \dot{x} &= f(x,t)\,, \\ x(0) &\sim \rho(x,0) \end{aligned}$$

Flow operator:
$$x(t) = \Phi_t(x_0)$$

Properties:
$$\begin{aligned} \Phi_t \cdot \Phi_s(x) &= \Phi_{t+s}(x) \\ \Phi_0(x) &= x \end{aligned}$$

Example: ODE with $x \in \mathcal{R}^2$



Deterministic Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R}$

Process:

$$\begin{aligned} \dot{x} &= f(x,t), \\ x(0) &\sim \rho(x,0) \end{aligned}$$

Flow operator:

$$x(t) = \Phi_t(x_0)$$

Properties:

$$\begin{aligned} \Phi_t \cdot \Phi_s(x) &= \Phi_{t+s}(x) \\ \Phi_0(x) &= x \end{aligned}$$

Liouville Theorem: let $\Phi_t(X)$ *be the flow (*$X \in \mathcal{R}^n$ *), given as a solution of*

$$\frac{d}{dt}\Phi_t(X) = f(\Phi_t(X), t)$$

If $\rho(X,t)$ *is defined as*

$$\int v(\Phi_t(X))\,\rho(X,0)\,dx = \int v(X)\,\rho(X,t)\,, \forall v \in \mathcal{H}$$

then the following equation is fulfilled:

$$\frac{\partial}{\partial t}\rho(X,t) + \text{div}\,(f(X,t)\rho(X,t)) = 0$$

Deterministic Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R}$

Process:

$$\begin{aligned} \dot{x} &= f(x,t), \\ x(0) &\sim \rho(x,0) \end{aligned}$$

Flow operator:

$$x(t) = \Phi_t(x_0)$$

Properties:

$$\begin{aligned} \Phi_t \cdot \Phi_s(x) &= \Phi_{t+s}(x) \\ \Phi_0(x) &= x \end{aligned}$$

Liouville Theorem: let $\Phi_t(X)$ *be the flow (*$X \in \mathcal{R}^n$ *), given as a solution of*

$$\frac{d}{dt}\Phi_t(X) = f(\Phi_t(X), t)$$

Excercise 1: think of the proof (hint: use the partial integration), think about the Hilbert space H and appropriate boundary conditions

$$\int v(\Phi_t(X))\,\rho(X,0)\,dx = \int v(X)\,\rho(X,t), \forall v \in \mathcal{H}$$

then the following equation is fulfilled:

$$\frac{\partial}{\partial t}\rho(X,t) + \text{div}\,(f(X,t)\rho(X,t)) = 0$$

Stochastic Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R} \qquad \{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$

Process:

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$
$$\pi(X,t) = \mathbb{P}[X_t \in X | X_{t-\tau}]$$

$$\pi(X, t+\tau) = \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau\partial_X f(X)\}\pi(X,t)$$

(follows from Ito‘s lemma)

This Equation is Deterministic

Stochastic Markov Process in **R**

Infenitisimal Generator:

$$\mathbf{G} = \lim_{\tau \to 0} \frac{\exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau\partial_X f(X)\} - \mathbf{I}}{\tau}$$

$$= \frac{\sigma^2}{2}\partial_X^2 + \partial_X f(X)$$

Markov Process Dynamics:

$$\dot{\pi}(X,t) = \frac{\sigma^2}{2}\partial_X^2 \pi(X,t) + \partial_X (f(X)\pi(X,t))$$

This Equation is Deterministic: PDE

Numerics of Stochastic Processes

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

Discrete State Space,
Discrete Time:
Markov Chain

$\tau \to 0$

$$\dot{\pi} = \pi\mathcal{G}(t)$$

Discrete State Space,
Continuous Time:
Markov Process

Continuous State Space,
Discrete Time:
Autoregressive Process

$$\pi(X,t+\tau) = \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau f(X)\partial_X\}\pi(X,t)$$

$\tau \to 0$

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$\partial_t \pi(X,t) = \mathbf{G}\pi(X,t)$$

Numerics of Stochastic Processes

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

Discrete State Space,
Discrete Time:
Markov Chain

$\tau \to 0$

$$\dot{\pi} = \pi \mathcal{G}(t)$$

Discrete State Space,
Continuous Time:
Markov Process

Deterministic Numerics of ODEs and PDEs!

Continuous State Space,
Discrete Time:
Autoregressive Process

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

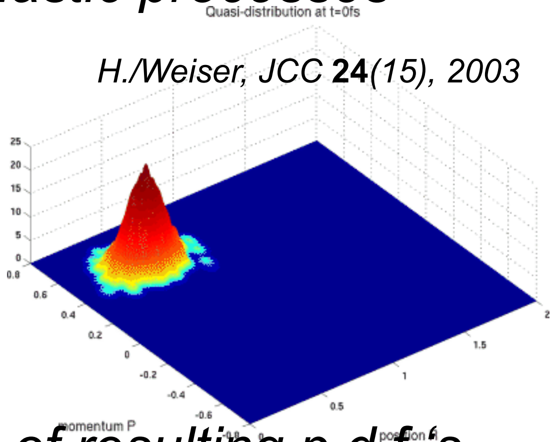
$$\pi(X, t+\tau) = \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau f(X)\partial_X\}\pi(X,t)$$

$\tau \to 0$

$$\partial_t \pi(X,t) = \mathbf{G}\pi(X,t)$$

Conclusions

1) *Numerical Methods from ODEs and (multidimensional) PDEs like Runge-Kutta-Methods, FEM and (adaptive) Rothe particle methods are applicable to stochastic processes*

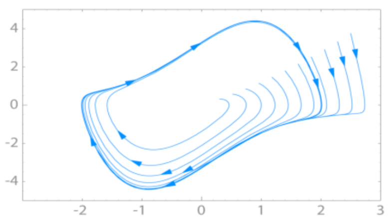


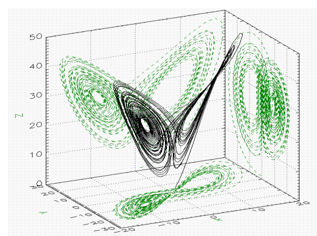
H./Weiser, JCC ***24****(15), 2003*

2) *Monte-Carlo-Sampling of resulting p.d.f.'s*

Stochastic Numerics = ODE/PDE numerics + Rand. Numb. Generator

3) *Concepts from the Theory of Dynamical Systems are Applicable (Whitney and Takens theorems, model reduction by identification of attractors)*





Plan for Today (stationary case)

- *“eagle eye“ perspective on* ***stochastic processes*** *from the viewpoint of* ***deterministic dynamical systems***

- *geometric model inference**: EOF/SSA***

- *multivariate dynamical model inference:* ***VARX***

- *handling the ill-posed problem*

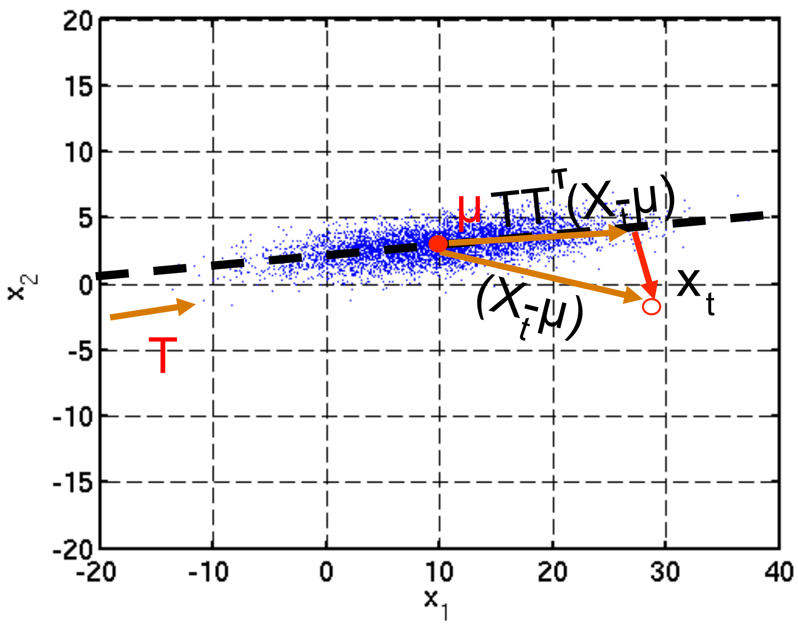
- *motivation for tomorrow : examples where stationarity assumption does not work*

EOF/PCA/SSA/...: Geometrical methods

For a given time series x_t^q we look for a

minimum of the *reconstruction error*

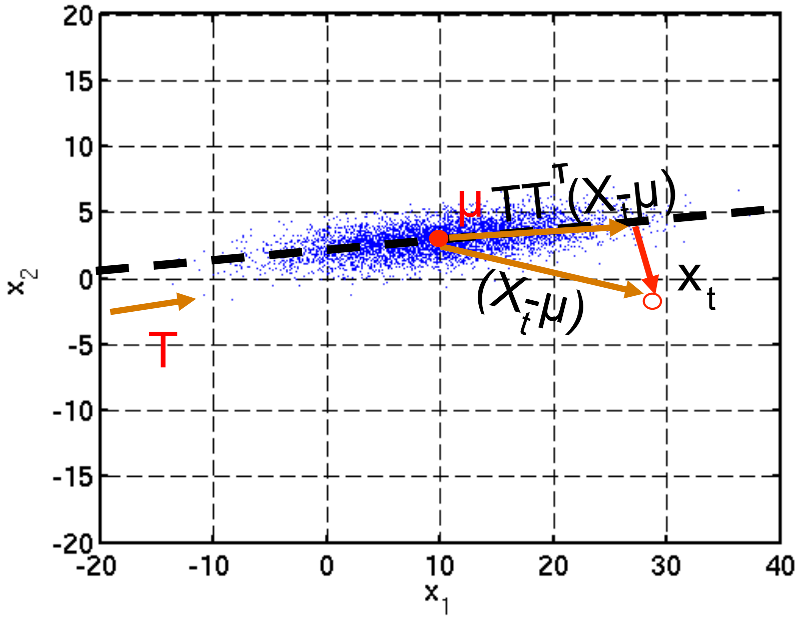
$$\left\| (x_t - \mu) - \mathbf{T}\,\mathbf{T}^{\mathsf{T}} (x_t - \mu) \right\|_2^2$$



EOF/PCA/SSA/...: Geometrical methods

For a given time series x_t^q we look for a minimum of the *reconstruction error*

$$\left\| (x_t - \mu) - \mathbf{T}\,\mathbf{T}^{\mathsf{T}} (x_t - \mu) \right\|_2^2$$



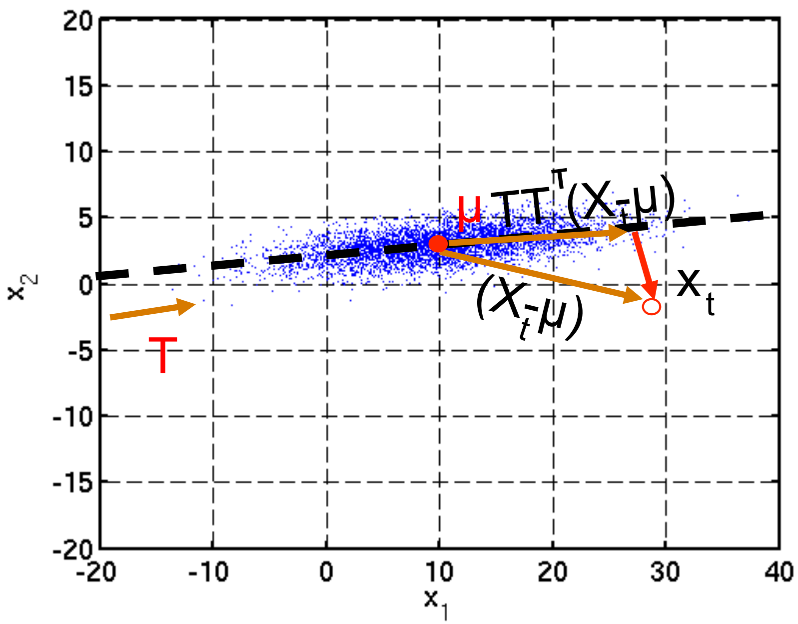
$$\sum_{t=1}^{T} \left\| x_t - \mathbf{T}\mathbf{T}^{\mathsf{T}} x_t \right\|_2^2 \to \min$$

$$\mathbf{T}^{\mathsf{T}}\mathbf{T} = Id,$$

EOF/PCA/SSA/...: Geometrical methods

For a given time series x_t^q we look for a minimum of the *reconstruction error*

$$\left\|(x_t-\mu)-\mathbf{T}\,\mathbf{T}^{\top}(x_t-\mu)\right\|_2^2$$



$$\sum_{t=1}^{T}\left\|x_t-\mathbf{T}\mathbf{T}^{\top}x_t\right\|_2^2 \longrightarrow \min$$

$$\mathbf{T}^{\top}\mathbf{T} = Id,$$

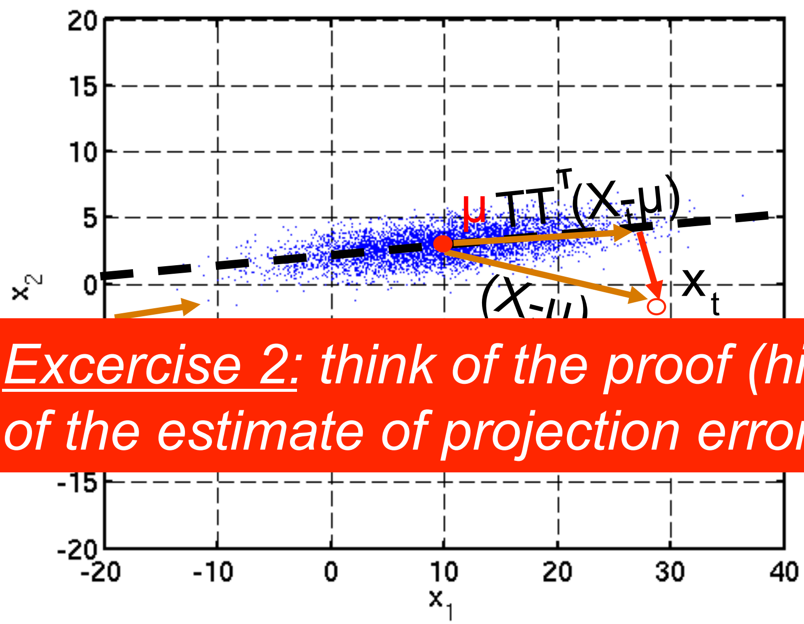
$$\mathrm{Cov}\,(x_t) = \sum_t x_t x_t^{\dagger}$$

$$\mathrm{Cov}\,(x_t)\,T = \lambda T$$

EOF/PCA/SSA/...: Geometrical methods

For a given time series x_t^q we look for a minimum of the *reconstruction error*

$$\left\|(x_t - \mu) - \mathbf{T}\,\mathbf{T}^\top (x_t - \mu)\right\|_2^2$$



$$\sum_{t=1}^{T}\left\|x_t - \mathbf{T}\mathbf{T}^\top x_t\right\|_2^2 \to \min$$

$$\mathbf{T}^\top\mathbf{T} = Id,$$

$$\mathrm{Cov}\,(x_t) = \sum_t x_t x_t^\top$$

$$\mathrm{Cov}\,(x_t)\,T = \lambda T$$

Excercise 2: think of the proof (hint: use the Lagrange multipliers), think of the estimate of projection error

Plan for Today (stationary case)

- *"eagle eye" perspective on* ***stochastic processes*** *from the viewpoint of* ***deterministic dynamical systems***

- *geometric model inference****: EOF/SSA***

- *multivariate dynamical model inference:* ***VARX***

- *handling the ill-posed problem*

- *motivation for tomorrow : examples where stationarity assumption does not work*

Predicting Stochastic Processes

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

Discrete State Space,
Discrete Time:
Markov Chain

$\tau \to 0$

$$\dot{\pi} = \pi\mathcal{G}(t)$$

Discrete State Space,
Continuous Time:
Markov Process

Deterministic Numerics of ODEs and PDEs!

Continuous State Space,
Discrete Time:
Autoregressive Process

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$\pi(X,t+\tau) \;=\; \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau\partial_X f(X)\}\pi(X,t)$$

$\tau \to 0$

$$\dot{\pi}(X,t) \;=\; \frac{\sigma^2}{2}\partial_X^2\pi(X,t) + \partial_X(f(X)\pi(X,t))$$

Predicting the Dynamics: VAR(p)

A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$\boxed{Y = BZ + U}$$

$$Z = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_{p-1} & y_p & \cdots & y_{T-1} \\ y_{p-2} & y_{p-1} & \cdots & y_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_0 & y_1 & \cdots & y_{T-p} \end{bmatrix}$$

$$Y = \begin{bmatrix} y_p & y_{p+1} & \cdots & y_T \end{bmatrix} = \begin{bmatrix} y_{1,p} & y_{1,p+1} & \cdots & y_{1,T} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{k,p} & y_{k,p+1} & \cdots & y_{k,T} \end{bmatrix}$$

$$B = \begin{bmatrix} c & A_1 & A_2 & \cdots & A_p \end{bmatrix} = \begin{bmatrix} c_1 & a^1_{1,1} & a^1_{1,2} & \cdots & a^1_{1,k} & \cdots & a^p_{1,1} & a^p_{1,2} & \cdots & a^p_{1,k} \\ c_2 & a^1_{2,1} & a^1_{2,2} & \cdots & a^1_{2,k} & \cdots & a^p_{2,1} & a^p_{2,2} & \cdots & a^p_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ c_k & a^1_{k,1} & a^1_{k,2} & \cdots & a^1_{k,k} & \cdots & a^p_{k,1} & a^p_{k,2} & \cdots & a^p_{k,k} \end{bmatrix}$$

$$U = \begin{bmatrix} e_p & e_{p+1} & \cdots & e_T \end{bmatrix}$$

Estimating VAR(p) : least squares

A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$Y = BZ + U$$

$$\|Y - BZ\| \rightarrow \min_B$$

Estimating VAR(p) : least squares

A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$\boxed{Y = BZ + U}$$

$$\|Y - BZ\| \to \min_B$$

$$L(B) = \mathrm{tr}\left[(Y - BZ)(Y - BZ)^T\right] \to \min_B$$

$$\frac{\partial L}{\partial B}(B_{est}) = 0$$

Estimating VAR(p) : least squares

A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$\boxed{Y = BZ + U}$$

$$\|Y - BZ\| \to \min_B$$

$$L(B) = \mathrm{tr}\left[(Y - BZ)(Y - BZ)^T\right] \to \min_B$$

$$\frac{\partial L}{\partial B}(B_{est}) = 0$$

$$\boxed{B_{est} = YZ^T\left(ZZ^T\right)^{-1}}$$

Estimating VAR(p) : least squares

A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$Y = BZ + U$$

$$\|Y - BZ\| \rightarrow \min_B$$

<u>Excercise 3:</u> think of the proof (hint: use the matrix function derivatives from the Matrix Coockbook: www2.imm.dtu.dk/pubdb/views/edoc_download.../imm3274.pdf)

$$B_{est} = YZ^T \left(ZZ^T\right)^{-1}$$

Plan for Today (stationary case)

- *“eagle eye“ perspective on* ***stochastic processes*** *from the viewpoint of* ***deterministic dynamical systems***

- *geometric model inference****: EOF/SSA***

- *multivariate dynamical model inference:* ***VARX***

- *handling the ill-posed problem*

- *motivation for tomorrow : examples where stationarity assumption does not work*

Estimating VAR(p) : regularization

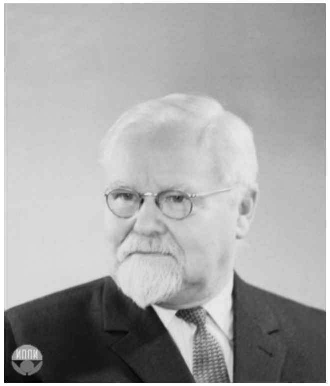
A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$Y = BZ + U$$



A. N. Tikhonov
(http://en.wikipedia.org)

$$\|Y - BZ\| + \|\alpha B\| \to \min_B$$

Estimating VAR(p) : regularization

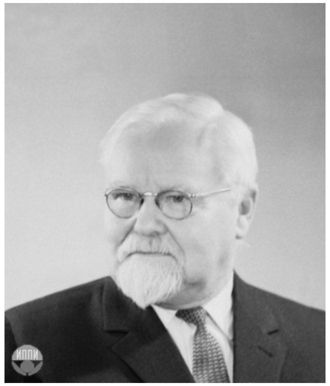
A ***(reduced) p-th order VAR***, denoted ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

where c is a $k \times 1$ vector of constants (**intercept**), A_i is a $k \times k$ matrix (for every $i = 1, \ldots, p$) and e_t is a $k \times 1$ vector of error terms satisfying

1. $\mathrm{E}(e_t) = 0$ — every error term has mean zero;
2. $\mathrm{E}(e_t e_t') = \Omega$ — the contemporaneous covariance matrix of error terms is Ω (a $n \times n$ positive definite matrix);
3. $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

$$Y = BZ + U$$



A. N. Tikhonov
(http://en.wikipedia.org)

$$\|Y - BZ\| + \|\alpha B\| \to \min_B$$

$$B_{est}(\alpha) = YZ^T\left(ZZ^T + \alpha\alpha^T\right)^{-1}$$

Plan for Today (stationary case)

- *“eagle eye“ perspective on* ***stochastic processes*** *from the viewpoint of* ***deterministic dynamical systems***

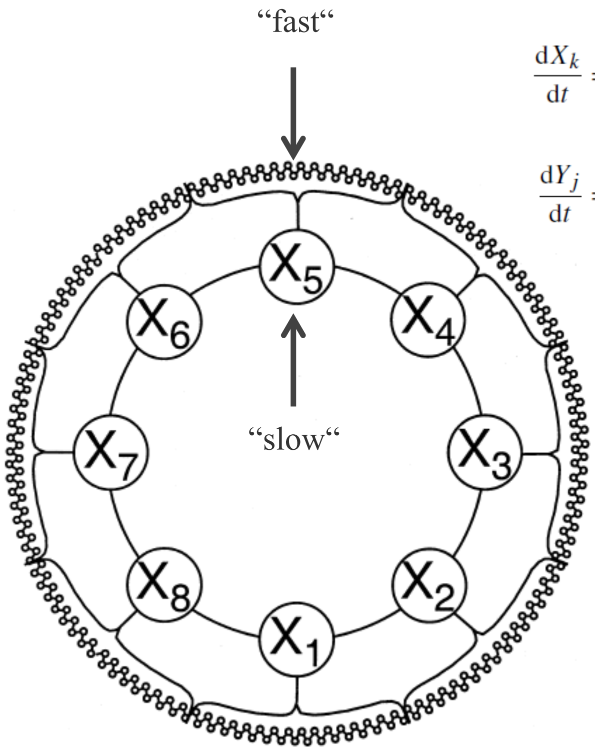
- *geometric model inference**: EOF/SSA***

- *multivariate dynamical model inference:* ***VARX***

- *handling the ill-posed problem*

- *motivation for tomorrow : examples where stationarity assumption does not work*

Lorenz96: “order zero“ atmosph. model



Wilks, *Quart.J.Royal Met.Soc., 2005*

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b} \sum_{j=J(k-1)+1}^{kJ} Y_j; \quad k = 1, \ldots, K$$

$$\frac{dY_j}{dt} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b} X_{\mathrm{int}[(j-1)/J]+1}; \quad j = 1, \ldots, JK.$$

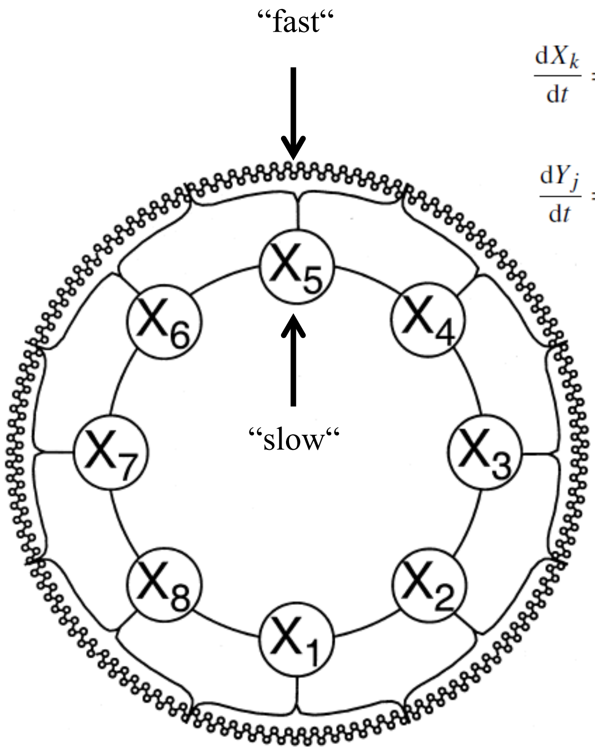
stationary model
(time-independent parameters)

VARX (standard stationary stochastic Model)

$$y_t \;=\; \mu + \sum_{q=1}^{m} \mathbf{A}_q y_{t-q\tau} + \mathbf{B}\phi\left(x_t\right) + \mathbf{C}\epsilon_t$$

- Lorenz, *Proc. Of. Sem. On Predict., 1996*
- Majda/Timoffev/V.-Eijnden, *PNAS, 1999*
- Orell, *JAS, 2003*
- Wilks, Quart.J.Royal Met.Soc., 2005
- Crommelin/V.-Eijnden, *Journal of Atmos. Sci., 2008*

Lorenz96: “order zero“ atmosph. model



Wilks, *Quart.J.Royal Met.Soc., 2005*

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} Y_j; \quad k = 1, \ldots, K$$

$$\frac{dY_j}{dt} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b}X_{\mathrm{int}[(j-1)/J]+1};$$

+ *F(t)*

+  **=** **non-stationary model (time-dependent parameters)**

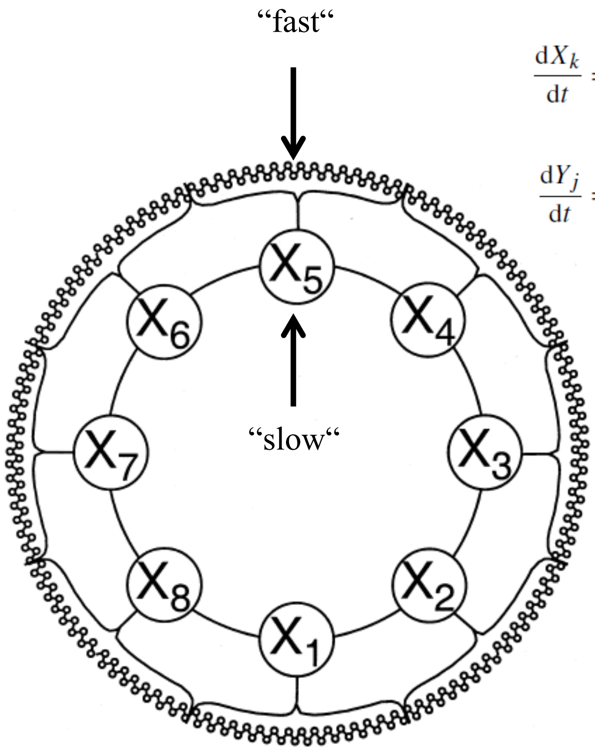
???

VARX (standard stationary stochastic Model)

$$y_t = \mu + \sum_{q=1}^{m} \mathrm{A}_q y_{t-q\tau} + \mathrm{B}\phi(x_t) + \mathrm{C}\epsilon_t$$

- Lorenz, *Proc. Of. Sem. On Predict., 1996*
- Majda/Timoffev/V.-Eijnden, *PNAS, 1999*
- Orell, *JAS, 2003*
- Wilks, Quart.J.Royal Met.Soc., 2005
- Crommelin/V.-Eijnden, *Journal of Atmos. Sci., 2008*

Lorenz96: “order zero“ atmosph. model



Wilks, *Quart.J.Royal Met.Soc.*, *2005*

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} Y_j; \quad k = 1, \ldots, K$$

$$\frac{dY_j}{dt} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b} X_{\mathrm{int}[(j-1)/J]+1};$$ **+ *F(t)***

+

=

non-stationary model (time-dependent parameters)

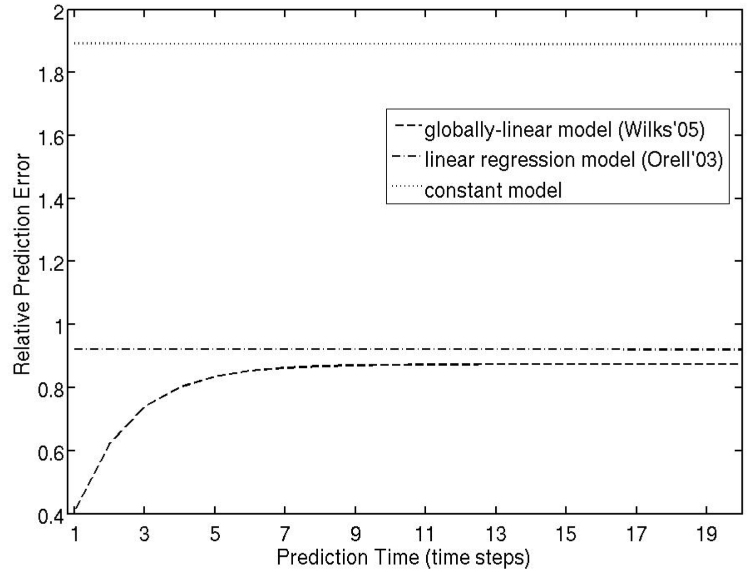
???

VARX (standard stationary stochastic Model)

$$y_t = \mu + \sum_{q=1}^{m} A_q y_{t-q\tau} + B\phi(x_t) + C\epsilon_t$$

- Lorenz, *Proc. Of. Sem. On Predict., 1996*
- Orell, *JAS, 2003*
- Wilks, Quart.J.Royal Met.Soc., 2005
- Crommelin/V.-Eijnden, *Journal of Atmos. Sci., 2008*

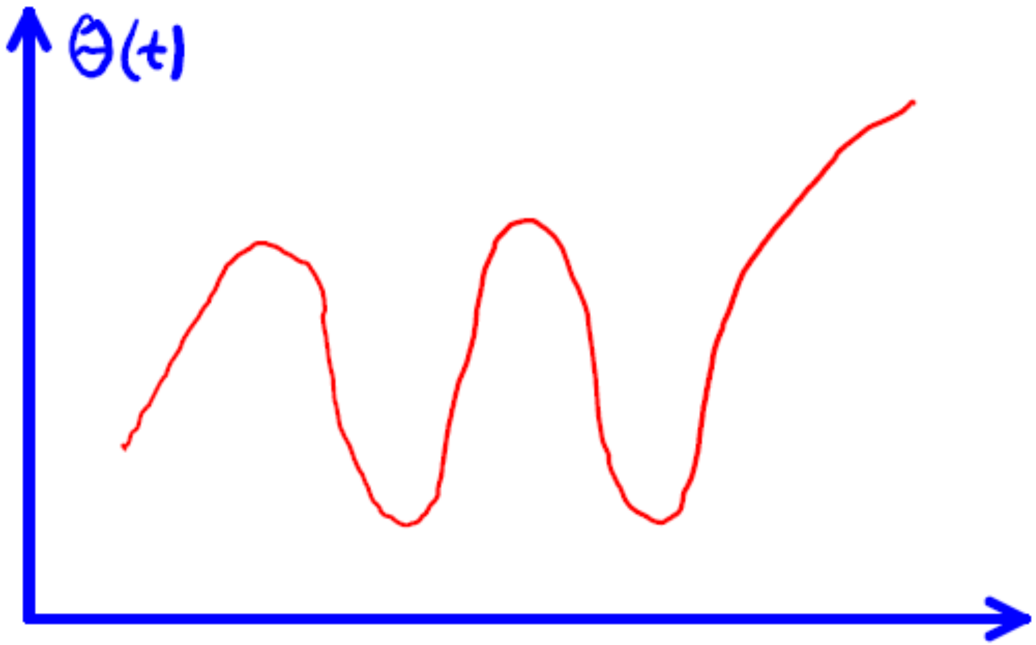
Predictors of Lorenz ‘96-Model



Tomorrow: Non-stationarity in TSA

Direct problem: $\mathbf{F}\left(x_t, \ldots, x_{t-m\tau}, \theta(t), t\right) = 0$

Inverse problem: $\int_0^T g(x_s, \theta(s))ds \rightarrow \min_{\theta(t)}$



<u>Example:</u>

$$x_t = \theta(t) + \epsilon_t \qquad \epsilon_t \text{ (i.i.d.) } \mathbb{E}\left[\epsilon_t\right] = 0$$

$$g\left(x_t, \theta(t)\right) = \|x_t - \theta_t\|_2$$