Methodology for statistical detection of climate change

Aurélien RIBES, Jean-Marc AZAÏS, Serge PLANTON

October 11, 2007

SEAMOCS Workshop

INSTITUT de MATHEMATIQUES de TOULOUSE

UNIVERSITE PAUL SABATIER TOULOUSE III

METEO FRANCE Toujours un temps d'avance

Detection

Definition, according to the IPCC AR3 :

"Detection is the process of demonstrating that an observed change is significantly different (in a statistical sense) than can be explained by natural internal variability"

Detection

Definition, according to the IPCC AR3 :

"Detection is the process of demonstrating that an observed change is significantly different (in a statistical sense) than can be explained by natural internal variability"

Detection of the anthropogenic climate change :

- Investigation of the presence of a particular signal, given by climate model simulation
- Double objectives :
  - Show the existence of a significant change of the observations
  - Validate the climate models

Statistical modelisation

Notations :

- ψ : random climate vector, of dimension p, taking one value per year
- C : the covariance matrix of ψ
- g : the climate change vector, of dimension p, given by a model

Framework

Assumptions

- $(\psi_i)_{i \in [\![1,n]\!]} \in \mathbb{R}^p$ rv iid having a $N(0, C)$ distribution
- $\psi_{n+1} \in \mathbb{R}^p$, independent of the $(\psi_i)_{i \in [\![1,n]\!]}$, having a $N(\mu g, C)$ distribution

With $\mu \in \mathbb{R}$, $g \in \mathbb{S}^{p-1} \subset \mathbb{R}^p$, and $C \in \mathcal{M}_p(\mathbb{R})$

Framework

Assumptions

- $(\psi_i)_{i\in[\![1,n]\!]} \in \mathbb{R}^p$ rv iid having a $N(0, C)$ distribution
- $\psi_{n+1} \in \mathbb{R}^p$, independent of the $(\psi_i)_{i\in[\![1,n]\!]}$, having a $N(\mu g, C)$ distribution

With $\mu \in \mathbb{R}$, $g \in \mathbb{S}^{p-1} \subset \mathbb{R}^p$, and $C \in \mathcal{M}_p(\mathbb{R})$

Test

One wants to test :

$$H_0 : \text{“}\mu = 0\text{”} \text{ vs } H_1 : \text{“}\mu > 0\text{”}$$

in large dimension, that is to say $n \sim p$.

Investigated test family

One considers the following tests $(T_f)_{f \in \mathbb{R}^p}$:

Test variable d_f

$$d_f = \langle \psi_{n+1}, f \rangle \quad \sim_{H_0} N(0, f' C f)$$

Rejection region

$$W_f = \left\{ \psi_{n+1} \in \mathbb{R}^p, d_f = \langle \psi_{n+1}, f \rangle \geq d_f^{(\alpha)} \right\}$$

with

$$d_f^{(\alpha)} = \Phi^{-1}(1 - \alpha) \sqrt{f' C f}$$

Optimal T_f

Question

Knowing C and g, is there an optimal T_f ?

Optimal T_f

Question

Knowing C and g, is there an optimal T_f ?

Answer : yes

$T_{C^{-1}g}$ is optimal, among the (T_f), within the following meaning

- $d_{C^{-1}g}$ maximise the signal-to-noise
- $T_{C^{-1}g}$ is the most powerful test
- $T_{C^{-1}g}$ is the likelihood ratio test

Notations

In the case "C unknown", one wants to approximate $T_{C^{-1}g}$.
Two new test families are considered.

Notations

In the case "C unknown", one wants to approximate $T_{C^{-1}g}$.
Two new test families are considered.

$\mathcal{T}_f$ tests

- f is "estimated" : depending on $(\psi_i)_{i\in[\![1,n]\!]}$, that are random, and on g
- The level, conditionally to $(\psi_i)_{i\in[\![1,n]\!]}$, is nominal;
  C is known, only for computing $d_f^{(\alpha)} = \Phi^{-1}(1-\alpha)\,\sqrt{f'Cf}$

Notations

In the case "C unknown", one wants to approximate $T_{C^{-1}g}$.
Two new test families are considered.

$\mathcal{T}_f$ tests

$\mathbb{T}_f$ tests

- f is estimated
- The threshold $d_f^{(\alpha)}$ is estimated depending on g and $(\psi_i)_{i\in[\![1,n]\!]}$; in the same way for the p-value
- The level is not necessarily nominal

Notations

In the case "C unknown", one wants to approximate $T_{C^{-1}g}$.
Two new test families are considered.

$\mathcal{T}_f$ tests

$\mathbb{T}_f$ tests

Naive test : T_g

Problem when C is not known

Objectives

One wants to construct a $\mathbb{T}_f$ test having "good" properties :

- Nominal level
- A power greater than the one of the naive test T_g

Problem when C is not known

Objectives

One wants to construct a $\mathbb{T}_f$ test having "good" properties :

- Nominal level
- A power greater than the one of the naive test T_g

Remark

It is possible to firstly study $\mathcal{T}_f$, noticing that "$\mathcal{T}_f > \mathbb{T}_f$"

Using the empirical covariance matrix

$$\hat{C} = \frac{1}{n} \sum_{i=1}^{n} \psi_i \psi_i'$$

Does $\mathcal{T}_{\hat{C}^{-1}g}$ have good properties ?

Using the empirical covariance matrix

$$\hat{C} = \frac{1}{n}\sum_{i=1}^{n} \psi_i \psi_i'$$

Does $\mathcal{T}_{\hat{C}^{-1}g}$ have good properties ?

No ! ($\hat{C}$ is quasi-singular)
The use of a pseudo-inverted truncation of $\hat{C}$ doesn't give an efficient test either

Regularisation in our case : shrinkage

Main idea

One uses $\gamma\hat{C} + \rho I_p$ instead of $\hat{C}$.

Justification

- Interpolation with the naive test T_g
- Ridge regression point of view
- Power of the tests $\mathcal{T}_{(\gamma\hat{C}+\rho I_p)^{-1}g}$

Estimation of the parameters (γ, ρ)

Optimal shrinkage method (O. Ledoit, *A Well-Conditioned Estimator for Large Dimensional Covariance Matrices*)

Framework :

- One reasons on $\mathcal{M}_p(\mathbb{R})$ with the norm

$$\|A\|^2_{\mathcal{M}_p} = \frac{\mathrm{Tr}(AA')}{p}$$

- A "general asymptotics" framework is used : n, p_n, $\frac{p_n}{n} \leq K$
- Estimators of the following family are investigated $C^* = \gamma \hat{C} + \rho I_p$. One researches the C^* minimising :

$$\mathrm{E}\left(\|C^* - C\|^2_{\mathcal{M}_p}\right)$$

Estimation of the parameters (γ, ρ)

Optimal shrinkage method (O. Ledoit, *A Well-Conditioned Estimator for Large Dimensional Covariance Matrices*)

n fixed : optimal γ_n^0 and ρ_n^0 depending on C

$$\gamma_n^0 = \frac{\alpha^2}{\delta^2}, \qquad \text{and} \qquad \rho_n^0 = \frac{\beta^2 \nu}{\delta^2}$$

where

$$\nu = \langle C, I_p \rangle_{\mathcal{M}_p} = \frac{\mathrm{Tr}(C)}{p}, \qquad \alpha^2 = \| C - \nu I_p \|^2_{\mathcal{M}_p},$$

$$\beta^2 = \mathrm{E}\left(\| \hat{C} - C \|^2_{\mathcal{M}_p} \right), \qquad \delta^2 = \mathrm{E}\left(\| \hat{C} - \nu I_p \|^2_{\mathcal{M}_p} \right),$$

Estimation of the parameters (γ, ρ)

Optimal shrinkage method (O. Ledoit, *A Well-Conditioned Estimator for Large Dimensional Covariance Matrices*)

n fixed : optimal γ_n^0 and ρ_n^0 depending on C

Under "general asymptotics" : convergent estimators $\hat{\gamma}_n^0$ and $\hat{\rho}_n^0$ of γ_n^0 and ρ_n^0

$$\hat{\gamma} = \frac{\hat{\alpha}^2}{\hat{\delta}^2}, \qquad \text{and} \qquad \hat{\rho} = \frac{\hat{\beta}^2 \hat{\nu}}{\hat{\delta}^2}$$

With

$$\hat{\nu} = \langle \hat{C}, I_p \rangle_{\mathcal{M}_p} = \frac{\mathrm{Tr}(\hat{C})}{p}, \qquad \hat{\beta}^2 = \min\left(\hat{\delta}^2, \frac{1}{n^2} \sum_{i=1}^{n} \| \psi_i \psi_i' - \hat{C} \|^2_{\mathcal{M}_p} \right),$$

$$\hat{\delta}^2 = \| \hat{C} - \hat{\nu} I_p \|^2_{\mathcal{M}_p}, \qquad \hat{\alpha}^2 = \hat{\delta}^2 - \hat{\beta}^2.$$

Estimation of the parameters (γ, ρ)

Optimal shrinkage method (O. Ledoit, *A Well-Conditioned Estimator for Large Dimensional Covariance Matrices*)

n fixed : optimal γ_n^0 and ρ_n^0 depending on C

Under "general asymptotics" : convergent estimators $\hat{\gamma}_n^0$ and $\hat{\rho}_n^0$ of γ_n^0 and ρ_n^0

A new estimator of C is defined :

$$\hat{C}_I = \hat{\gamma}_n^0 \hat{C} + \hat{\rho}_n^0 I_p$$

Bootstrap's motivations

One has to achieve the construction of $\mathbb{T}_{\hat{C}_l^{-1} g}$, by :

- computing the test threshold (and more generally the *p*-value),
- verifying that this computation is correct, or that the level is close to the nominal value,
- computing the power.

Additional treatment

- Choice of a learning sample,
- Temporal treatment (mobile average),
- Spatial centering.

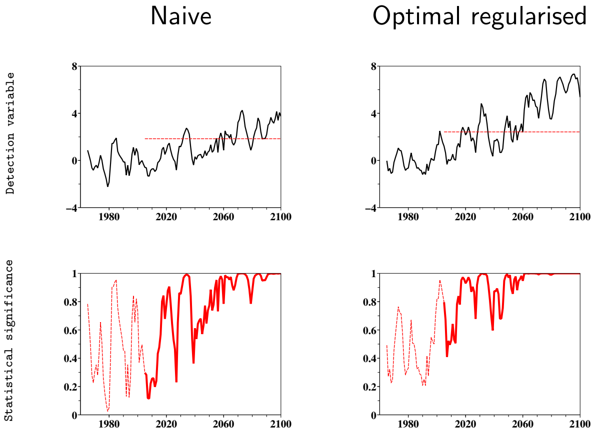

Figure: Summer minimum temperatures in the model : Comparison between the naive test T_g and the optimal regularised $\mathbb{T}_{\hat{C}_I^{-1}g}$ for summer minimum temperatures taken from a climate scenario.

Figure: Summer temperatures in observations : Results for minimum and maximum temperatures

Methodology for statistical detection of climate change

Aurélien RIBES, Jean-Marc AZAÏS, Serge PLANTON

October 11, 2007

Questions