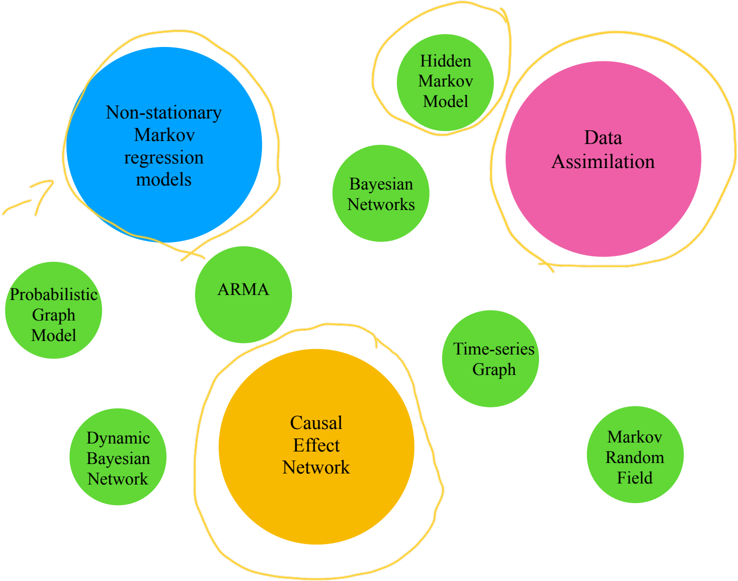
Statistical Data Analysis

Dr. Jana de Wiljes

19. Januar 2022

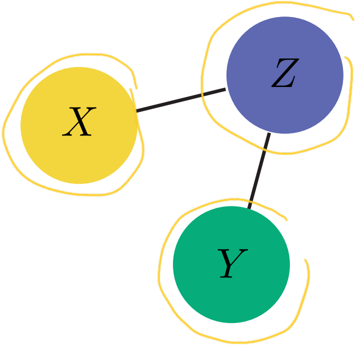
Universität Potsdam

Connecting the dots....



Probabilistic graph model

Probabilistic graph model (PGM)

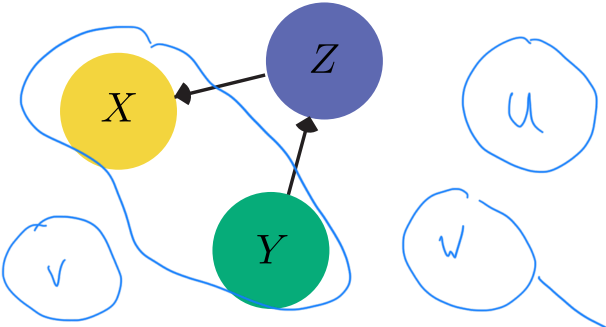


- **Nodes** are associated with random variables $X, Y, Z : \Omega \to \mathbb{R}^N$
- Absence of **edge** between X and Y indicates independence of variables X and Y

X and Y are independent

Bayesian Network

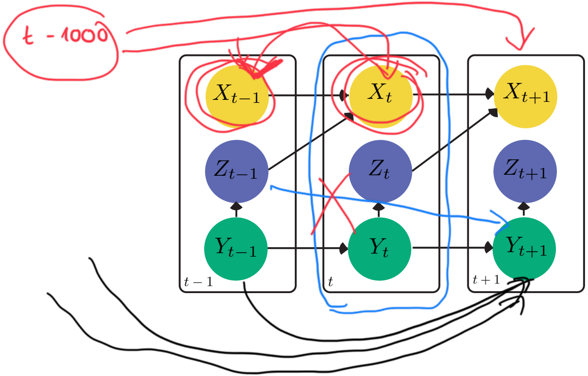
Bayesian Network (BN)



- BN is directed acyclic graph
- edges corresponds to conditional dependency, i.e., X and Y are conditionally independent if $P(X, Y|Z) = P(X|Z)P(Y|Z)$
- Nodes corresponds to random variables $X, Y, Z : \Omega \to \mathbb{R}^N$

Dynamic Bayesian Network

Dynamic Bayesian Network (DBN)

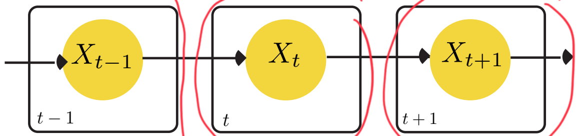


$X_{t-1} \longrightarrow X_t$

- Adding concept of time to BNs, i.e., random variables are time dependent $X_t : \Omega \to \mathbb{R}^N$, $t \in \mathbb{Z}$
- allows us to model time series or sequences

Dynamic Bayesian Network: Markov process

...



X_{t+1}

$X_{t-\tau+1}$

- Markov model of order 1
- Markov property is fulfilled:

$$P(X_{t+1}|X_1 = x_1, \ldots, X_t = x_t) = P(X_{t+1}|X_t = x_t) \quad (1)$$

1   t

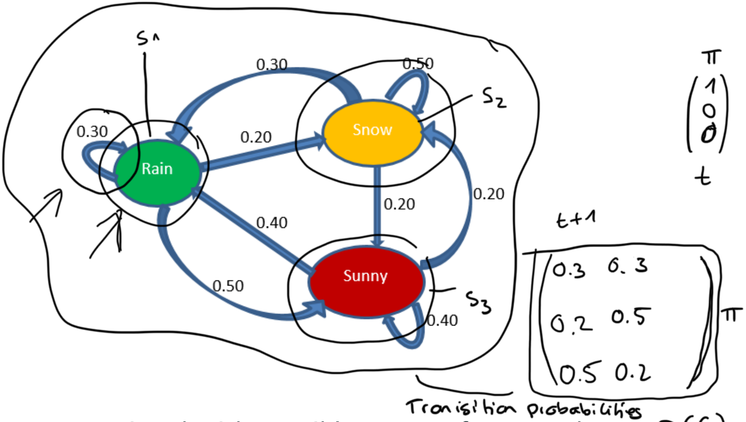
- higher oder Markov models can also be represented as DBNs

$$P(X_{t+1} | X_1 = x_1, \ldots, X_{t-\tau}, \ldots, X_t = x_t)$$

$$= P(X_{t+1} | X_{t-\tau}, \ldots, X_t = x_t)$$ of order τ

Discrete Markov Process

Graph representation of state transitions



O(6)

- **Nodes** are associated with possible states of one random variable $X : \Omega \to \{s_1, \ldots, s_n\}$ e.g., $X(\omega) \in \{\text{rain, snow, sunny}\}$
- **Edges** between states indicate that the probability to go from one state to the next is larger the zero

Markov regression model

Assume: Assume Markov process X_t is influenced by external factors and transition matrix M is given by

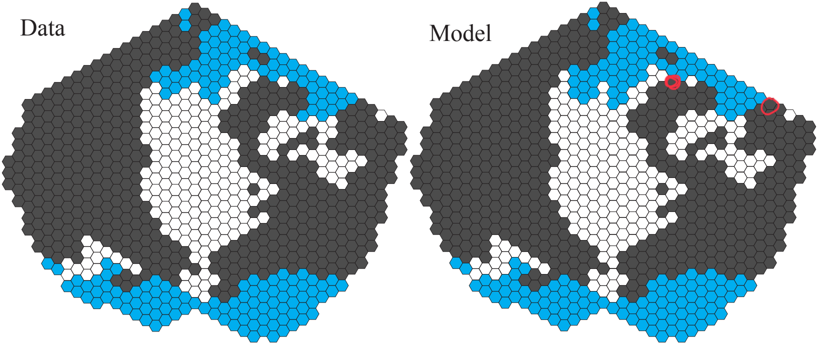
$$\boxed{M(t)} = \sum_{j=1}^{N} A_j u_j(t) \tag{2}$$

where $u_j(t)$ is one of N external factors at time t.

Goal: fit $M(t)$ according to time series x_t and external factors $u_j(t)$

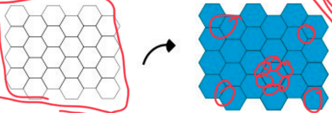
Problem: ill-posed $\rightarrow$ need to assume local stationarity

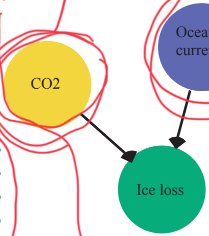
Arctic sea ice coverage



Tipping point scenarios

today from year





0.5

- Manipulate certain causes and see the effects via fitted regression model
- Create tipping points/ certain outcomes
- compute corresponding conditional probabilities

[0, 1]

0.3

Link to Data assimilation

Model:

$$X_{k+1} = AX_k + \epsilon_k, \ \epsilon_k \sim N(0, \mathbf{Q}) \tag{3}$$

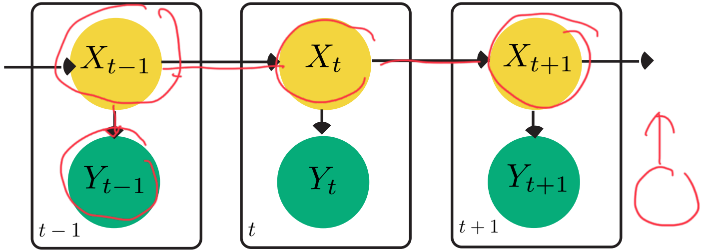
where $X_{k+1} \in \mathbb{R}^{N_x}$ and $\mathbf{Q} \in \mathbb{R}^{N_x \times N_x}$.

Observations:

$$Y_{k+1} = HX_{k+1} + \nu_k, \ \nu_k \sim N(0, \mathbf{R}) \tag{4}$$

where $Y_{k+1} \in \mathbb{R}^{N_y}$ and $\mathbf{R} \in \mathbb{R}^{N_y \times N_y}$.

Kalman Filter as Dynamic Bayesian Networks



- the current state of the system X_t depends on the previous state of the system X_{t-1}, dependence is given by evolution model
- Observation Y_t depends on the current state X_t, dependence given by observation operator H

Causality

Causality

Granger Causality: predictive causality, i.e., how good can X_t be predicted knowing $X_{t-\tau}$. Can be tested via fitting VAR model with parameters $A(\tau)$

$$X_t = \sum_{\tau=1}^{\tau_{max}} \overset{\beta}{\overset{\shortparallel}{A(\tau)}} \boxed{X_{t-\tau}} + \eta_t \tag{5}$$

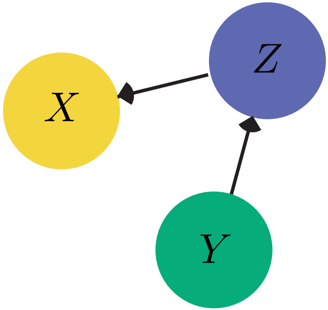
Then $X_{t-\tau}$ is called a Granger cause of X_t if at least one entry of $A(\tau)$ is significantly larger than zero.

(β)

Causality: X causes Y if and only if an intervention or manipulation in X has an effect on Y

Causal Network

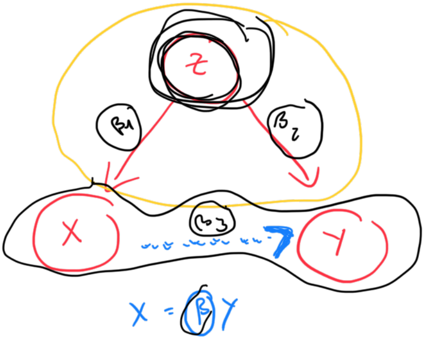
Causal Network



- A causal network is a Bayesian network with the added property that the parents of each node are its direct cause

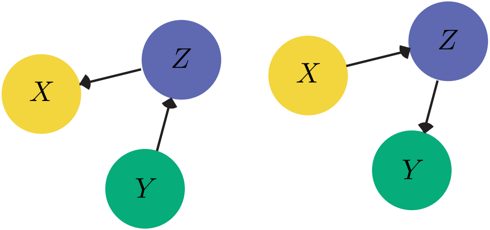
Causality and Correlation

- X, Y are independent $\implies$ X, Y are not correlated $\implies$ there is no causal link between X and Y
- There is a causal link between X and Y $\implies$ X, Y are correlated
- X, Y are correlated $\not\Rightarrow$ there is a causal link between X and Y



Causal discovery

Causal discovery



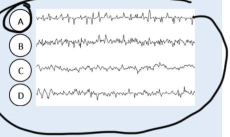
Problem: Can not just use fitted DBN because correlation does not imply causation

Approach: If a certain independence structure is given, there is usually a small number of cases that needs to be checked

Causal Discovery : learn causal relations from observations

t
X
Y
Z

Input: Time-series data



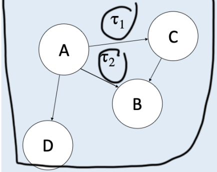
Causal Discovery

Correlation($A_{t-\tau}$, B_t | *Iterate through combinations of conditions*)

PCMCI Algorithm

can deal with auto-correlation, high dimensionality

Output: Causal Structure



Discrete-time structural causal model (SCM)

Given: $\{X_t\}_{t\in\mathbb{Z}}$ be a sequence of real-valued N_X dimensional random variables

Causal links $X^i_{t-\tau} \to X^j_t$

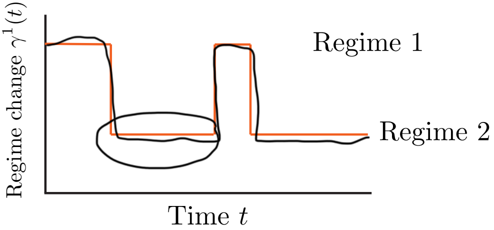
- $X^i_{t-\tau}$ and X^j_t are linked if $X^i_{t-\tau}$ is not conditionally independent of X^j_t given the past of all variables, i.e.,

$$X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t \setminus \{X^i_{t-\tau}\} \tag{6}$$

  with $\not\perp\!\!\!\perp$ denoting the absence of a (conditional) independence
- $X^i_{t-\tau} \in \mathcal{P}^j_t$ if $X^i_{t-\tau}$ and X^j_t are connected by a lag-specific directed link
- parents set $\mathcal{P}^j_t \subset (X_{t-1}, X_{t-2}, \ldots)$

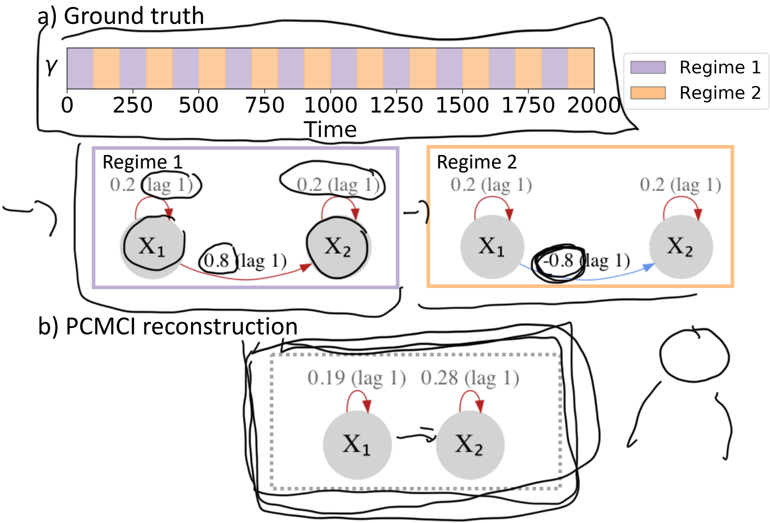
Non-stationary approach

Local stationarity assumption



- some form of persistency is assumed, i.e., for some time the underlying model is stationary
- if the structure of the local stationarity is know one can separate the data accordingly
- if the structure of the local stationarity unknown one needs to detect the hidden switching process $\Gamma \rightarrow$ combine ideas from Runge 2018 and deWiljes et al 2014

Regime-dependent causal relationships



Problem setting

Inverse Problems

Given: time series or data set $\mathbf{x}_t$

Aim: find Θ_t with

$$\mathbf{x}_t = \widehat{\mathbf{G}}_t\left(\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-\tau_{\max}}; \Theta_t\right)$$

with $\widehat{\mathbf{G}}_t = [\widehat{g}_t^1, \ldots, \widehat{g}_t^{N_X}]$

Ansatz:

- choose appropriate distance measure $d(\cdot)$
- assume model structure for Θ_t
- solve optimisation problem

Problem setting:

Assumption: parents and functional dependencies are stationary for an average of N_M consecutive time steps t, and finite number of regimes on the whole time domain

regression parameter

Find: unknown $\Theta_t = [\Gamma(t), \mathcal{P}, \Phi]$

switching between regimes

parents sets (underlying causal) structure

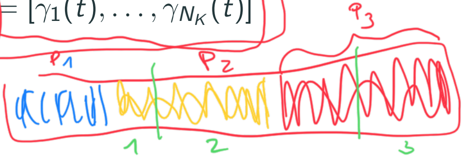
1. a set of regimes' network parameters

$$\mathcal{P}, \Phi = \{\mathcal{P}_1, \dots, \mathcal{P}_{N_K}, \Phi_1, \dots, \Phi_{N_K}\}$$

2. the change points between the regimes given by the regime-assigning process

$$\Gamma(t) = [\gamma_1(t), \dots, \gamma_{N_K}(t)]$$

with $\Gamma(t) \in [0,1]^{N_K \times T}$.



Optimisation problem

$$\mathbf{L}(\Gamma, \mathcal{P}, \Phi) = \sum_{t=0}^{T} \sum_{k=1}^{N_K} \gamma_k(t) d(\mathbf{x}_t - \widehat{\mathbf{G}}_t(\mathcal{P}_k; \Phi_k)) \tag{7}$$

subject to constraints

$$\sum_{k=1}^{N_K} \gamma_k(t) = 1 \quad \forall\ t,\ \text{with } \gamma_k(t) \in [0,1] \tag{8}$$

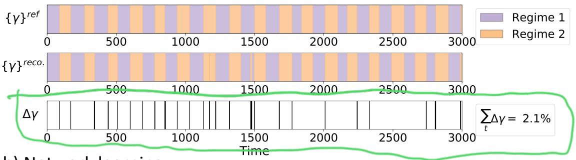
and

$$\sum_{t=1}^{T-1} |\gamma_k(t+1) - \gamma_k(t)| \leq N_C \quad \forall k \tag{9}$$

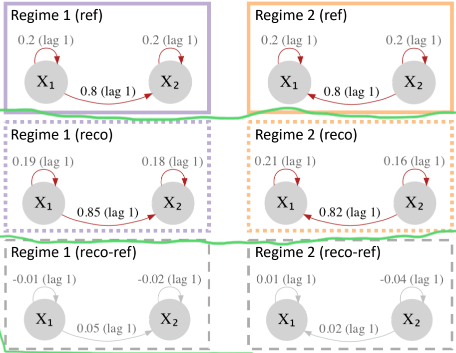
Numerical investigation for different causal scenarios

Arrow direction

a) Regime learning



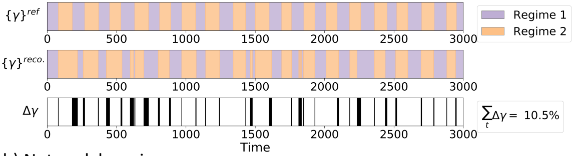
b) Network learning



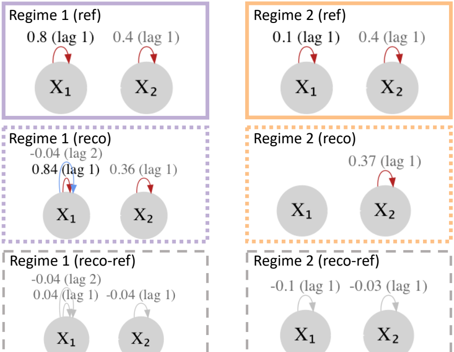
Numerical investigation for different causal scenarios

Causal effect

a) Regime learning



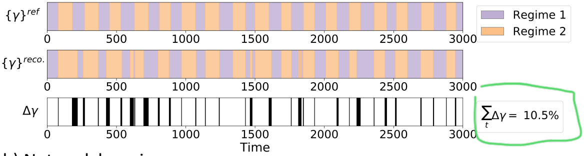
b) Network learning



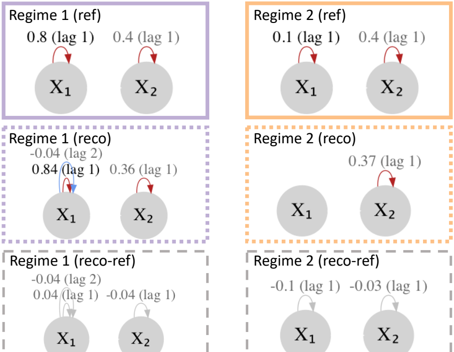
Numerical investigation for different causal scenarios

Causal effect

a) Regime learning



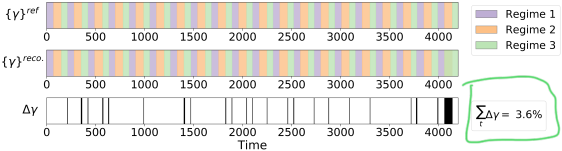
b) Network learning



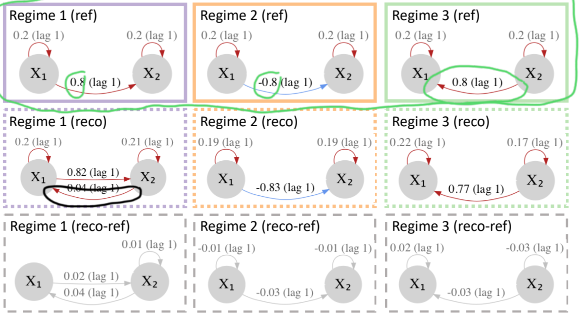
Numerical investigation for different causal scenarios

Sign X^1X^2 and arrow direction

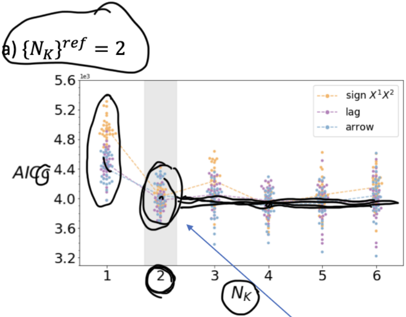
a) Regime learning

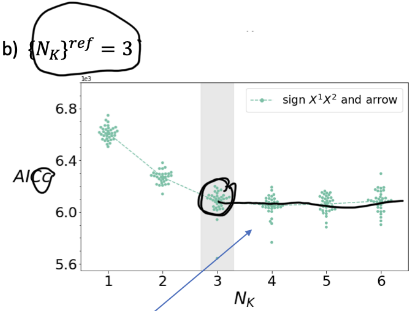


b) Network learning

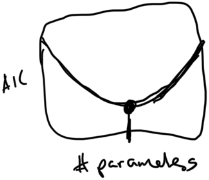


Model selection





AICc flattens for N_K larger than true value.



Model selection

Information criterion:

correction

$$AICc = -2\log(\mathcal{L}) + 2N_{\text{para}} + \frac{2N_{\text{para}}(N_{\text{para}}+1)}{T - N_{\text{para}} - 1} \tag{10}$$

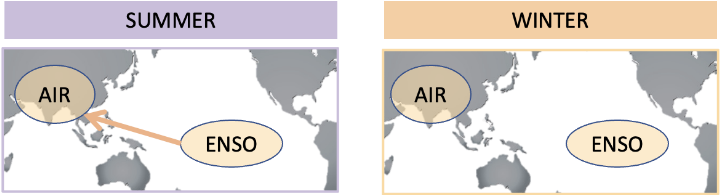
'accuracy

Number of parameters

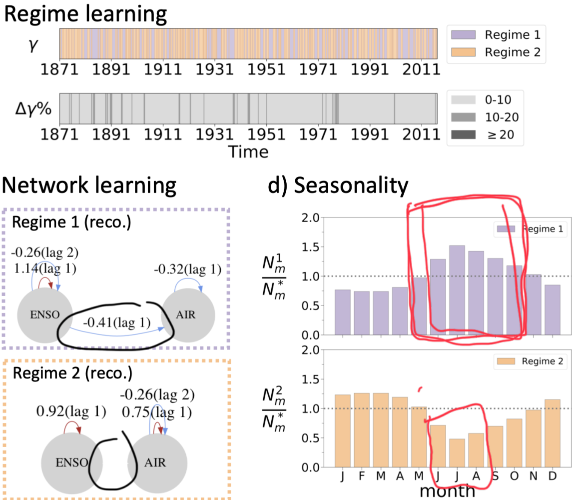
where the number of parameters are

$$N_{\text{para}} = (N_K - 1)N_C + \sum_{k=1}^{N_K}\sum_{j=1}^{N_X}|\mathcal{P}_k^j|. \tag{11}$$

Real data: the effect of El Niño Southern Oscillation on Indian rainfall



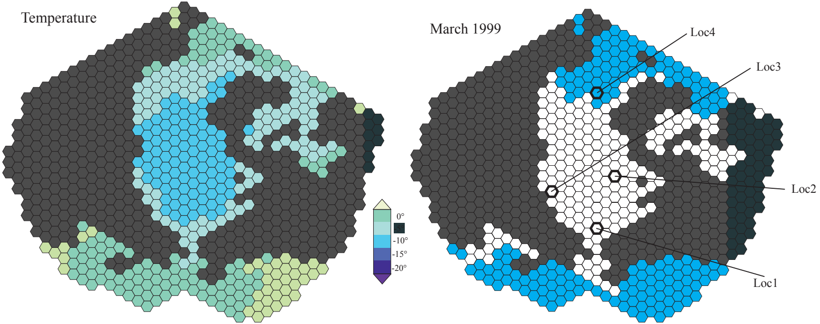
Real data: the effect of El Niño Southern Oscillation on Indian rainfall



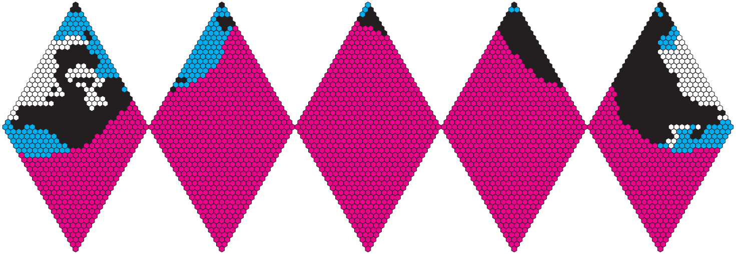
Thanks

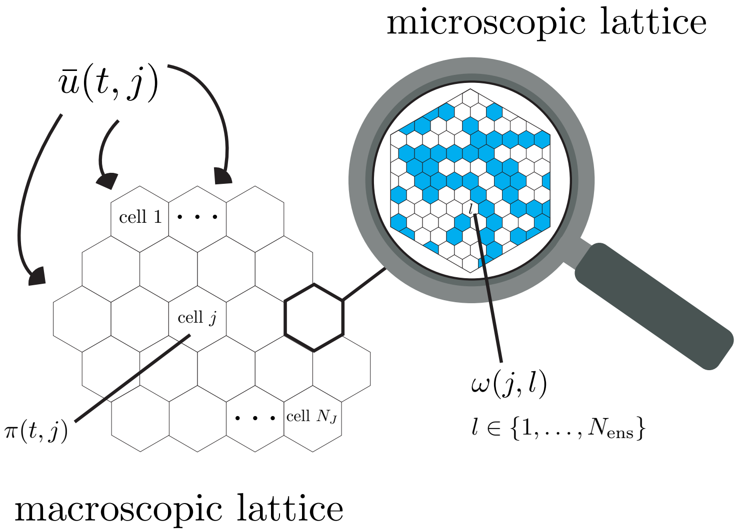
Non-stationary Non-homogenous Markov regression

Arctic sea ice coverage



Global grid




microscopic lattice
$\bar{u}(t, j)$
cell 1
. . .
cell j
. . .
cell N_J
$\pi(t, j)$
l
$\omega(j, l)$
$l \in \{1, \ldots, N_{\text{ens}}\}$
macroscopic lattice

Optimisation problem

$$\mathbf{L}(\Gamma(t,j), P(u(t,j)))$$
$$= \sum_{j=1}^{N_J}\sum_{t=1}^{N_T}\sum_{k=1}^{N_K}\gamma_k(t,j)\left\|\pi(t+1,j)^\top - \pi(t,j)^\top P^k(u(t,j))\right\|_2^2 \rightarrow \min_{\Gamma(t,j),P(u$$

with

$$P^k(u(t,j)) = P_0^k + \sum_{e=1}^{N_E} P_e^k u_e(t,j) \quad \forall\, k \in \{1,\dots,N_K\}.$$

$$\sum_{k=1}^{N_K}\gamma_k(t,j) = 1 \quad \text{for } j \in \{1,\dots,N_J\}, t \in \{1,\dots,N_T\},$$
$$\gamma_k(t,j) \geq 0 \quad \text{for } j \in \{1,\dots,N_J\}, t \in \{1,\dots,N_T\}, k \in \{1,\dots,N_K\}$$

References i