Statistical modelling in climate science

Nikola Jajcay

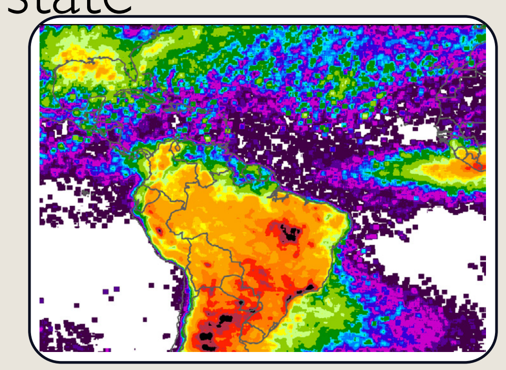
supervisor Milan Paluš

Seminář strojového učení a modelovaní
MFF UK

Introduction modelling in climate science

dynamical model

initial state



eqs.

$$-\eta\nabla^2\mathbf{u} + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p = \mathbf{F}$$

$$\nabla\cdot\mathbf{u} = 0$$

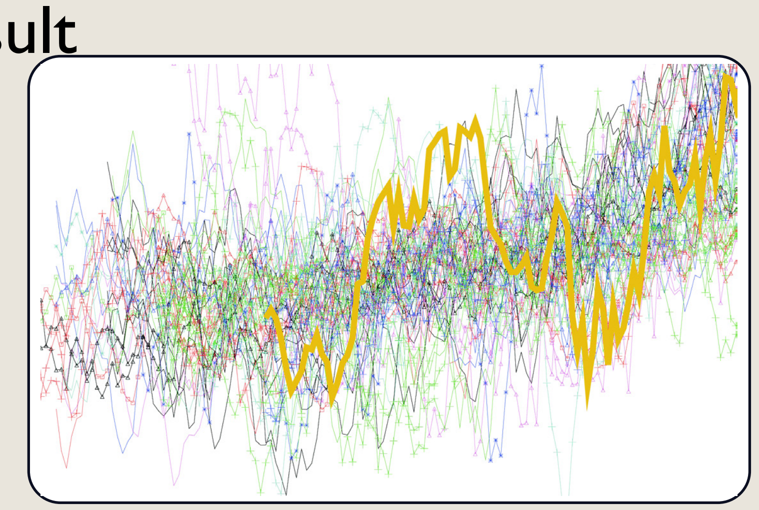
$$\frac{\partial}{\partial t}\left[\rho\left(e+\frac{\mathbf{u}^2}{2}\right)\right] + \nabla\cdot\left[\rho\left(e+\frac{\mathbf{u}^2}{2}\right)\mathbf{u}\right] =$$

$$\rho\dot{q} - \frac{\partial(up)}{\partial x} - \frac{\partial(vp)}{\partial y} - \frac{\partial(wp)}{\partial z} + \rho\mathbf{F}\cdot\mathbf{u}$$

$$\dots$$

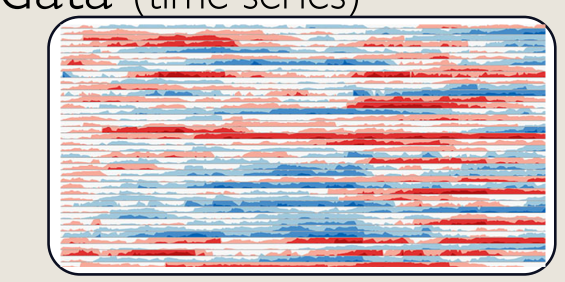
$$\Downarrow\Downarrow \int \dots \mathrm{d}t$$

result



statistical model

past data (time series)



$\Downarrow$ train \ estimate

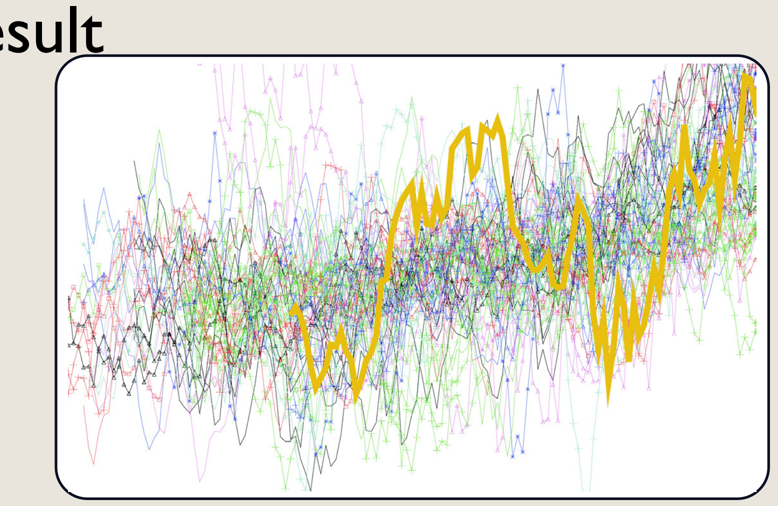
stat. model

$$\mathrm{d}x_i = \left(\mathbf{x}^T\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\mathbf{x} + c_i\right)\mathrm{d}t + \mathrm{d}r_i$$

$$\mathrm{d}r^{(L)} = \mathbf{b}_i^{(L)}\left[\mathbf{x}, \mathbf{r}^{(0)}, \dots, \mathbf{r}^{(L)}\right]\mathrm{d}t + \mathrm{d}r_i^{(L+1)}$$

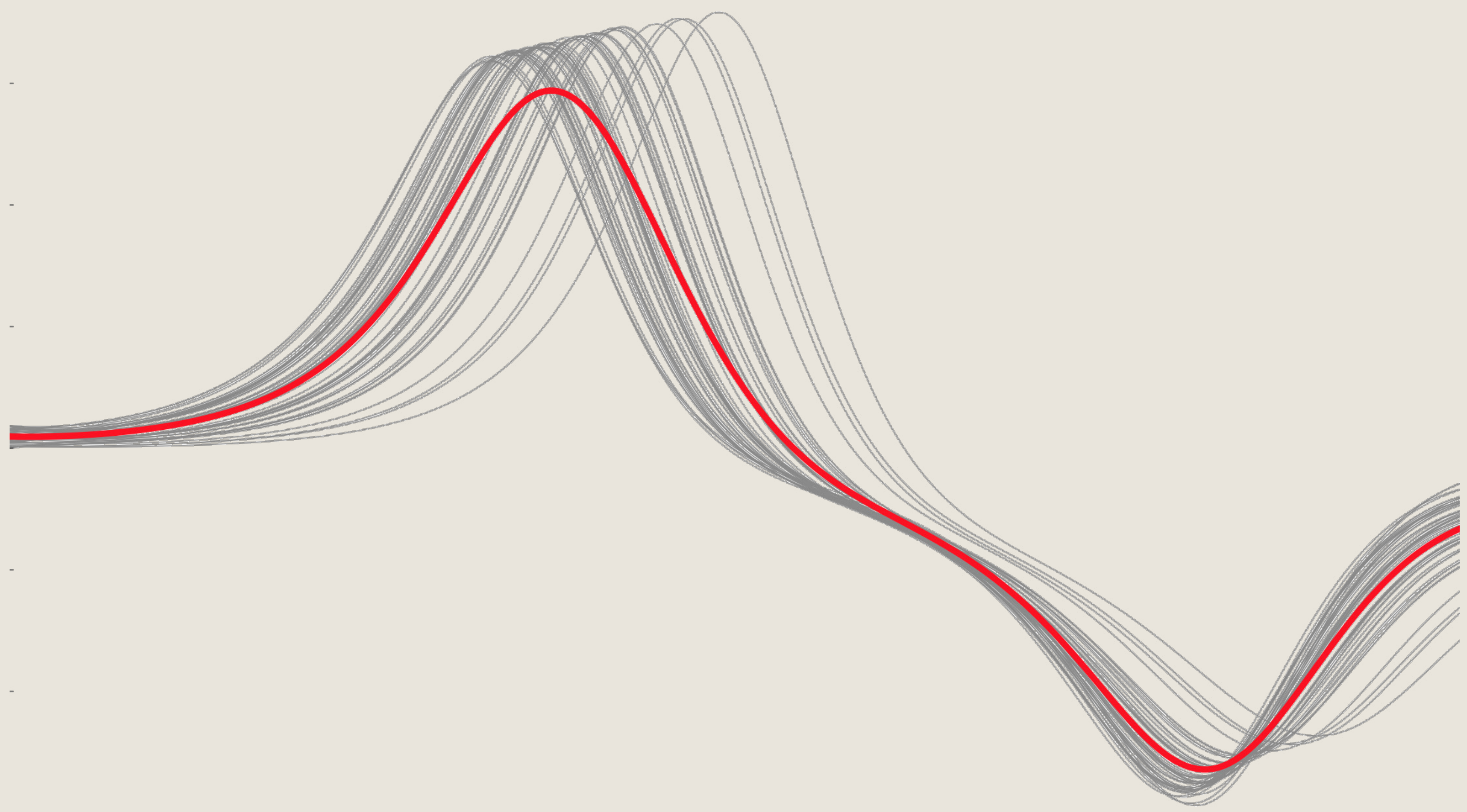
$$\Downarrow\Downarrow \int \dots \mathrm{d}t$$

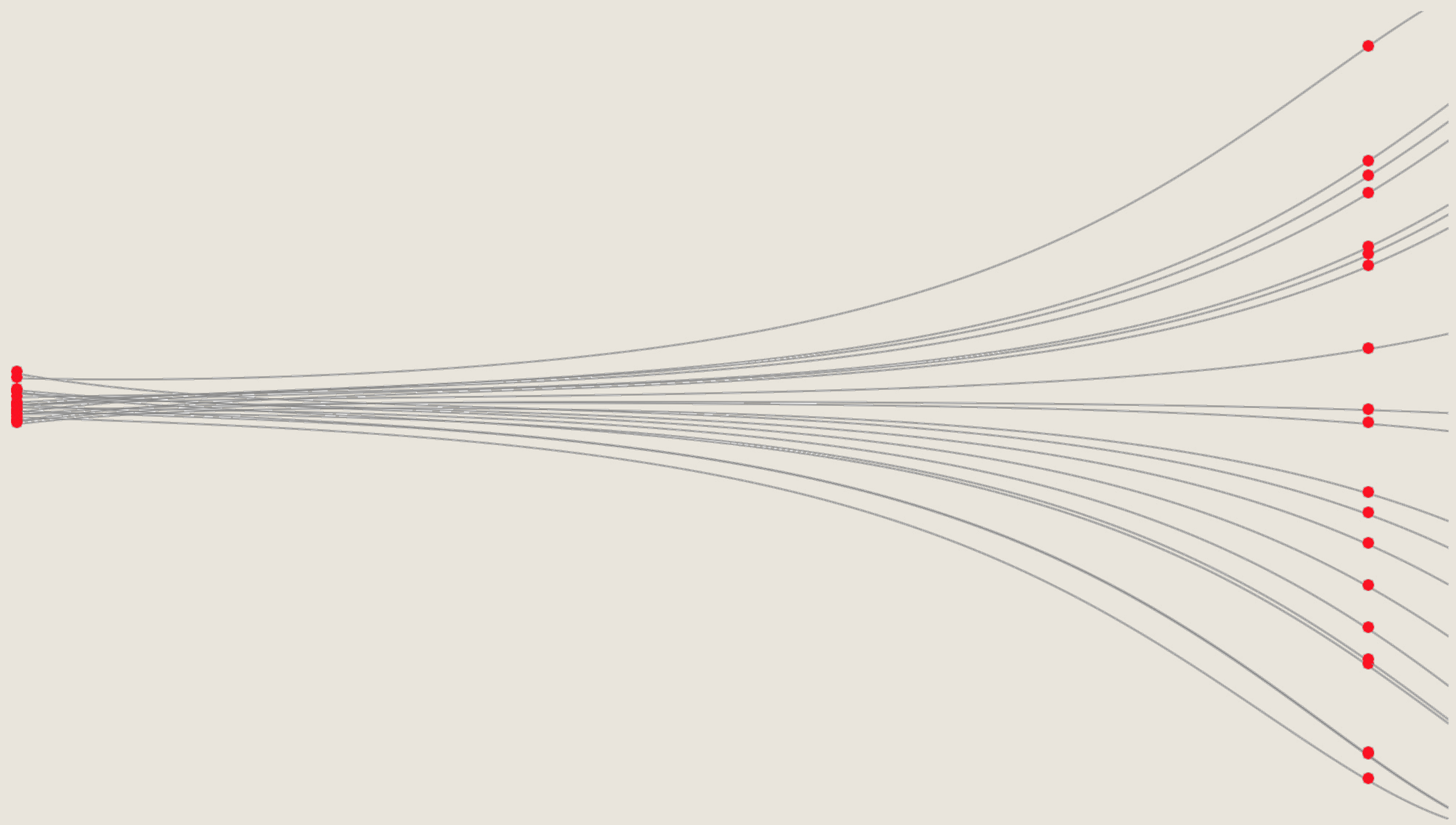
result



Introduction dynamical models

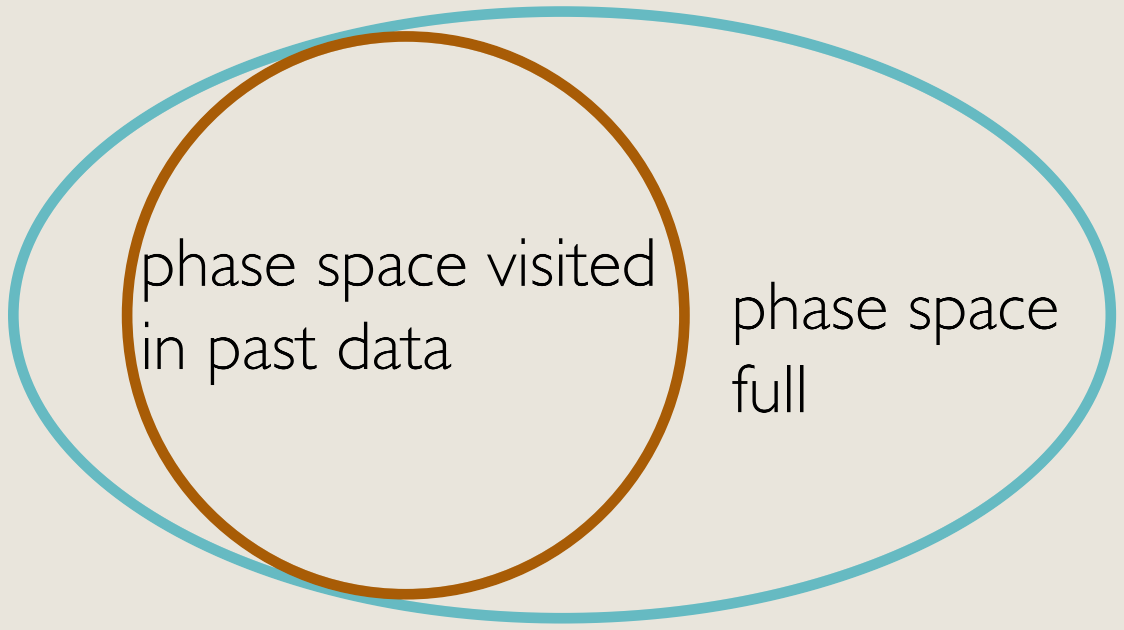
- discretized partial differential equations + current state of the climate (initial cond.)
- general circulation models (GCMs)
- used in numerical weather prediction, reanalysis data-sets and future climate intercomparisons
- uncertainties: initial errors and model errors





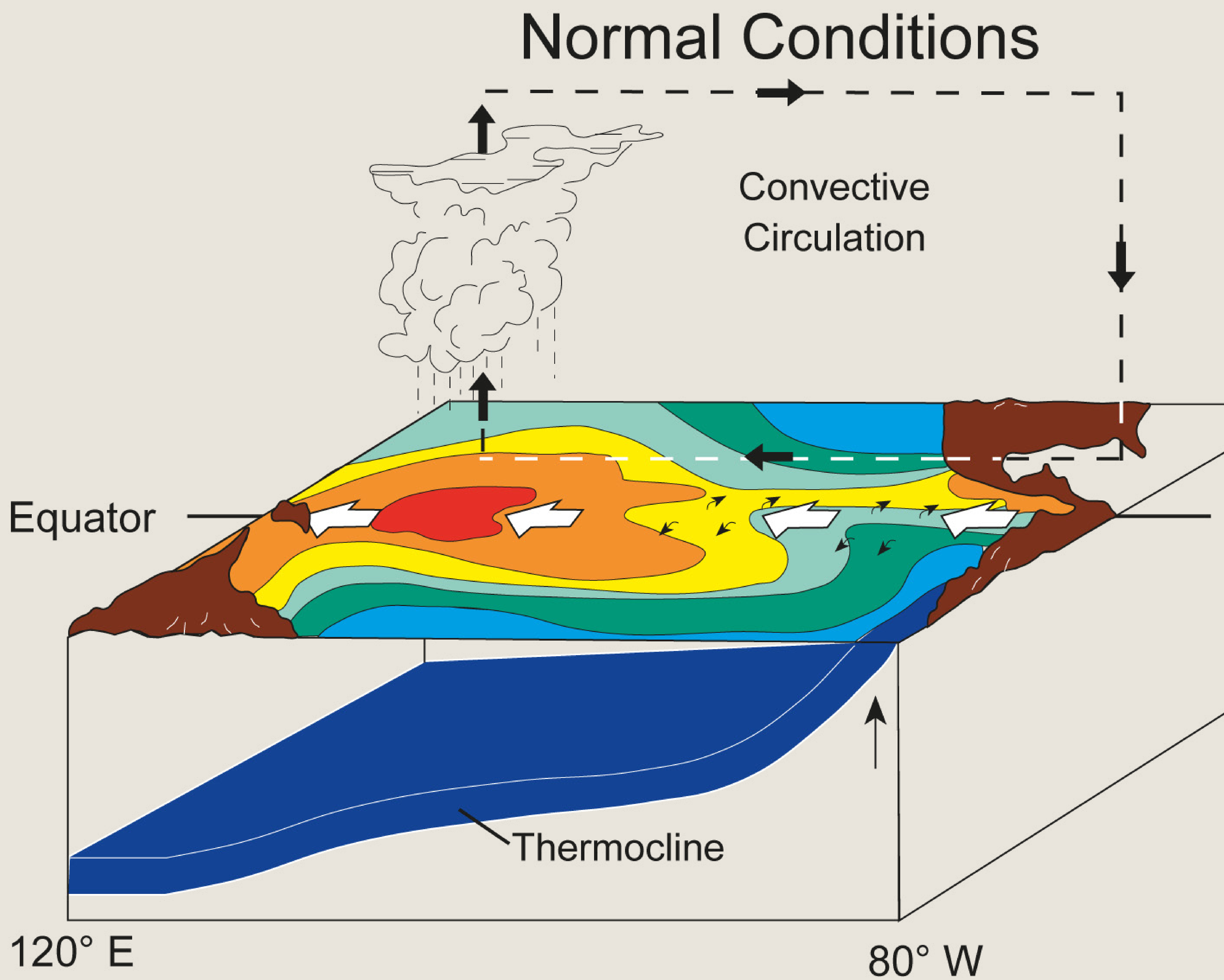
Introduction statistical models

- motivation: forecasting, scaling down the complexity
- NOT based on physical mechanisms underlying the dynamics, but derived from past weather patterns
- inverse stochastic models
- model is designed, its parameters estimated / trained using past weather data and stochastically integrated
- uncertainties: which variables and non-stationarity
  - temperature?
  - atm. pressure?
  - sea-surface temperature?
  - cloudiness?
  - latent heat flux?



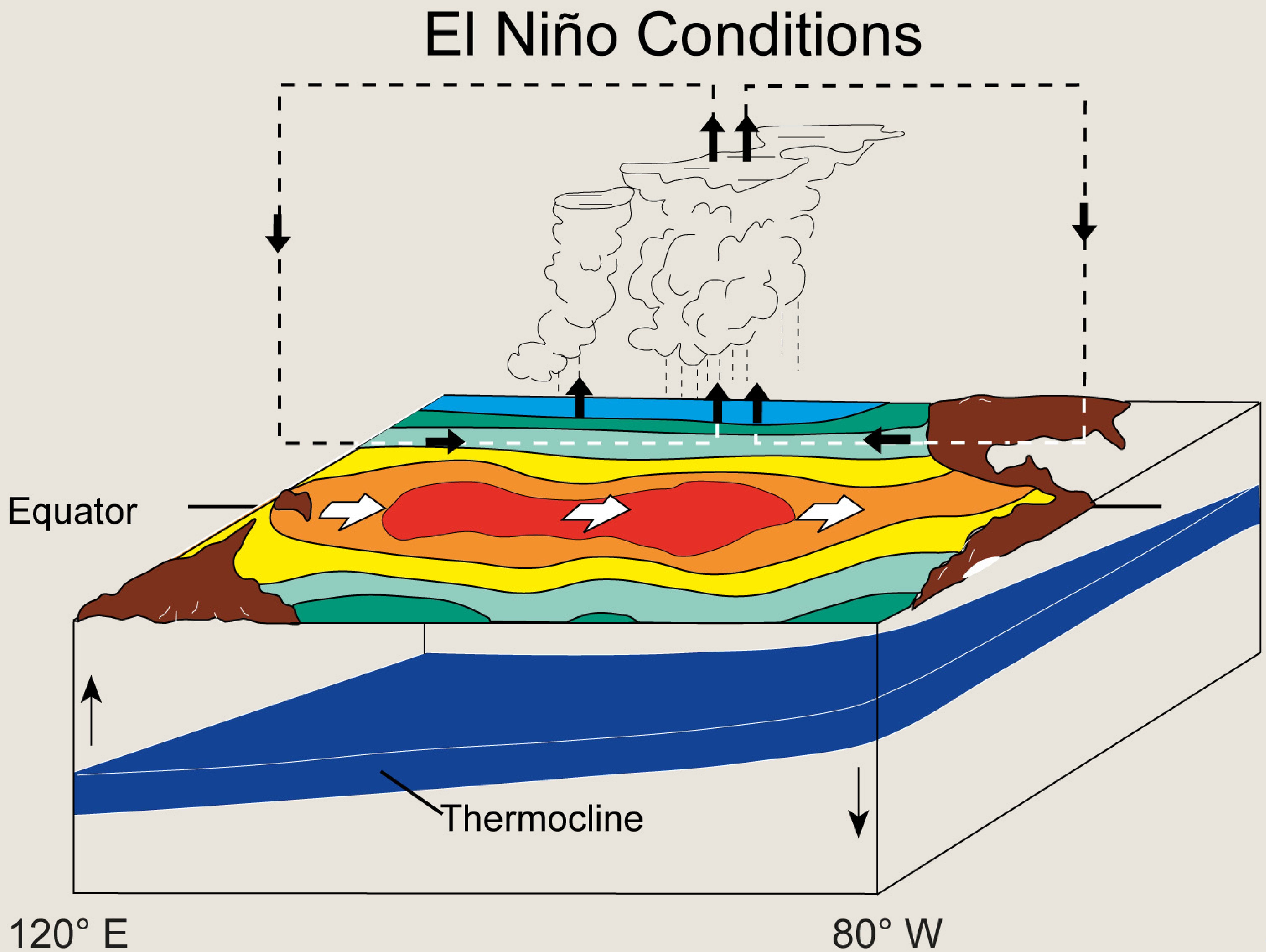
ENSO neutral

- strong interannual signal with great economic and so-cietal impact



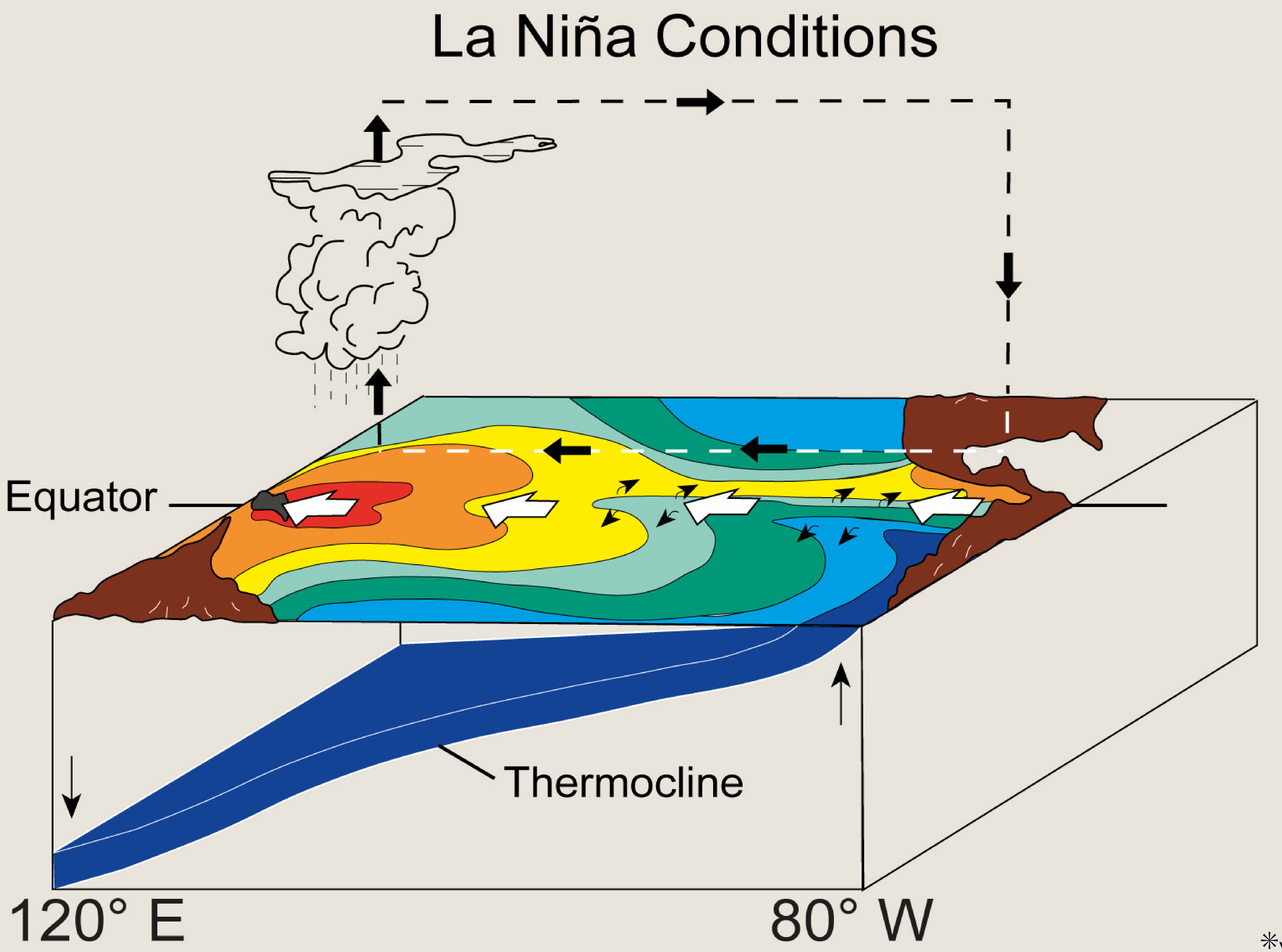
*wikipedia.org

ENSO positive



*wikipedia.org

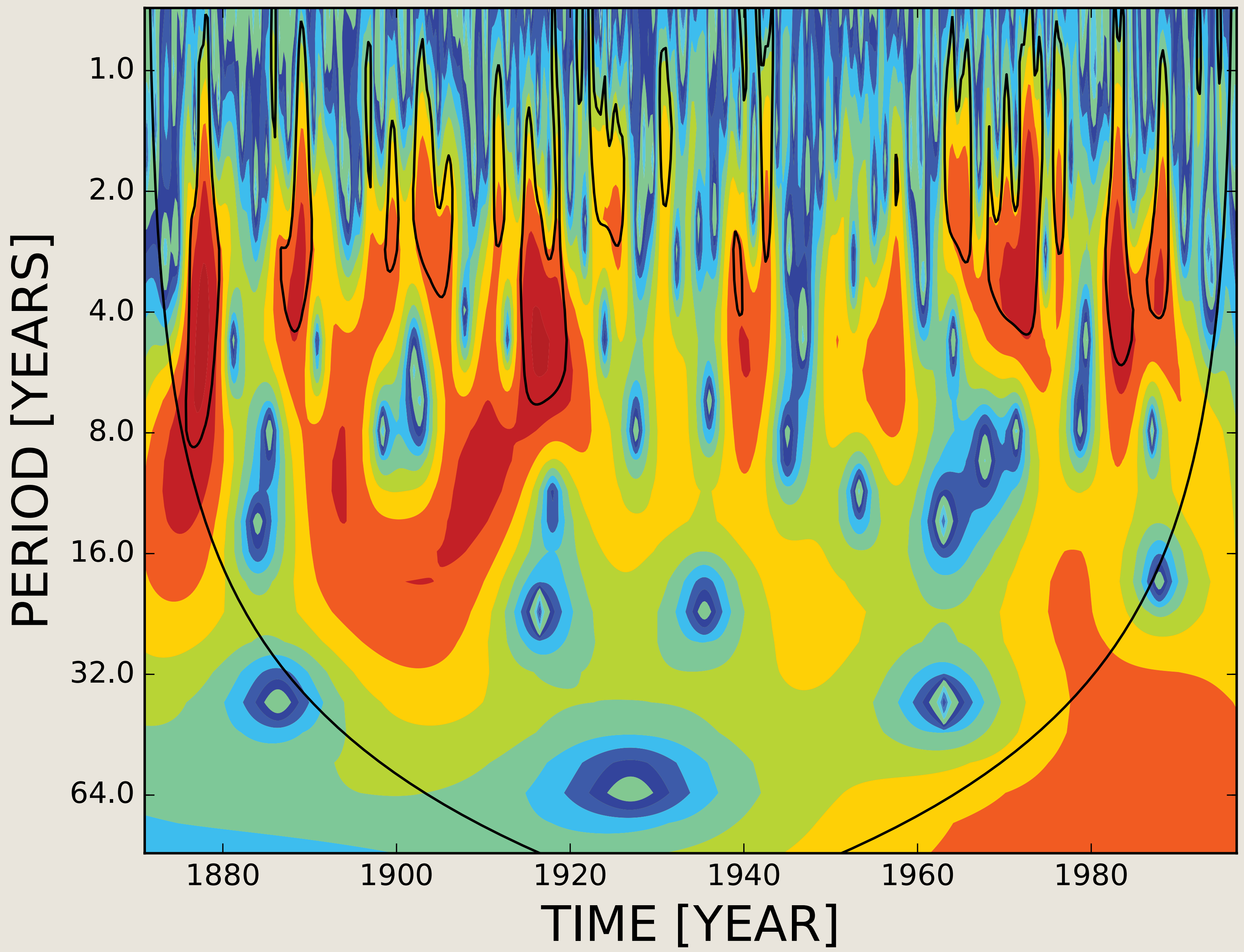
ENSO negative



*wikipedia.org

ENSO overview

- naturally oscilates between phases without a distinct period



- reasons why are still largely unknown
- positive phase characterised by a larger magnitude than negative phase -- nonlinear interactions

Statistical model inverse nonlinear model

- evolution of anomalies as
  $\dot{\mathbf{x}} = \mathbf{L}\mathbf{x} + \mathbf{N}(\mathbf{x})$
- linear inverse models by assuming linear form
  $\mathbf{N}(\mathbf{x})\mathrm{d}x \approx \mathbf{T}\mathbf{x}\mathrm{d}t + \mathrm{d}\mathbf{r}^{(0)}$
- describes linear feedback of hidden processes
- assume polynomial form
  $N_i(\mathbf{x})\mathrm{d}x \approx (\mathbf{x}^T\mathbf{A}_i\mathbf{x} + \mathbf{t}_i\mathbf{x} + c_i^{(0)})\mathrm{d}t + \mathrm{d}r_i^{(0)}$
  $\mathbf{b}_i^{(0)} = \mathbf{l}_i + \mathbf{t}_i,\ \mathbf{B}^{(0)} = \mathbf{L} + \mathbf{T}$
- so that the main level of our model is
  $\mathrm{d}x_i = \left(\mathbf{x}^T\mathbf{A}_i\mathbf{x} + \mathbf{b}_i^{(0)} + c_i^{(0)}\right)\mathrm{d}t + \mathrm{d}r_i^{(0)}$

Statistical model multilevel models

- stochastic forcing still involves serial correlations and might also depend on modelled process
- additional levels included to express the known time increments as linear function of extended state vector

$$\mathrm{d}r_i^{(0)} = \mathbf{b}_i^{(1)}\left[\mathbf{x}, \mathbf{r}^{(0)}\right]\mathrm{d}t + r_i^{(1)}\mathrm{d}t$$
$$\dots$$
$$\mathrm{d}r_i^{(L)} = \mathbf{b}_i^{(L+1)}\left[\mathbf{x}, \mathbf{r}^{(0)}, \dots, \mathbf{r}^{(L)}\right]\mathrm{d}t + r_i^{(L+1)}\mathrm{d}t$$

Statistical model ENSO model

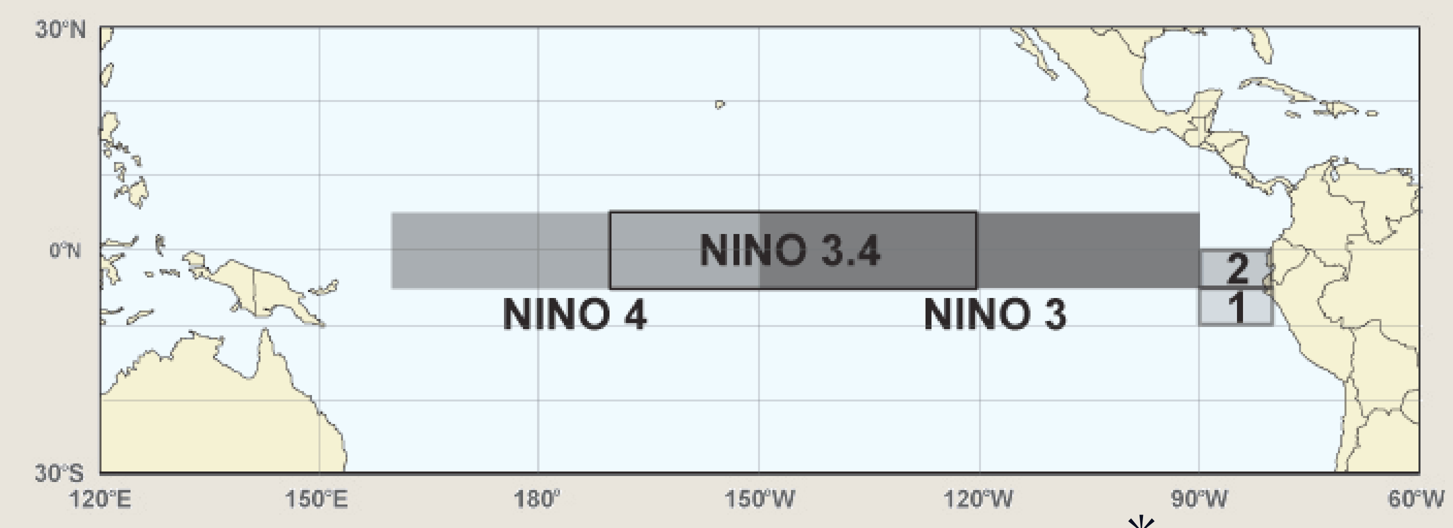
- it is known that extreme ENSO events tend to occur in boreal winter, we include seasonality as

$$\mathbf{B}^{(0)} = \mathbf{B}_0 + \mathbf{B}_s \sin(2\pi t/T) + \mathbf{B}_c \cos(2\pi t/T)$$
$$\mathbf{c}^{(0)} = \mathbf{c}_0 + \mathbf{c}_s \sin(2\pi t/T) + \mathbf{c}_c \cos(2\pi t/T)$$

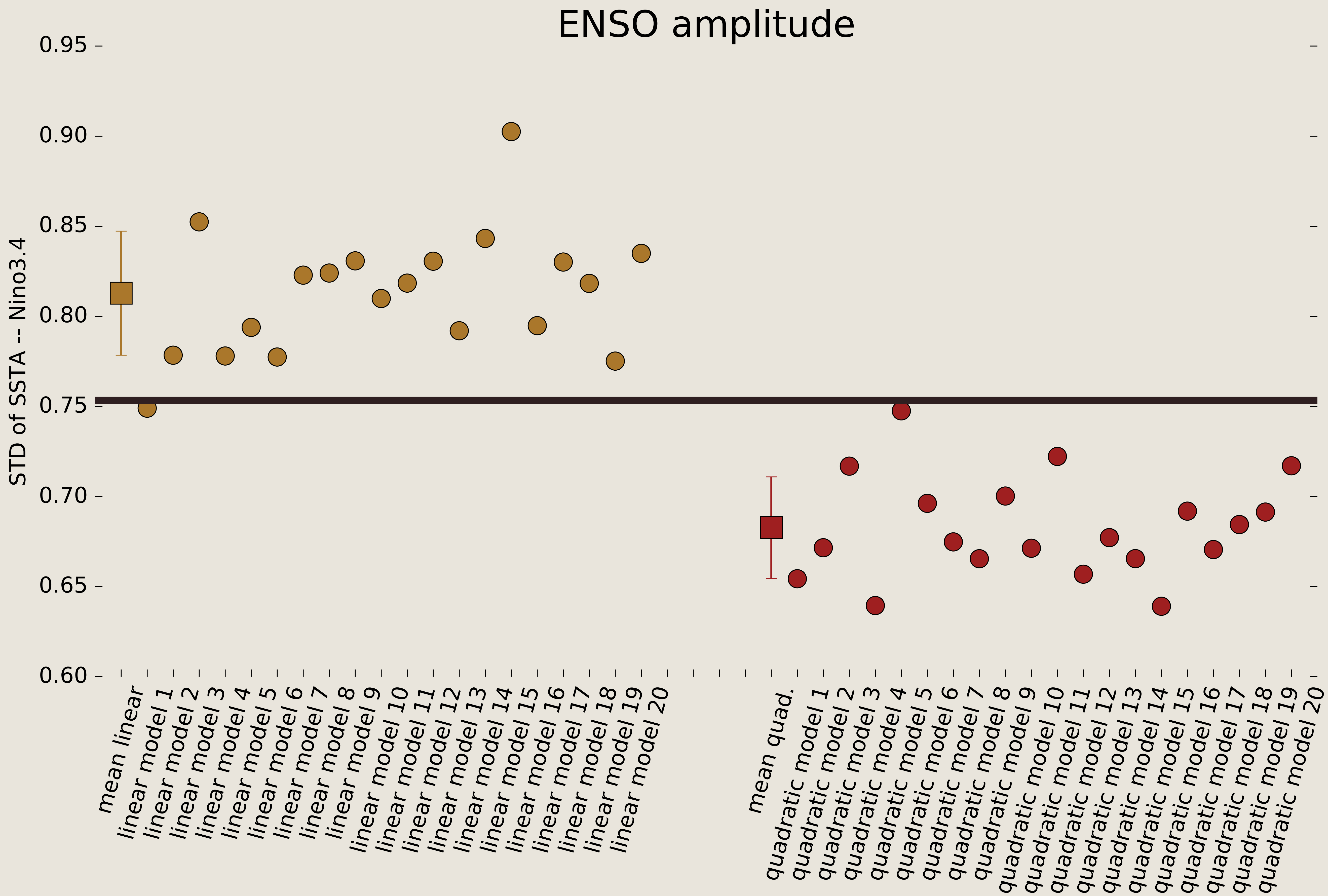
- model is estimated in the leading EOF space of Pacific sea surface temperature anomalies
- optimal number of state vector variables and degree of nonlinearity has to be assessed by cross-validation

Results basic ENSO metrics

- *NINO3.4* index - spatial average
- **amplitude** - STD of NINO3.4

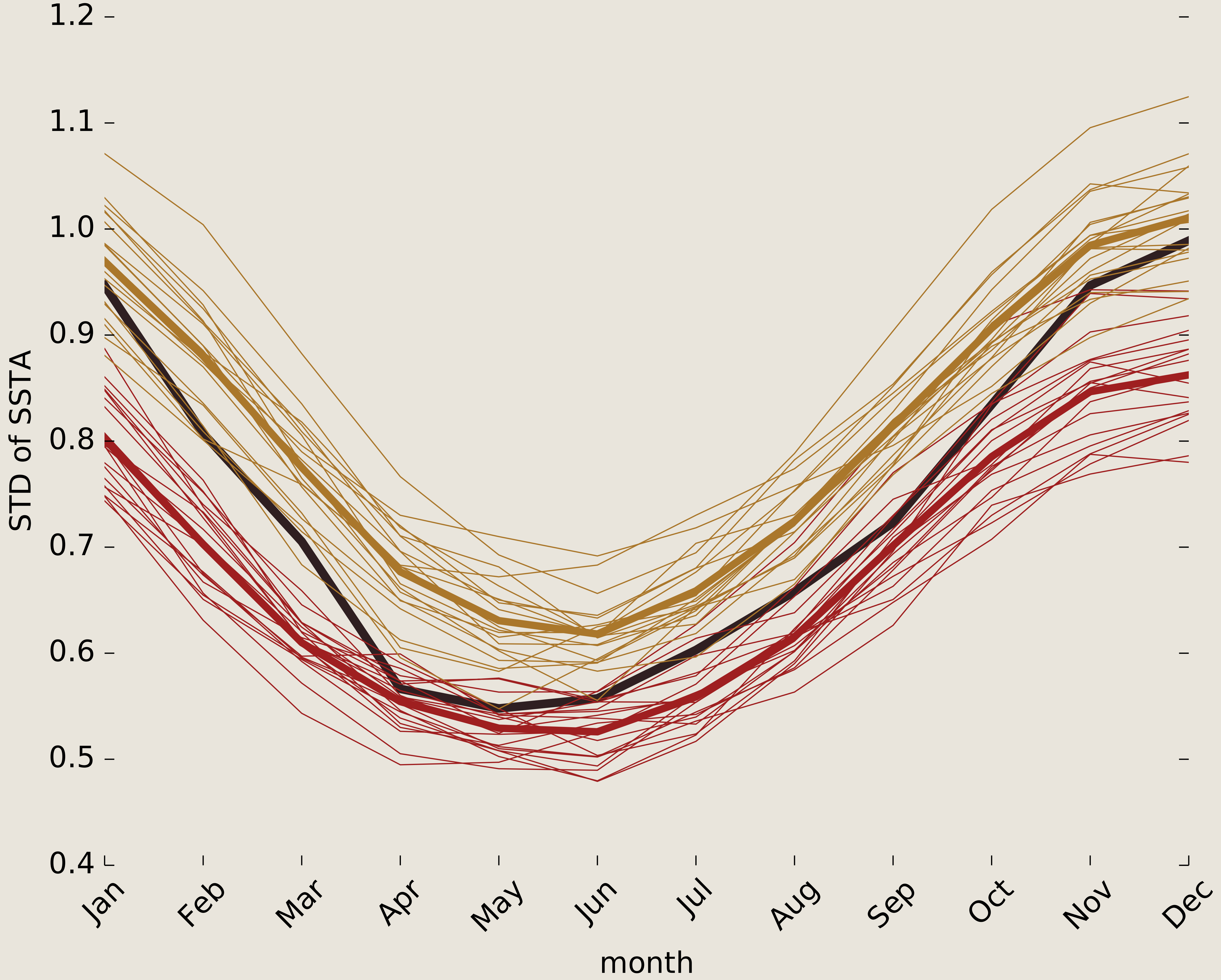


*noaa.gov



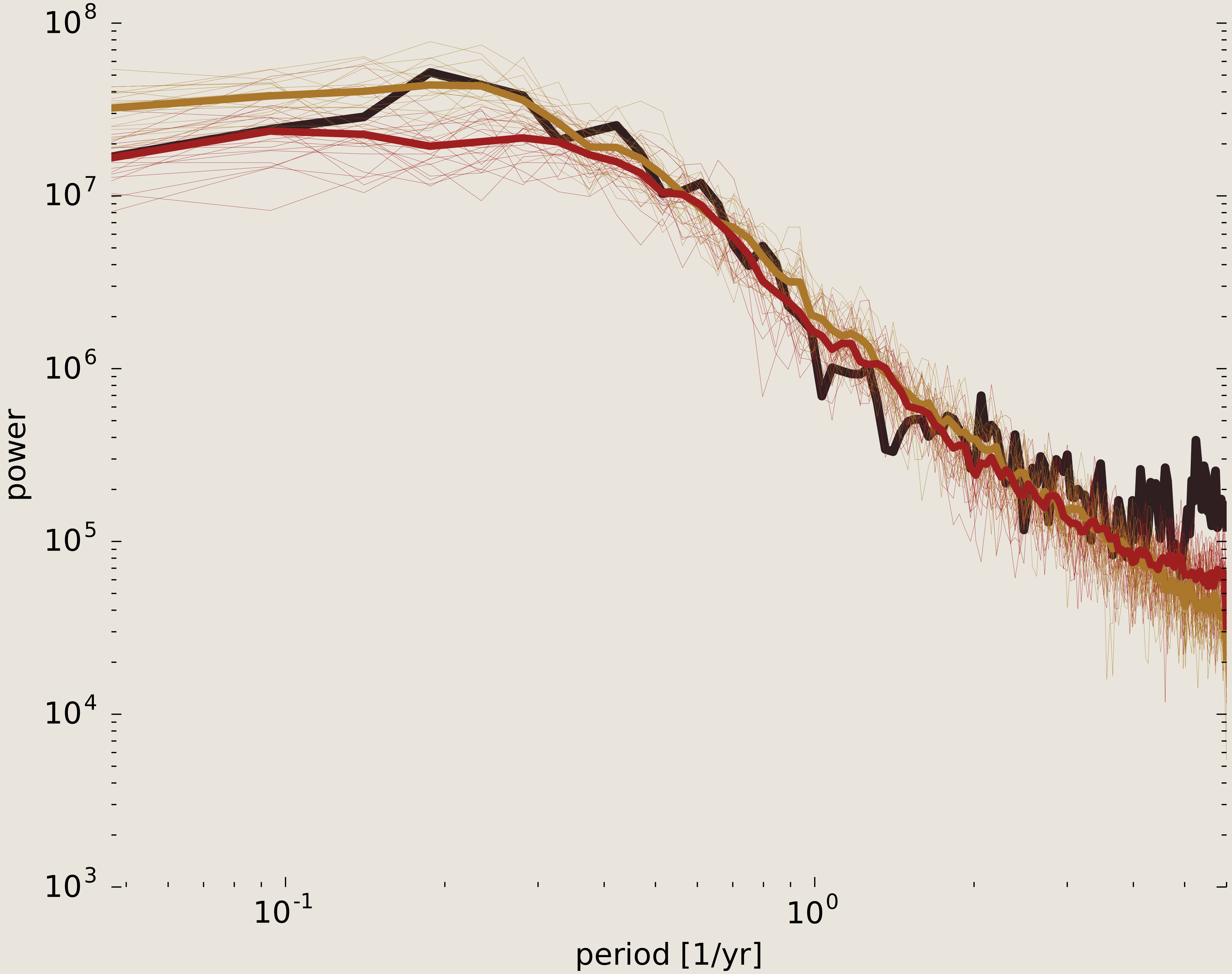
Results basic ENSO metrics

- **seasonality** - monthly STD of NINO3.4



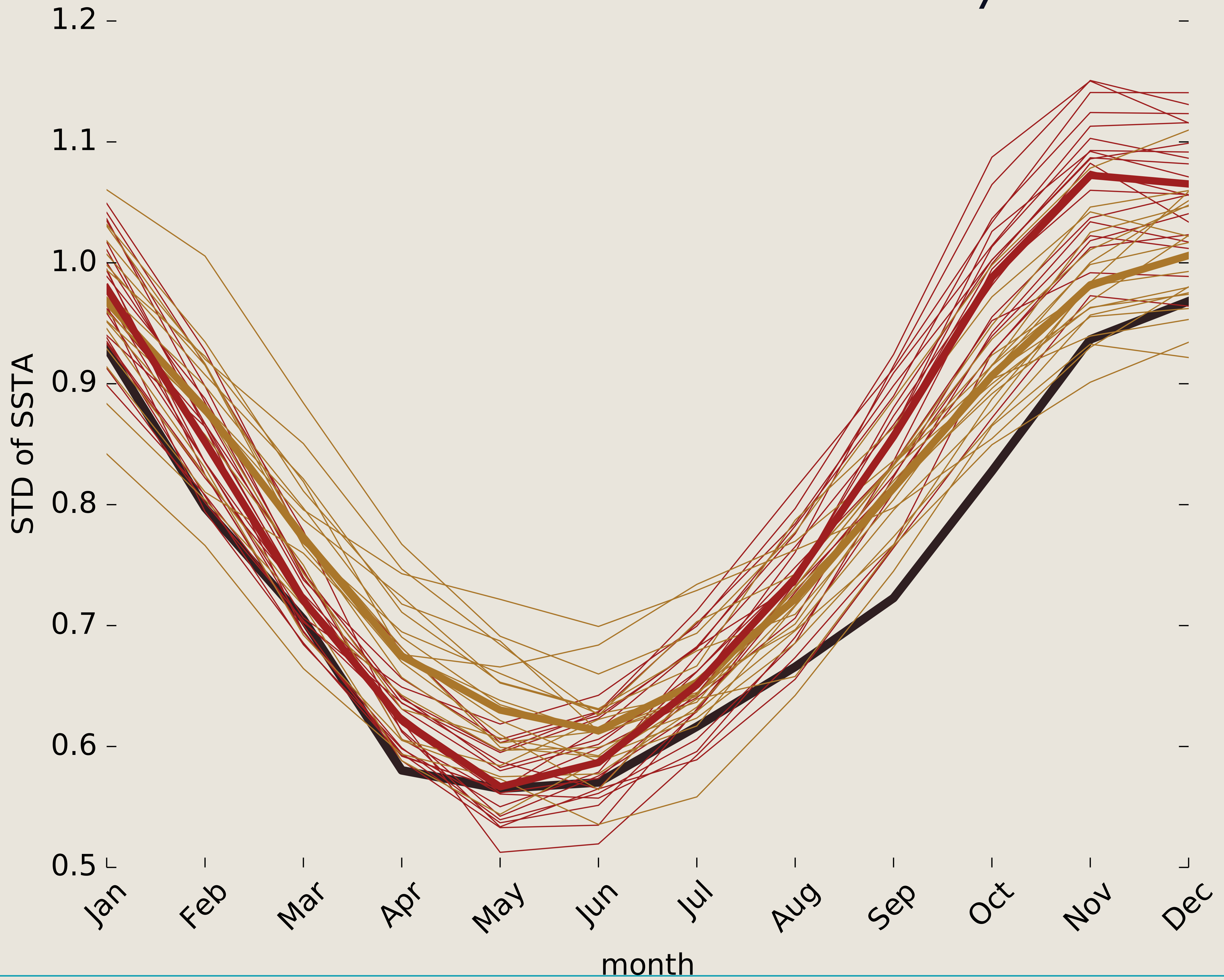
Results basic ENSO metrics

- **spectrum** - estimated using Welch method



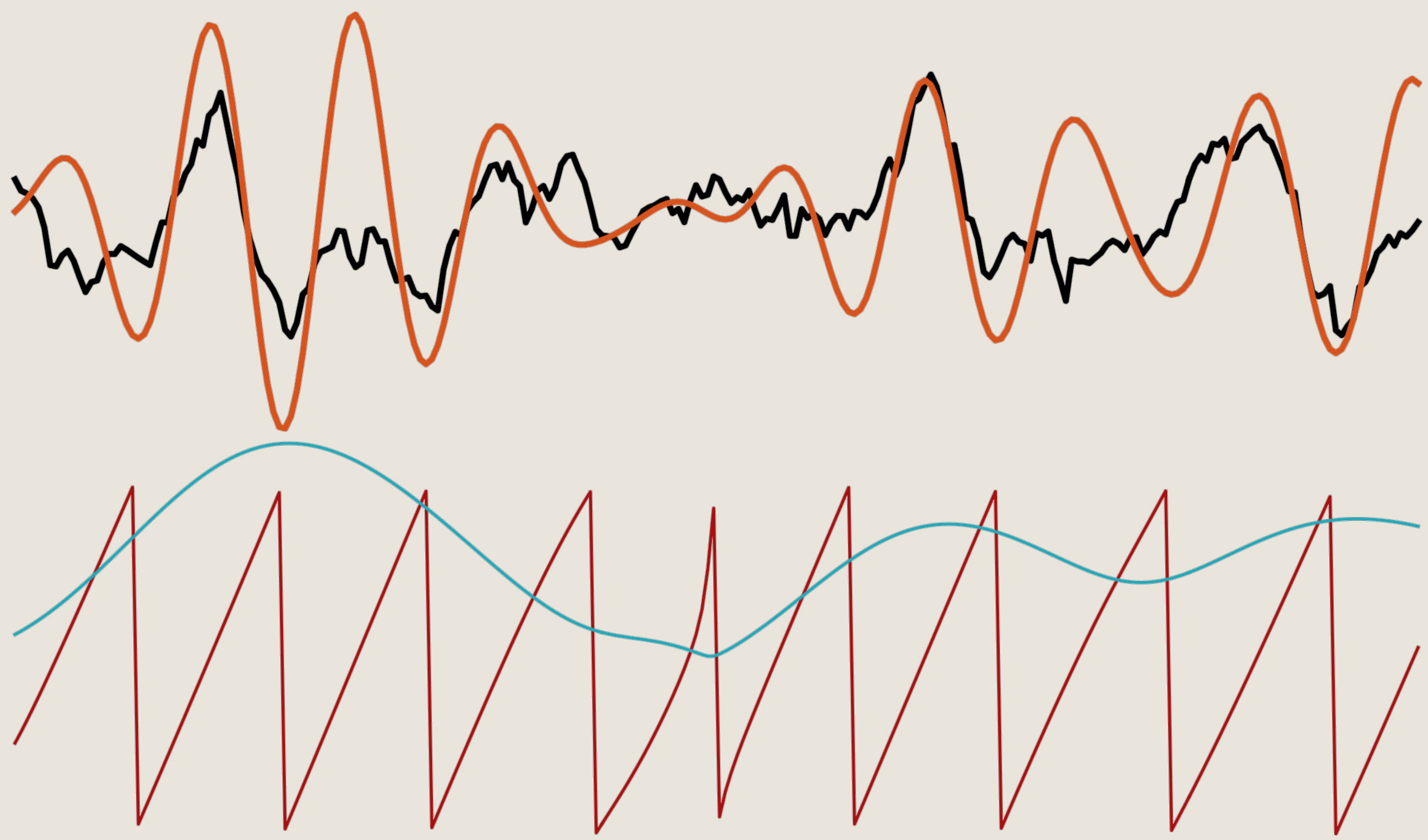
Noise parametrization seasonality

- even multi-level model exhibit serial correlations and seasonal dependence
- noise is conditioned on system's state



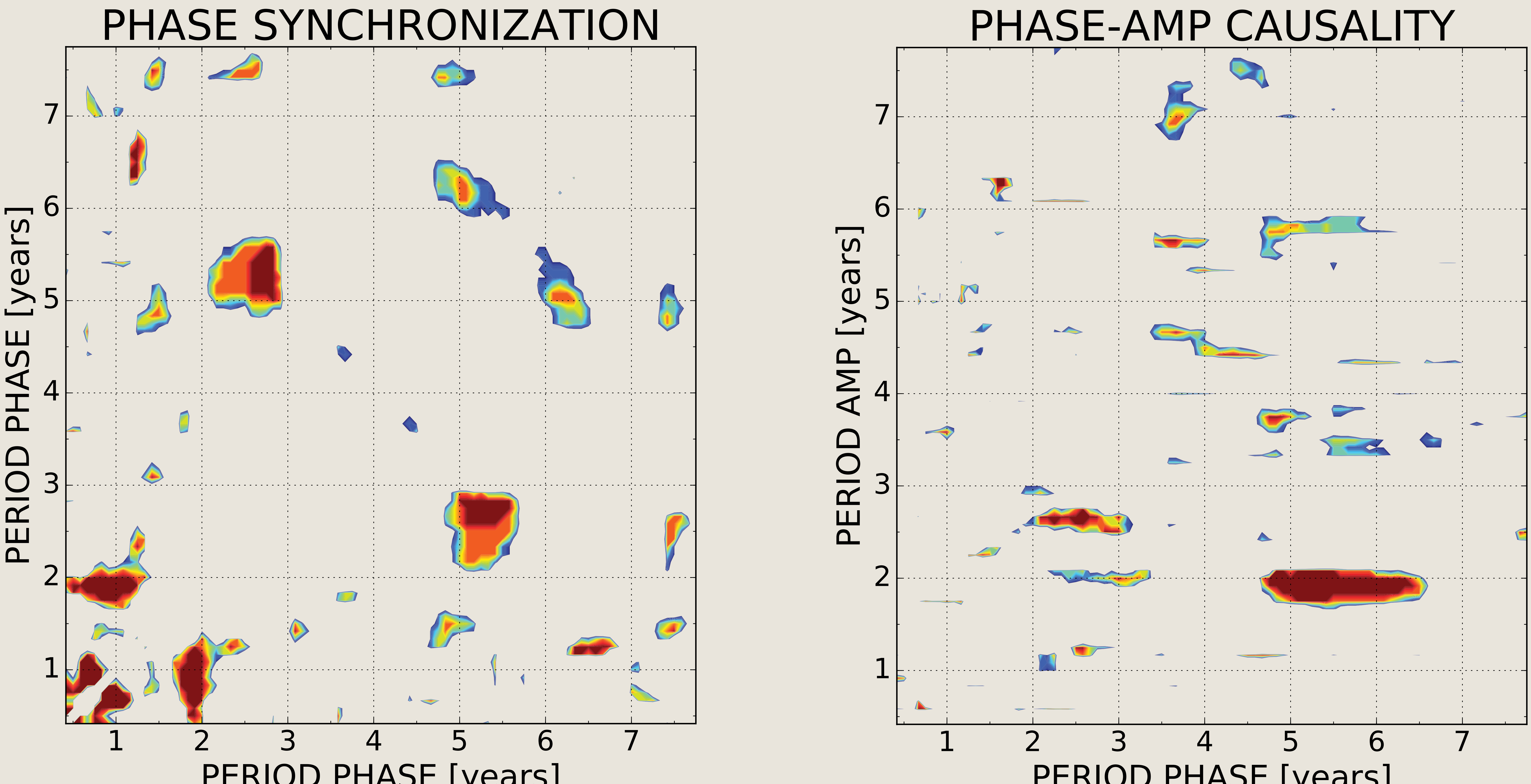
Synchronization and causality concept

- causal relations or information flow between various scales in the same variable / process
- using wavelet transform to infer instantaneous phase and amplitude of the signal with selected period



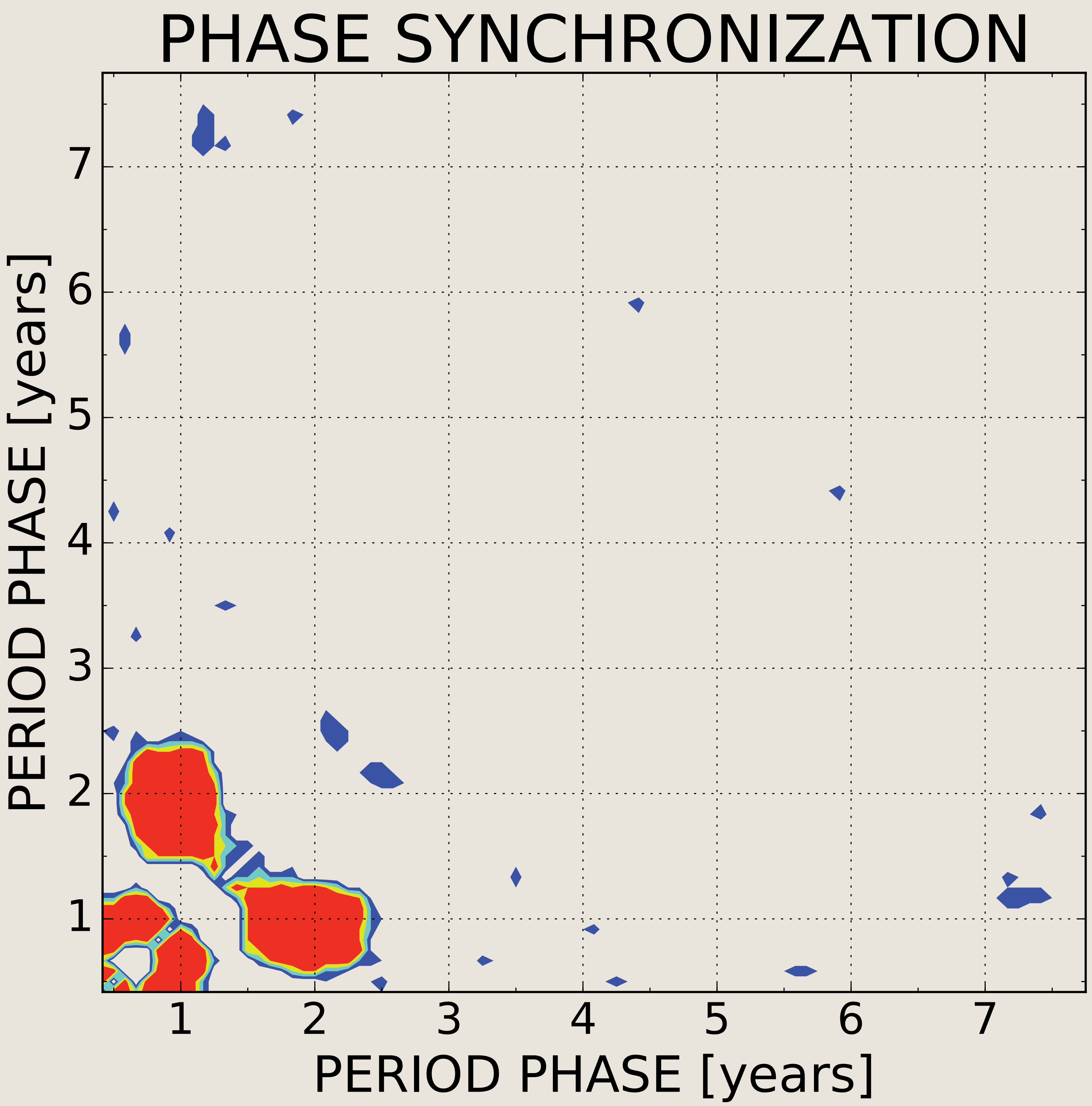
Synchronization and causality data

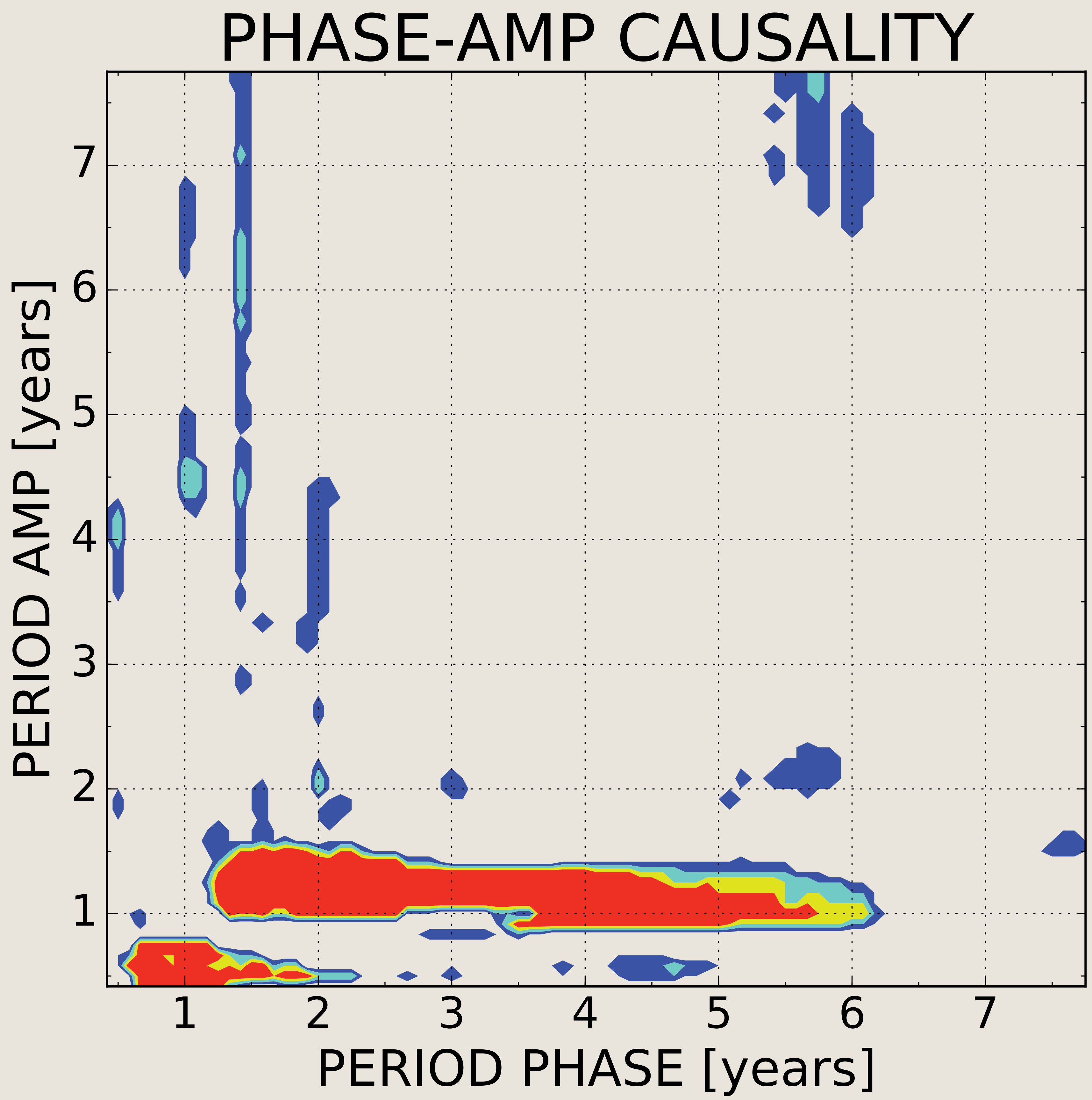
- using (conditional) mutual information to infer synchronization and causality measures



Synchronization and causality model

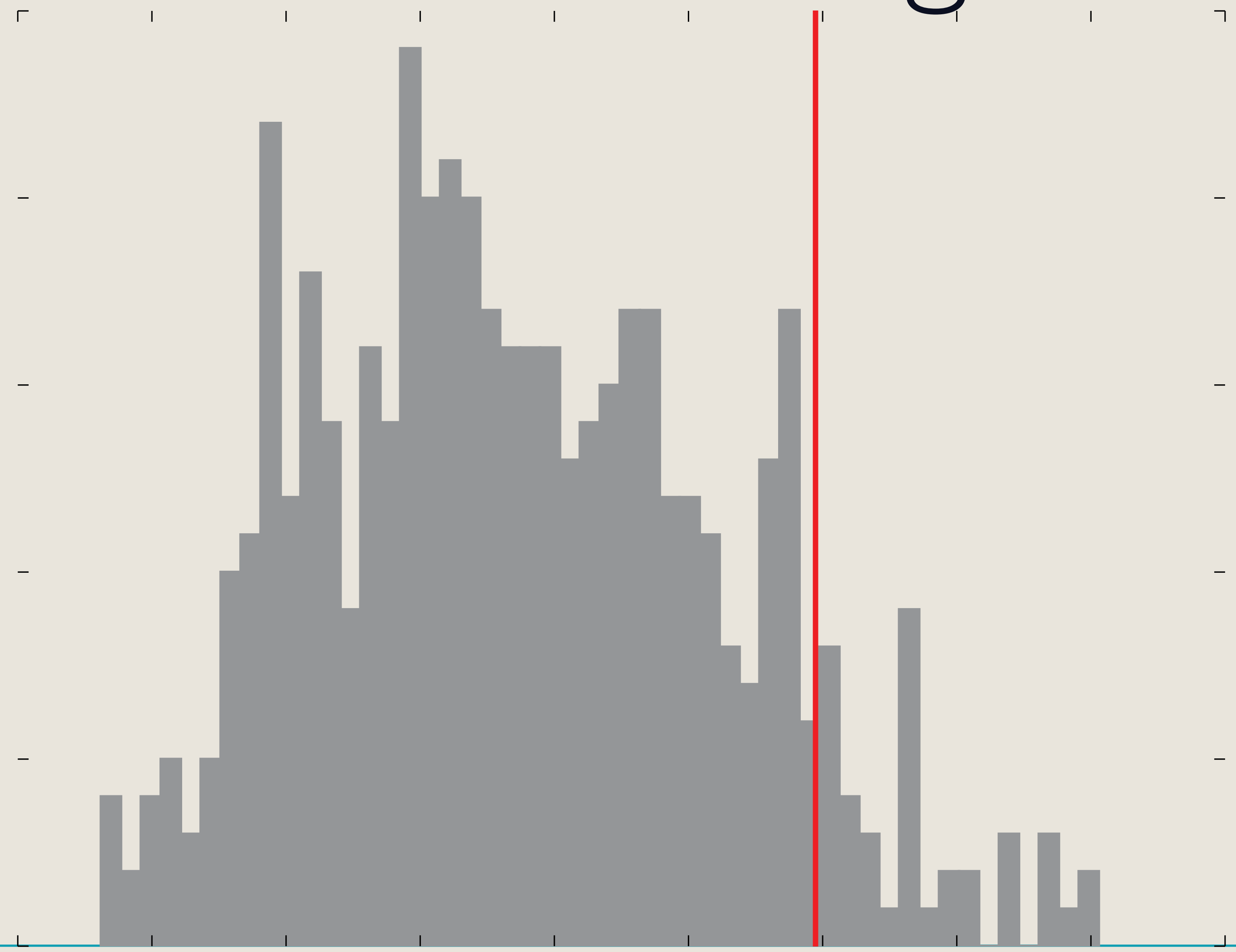
- simulate synchronization and causality in modelled time series to uncover the mechanisms





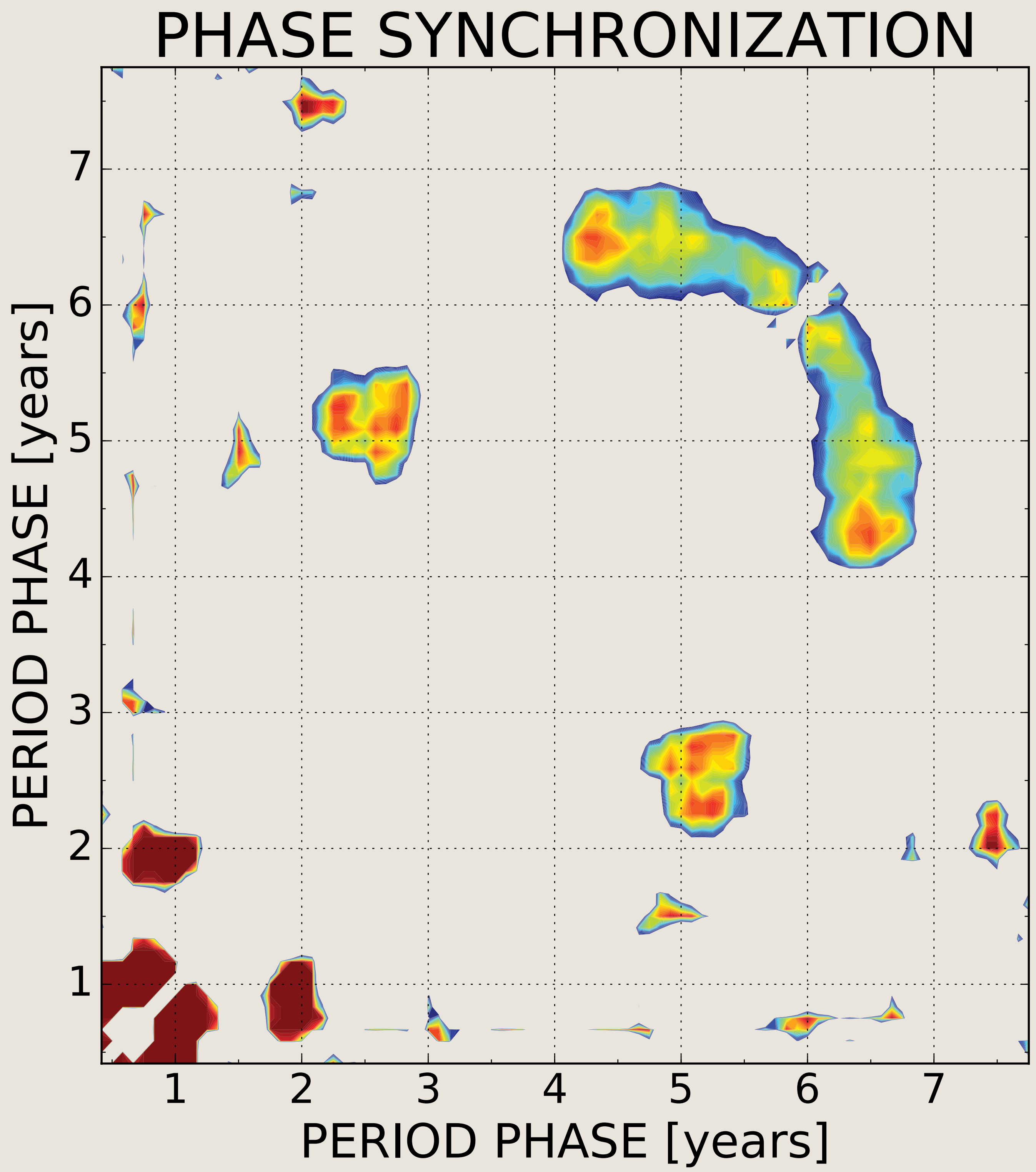
Surrogate data modelling with statistical model

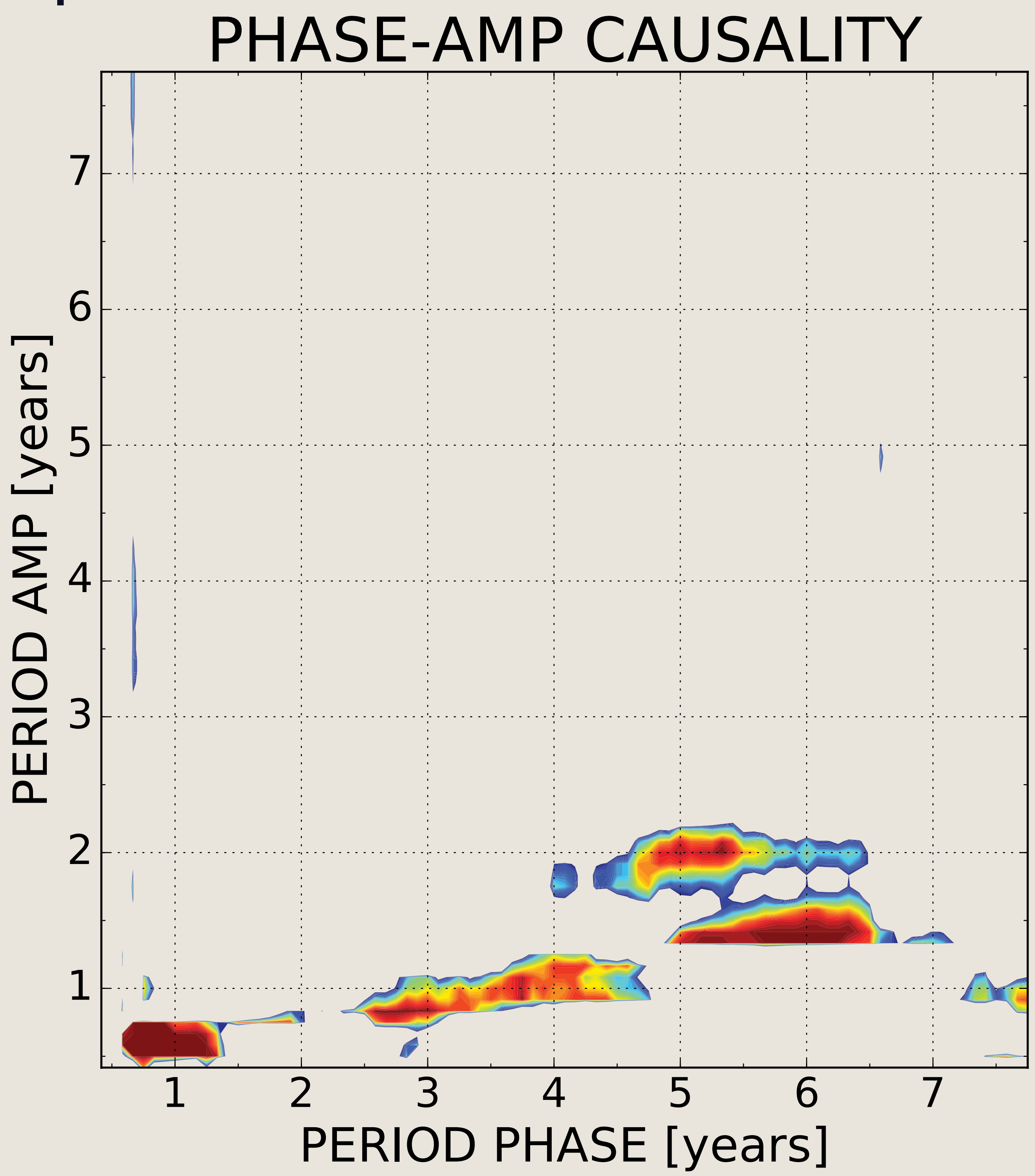
- method to generate synthetic that preserve some of the properties of the original data, while omitting the others
- use to test statistical significance by contradiction
- pose a null hypothesis and then generate an ensemble of surrogate time series using MC methods



Surrogate data modelling with statistical model

- more sophisticated null hypothesis: exploit the options of data-based model -- create surrogate ensemble statistical model with low complexity
- our case: linear, no seasonal dependence, white noise





Conclusions and outlook

- statistical models for scaling down the complexity
- modelling linear and non-linear interactions
- various noise parametrizations
- possible usage as models for generating ensembles of surrogate data for statistical testing
- two paths: focusing on a model itself (various settings, multi variables, etc..) or connection with dynamical models (e.g. for parametrization of sub-grid phenomena etc)

Dept. of Nonlinear Dynamics and Complex Systems, Institute of Computer Science, CAS
Dept. of Atmospheric Physics, Faculty of Mathematics and Physics, Charles University in Prague

Thanks for your attention!

Nikola Jajcay

jajcay@cs.cas.cz

github.com/jajcayn

Seminář strojového učení a modelovaní

MFF UK