On associating significance levels with temporal changes in empirical orthogonal function analysis: a case study for ENSO

Gábor Drótos

IFISC (UIB-CSIC), Palma de Mallorca, Spain,
and
MTA–ELTE Theoretical Physics Research Group, Eötvös University, Budapest, Hungary

EGU2020

funded by the European Social Fund

Outline

Outline

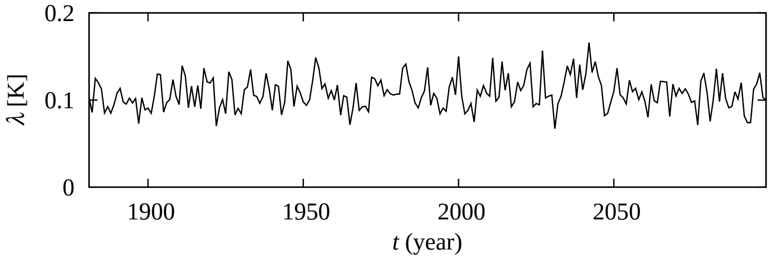
Temporal changes of what?

- The variability in a field $T(\mathbf{x};t)$ is described correctly by an ensemble of different realizations $T_r(\mathbf{x};t)$ at any given fixed time instant t.
- This variability can be decomposed to EOFs: $T_r(\mathbf{x}) = \sum_k \mathtt{PC}_r^{(k)}\mathtt{EOF}^{(k)}(\mathbf{x})$.
- An amplitude can be associated with each EOF: $\lambda^{(k)}$.
- The EOFs and the amplitudes can be different in different "instants of time" (years): e.g., we have $\lambda^{(k)}(t)$ (cf. EGU2020-2894; EGU2020-7527; EGU2020-12061; Maher et al., Geophys. Res. Lett. 45, 11,390, 2018; Haszpra et al., J. Climate 33, 3107, 2020; Haszpra et al., Earth System Dynam. 11, 267, 2020).
- How do we detect this time dependence if we can only estimate these quantities from a finite number of ensemble members?
- The methodology is the same for any k, we thus omit the index k.

Outline

Example: $\lambda^{(1)}$ of the Pacific SST field

MPI-GE historical+RCP8.5



Do we have a trend here?

Linear fit:

$$(2.08 \pm 2.06)^{-5}\ [\mathrm{K/yr}]$$

$\Longrightarrow$ Yes, we have?

Concept for testing

- A linear fit provides with an upper bound on the significance level for a temporal change, since
  - either the dependence is really linear (unlikely),
  - or not, so that the numerical result is meaningless,
    but then λ cannot be constant: there is a temporal change!
    (Thanks to T. Bodai.)
- However: the errors $\Delta\lambda$ are heteroskedastic and autocorrelated
  $\Longrightarrow$ the traditional significance level is incorrect
  $\Longrightarrow$ generalized least squares fit must be used
  based on the variance-covariance matrix $\sigma_{ij}^2 = \langle \Delta\lambda(t_i) \Delta\lambda(t_j) \rangle$.

Methodology

- Estimating the variance-covariance matrix of the errors: between years t_i and t_j:

$$\sigma_{ij}^2 = \langle \Delta\lambda(t_i)\Delta\lambda(t_j)\rangle = \frac{1}{N}\langle \mathtt{PC}(t_i)\mathtt{PC}(t_j)\rangle,$$

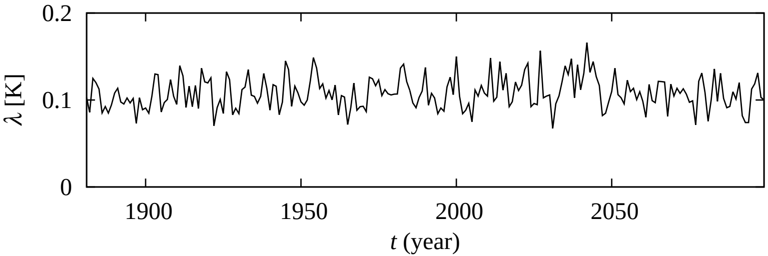
  and the latter can be estimated just by using the ensemble!

- Improving the estimation: fitting a Markovian form to the corresponding correlation matrix:

$$\begin{pmatrix} 1 & \rho_1 & \rho_1\rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_1\rho_2 & \rho_2 & 1 \end{pmatrix}$$

  (The fit is performed in a sliding window that is long enough to ensure independence on the window size.)

Example revisited



Do we have a trend here?

Linear fit:

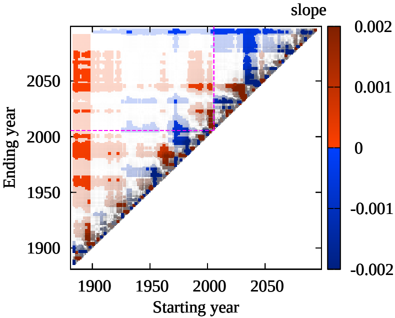
$$(2.08 \pm 2.06) \times 10^{-5}\ [\mathrm{K/yr}]$$

$$(0.11 \pm 2.39) \times 10^{-5}\ [\mathrm{K/yr}]$$

$\Longrightarrow$ Not significant.

Subintervals in the example

Grayscale: insignificant, saturated: 5% significant slope.



Yes, there are significant trends! (Warning: multiple testing, see Wilks, BAMS 97, 2263, 2016.)

- Strengthening ENSO before 1900,
- forced fluctuations until 2050
- weakening in the late 21st century.

Outline

Future work. . .