Peut-on prévoir le climat du futur avec des statistiques ?

Aurélien Ribes
CNRM - Météo-France and CNRS

Journées MAS, 29 Août 2016

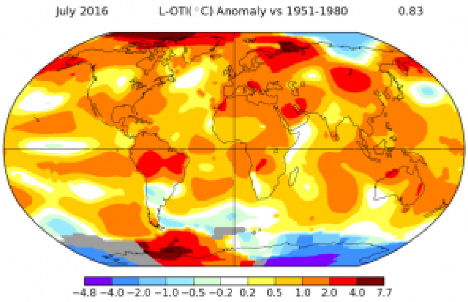
METEO FRANCE

cnrs

Introduction

16 août 2016

Climat : juillet 2016 nouveau record



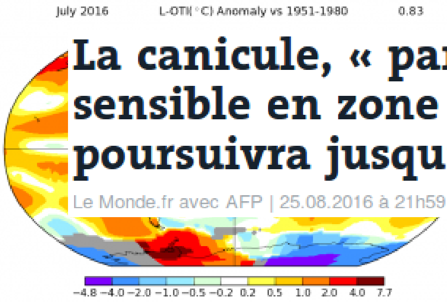
La série des mois records de chaleur planétaire s'est poursuivie en juillet. C'est ce qu'annonce l'équipe Nasa/Université Columbia de New York. Avec un écart à la moyenne calculée sur la période 1951/1980 de 0,83°C, juillet 2016 détient le nouveau record du mois.

La carte de ces écarts à la

Introduction

16 août 2016

Climat : juillet 2016 nouveau record



La canicule, « particulièrement sensible en zone urbaine », se poursuivra jusqu'à dimanche

Le Monde.fr avec AFP | 25.08.2016 à 21h59 • Mis à jour le 26.08.2016 à 09h53

La série des mois records de
ursuivie
annonce
olumbia
art à la
période
et 2016
détient le nouveau record du mois.

La carte de ces écarts à la

Introduction

16 août 2016

Climat : juillet 2016 nouveau record

La série des mois records de
ursuivie
annonce
olumbia
art à la
période
let 2016
détient le nouveau record du mois.

La canicule, « particulièrement sensible en zone urbaine », se poursuivra jusqu'à dimanche

Le Monde.fr avec AFP | 25.08.2016 à 21h59 • Mis à jour le 26.08.2016 à 09h53

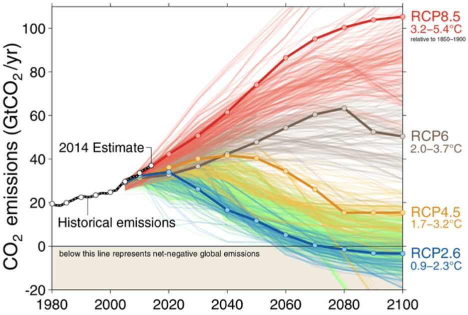
Climat : la chimère du 1,5 °C

LE MONDE | 22.08.2016 à 06h38 • Mis à jour le 22.08.2016 à 11h38 |

Par **Stéphane Foucart**

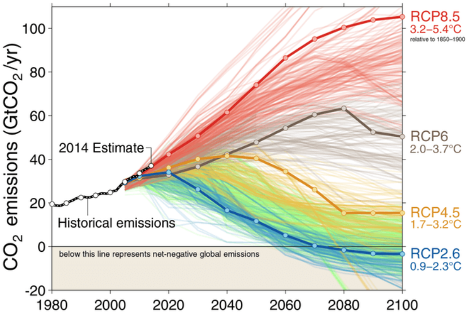
Problématique principale

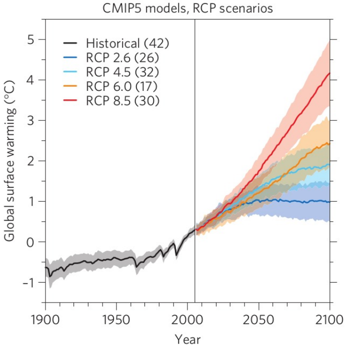
Décrire le climat de demain, conditionnellement aux émissions de GES



Problématique principale

Décrire le climat de demain, conditionnellement aux émissions de GES



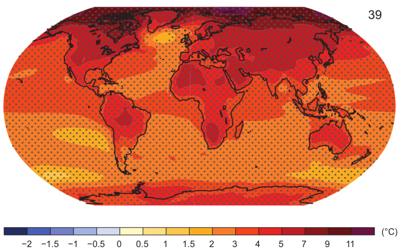


- Notion de sensibilité climatique

Problématique principale (2)

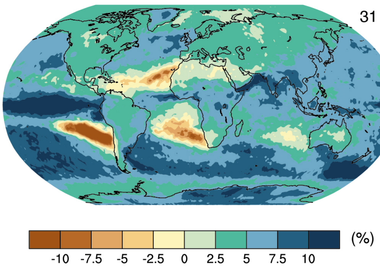
- Question posée identiquement pour d'autres variables, échelles, caractéristiques (e.g. événements extrêmes), etc.

Température moyenne

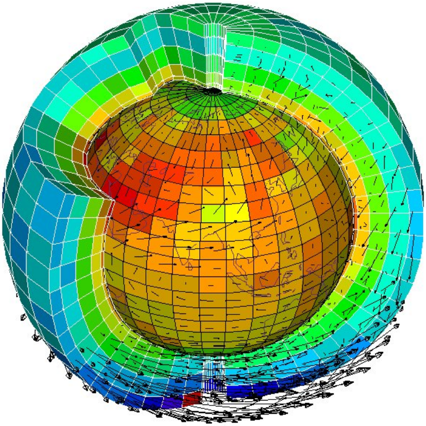


Extremes de précipitations

Daily precipitation 20-yr RV change per 1°C warming



Approche historique : la modélisation



Source : IPSL

Les modèles reposent sur:

- équations physiques
- solutions numériques approchées des EDP (schémas numerques)

Résolution :

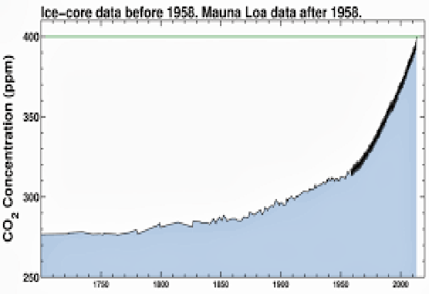
- spatiale (H) $\sim$ 100 km,
- spatiale (V) $\sim$ 500 m,
- temporelle $\sim$ 15'.

Résultats :

- nombreux progrès dans la connaissance du CC,
- pas de convergence sur la sensibilité climatique depuis $\sim$30ans.

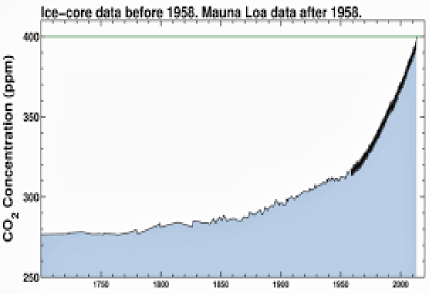
Alternative : les statistiques !

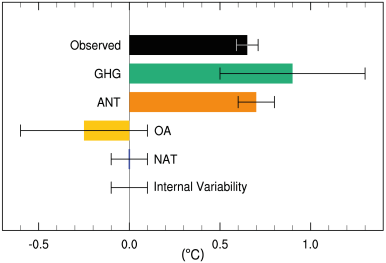
Alternative : étudier les 150 ans de données (observations) disponibles, avec des outils statistiques.



Alternative : les statistiques !

Alternative : étudier les 150 ans de données (observations) disponibles, avec des outils statistiques.





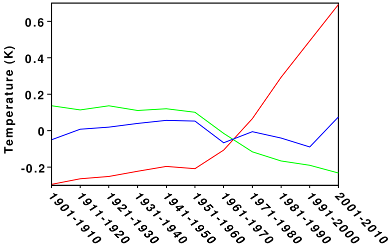
Réchauffement (1951 – 2010) attribuable à différents forçages externes

Étude des tendances observées de toutes façons nécessaire pour vérifier les résultats des modèles.

Comment distinguer GHG et OA ?

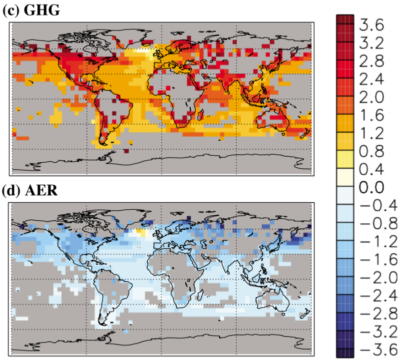
On utilise l'information disponible (modèles) sur leur distribution spatiale / temporelle

Séries temporelles



Variations de température moyenne ; rouge:GHG, bleu:NAT, vert OA (i.e. autres anthropiques).

Distributions spatiales



Tendance de réchauffement 1851-2010

1. Introduction - Problématique
2. Modèles et inférence statistique
   - OLS
   - TLS
   - EIV
   - Alternative
3. Problèmes et défis
4. Conclusion

Regression based models

$$Y_\ell = \sum_{i=1}^{k} \beta_i X_\ell^{(i)} + \varepsilon_\ell, \qquad Cov(\varepsilon) = \Sigma,$$

$$\underset{n \times 1}{Y} = \underset{n \times k,}{X} \; \underset{k \times 1}{\beta} + \underset{n \times 1}{\varepsilon,} \qquad \underset{n \times n}{Cov(\varepsilon) = \Sigma,}$$

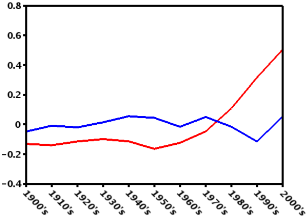
- ℓ: location (space-time),
- Y: observations (space-time vector),
- β_i: scaling factor (scalar), unknown,
- $X^{(i)}$: expected response to forcing i (space-time vector), known,
- ε: internal variability (space-time vector),
- Σ: IV covariance matrix (matrix).

Regression based models

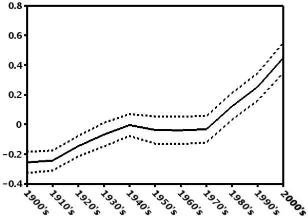
$$Y_\ell = \sum_{i=1}^{k} \beta_i X_\ell^{(i)} + \varepsilon_\ell, \qquad Cov(\varepsilon) = \Sigma,$$

$$\underset{n\times 1}{Y} = \underset{n\times k,}{X} \; \underset{k\times 1}{\beta} + \underset{n\times 1}{\varepsilon,} \qquad \underset{n\times n}{Cov(\varepsilon) = \Sigma,}$$

X : simulations



Y : observations



Regression based models

$$Y_\ell = \sum_{i=1}^{k} \beta_i X_\ell^{(i)} + \varepsilon_\ell, \qquad Cov(\varepsilon) = \Sigma,$$

$$\underset{n\times 1}{Y} = \underset{n\times k,}{X} \; \underset{k\times 1}{\beta} + \underset{n\times 1}{\varepsilon,} \qquad \underset{n\times n}{Cov(\varepsilon) = \Sigma,}$$

General philosophy

- The space-time pattern of the response to each forcing is known (eg GHG, AER),
- We have some uncertainty on the magnitude of the response (eg a factor 3 on climate sensitivity),
- So, based on the pattern, we want to estimate the amplitude of each response, from observations.

OLS

Philosophy: The response patterns X are perfectly known.

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \Sigma), \qquad (\Sigma \text{ known})$$

Likelihood (-2 log-): $\ell_{\text{OLS}}(\beta) = (Y - X\beta)'\Sigma^{-1}(Y - X\beta).$

Inference

- If Σ known, then multiply by $\Sigma^{-1/2}$!
- $\widehat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \qquad \sim N(\beta, (X'\Sigma^{-1}X)^{-1}),$
- The issue of how estimating Σ to efficiently approximate $\widehat{\beta}$ is uncommon.

TLS model (Allen & Stott, 2003)

Philosophy: Taking into account the internal variability within the climate simulations leading to X.

$$
\begin{aligned}
Y &= X^*\beta + \varepsilon, \quad \varepsilon \sim N(0, \Sigma) && (1)\\
X &= X^* + \varepsilon_X, \quad \varepsilon_X \sim N(0, \Sigma/n_X) && (2)
\end{aligned}
$$

TLS: model and likelihood

TLS model

Regression equation $$Y^* = X^*\beta$$

One observes $$\begin{cases} X = X^* + \varepsilon_X, & \mathrm{Cov}(\varepsilon_X) = \Sigma_X \\ Y = Y^* + \varepsilon_Y, & \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y. \end{cases}$$

$$\Sigma_X = \lambda \Sigma_Y.$$

TLS: model and likelihood

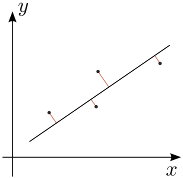
TLS model

Regression equation $$Y^* = X^* \beta$$

One observes $$\begin{cases} X = X^* + \varepsilon_X, & \mathrm{Cov}(\varepsilon_X) = \Sigma_X \\ Y = Y^* + \varepsilon_Y, & \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y. \end{cases}$$

$$\Sigma_X = \lambda \Sigma_Y.$$

$$\ell_{\mathrm{TLS}}(\beta, X^*) = (Y - X^*\beta)'\Sigma_Y^{-1}(Y - X^*\beta) + (X - X^*)'\Sigma_X^{-1}(X - X^*).$$



TLS: model and likelihood

TLS model

Regression equation $$Y^* = X^*\beta$$

One observes $$\begin{cases} X = X^* + \varepsilon_X, & \mathrm{Cov}(\varepsilon_X) = \Sigma_X \\ Y = Y^* + \varepsilon_Y, & \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y. \end{cases}$$

$$\Sigma_X = \lambda \Sigma_Y.$$

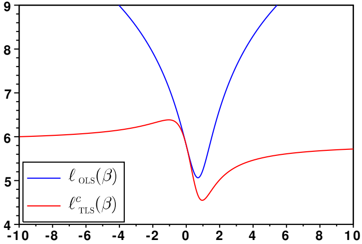
$$\begin{aligned} \ell_{\mathrm{TLS}}(\beta, X^*) &= (Y - X^*\beta)'\Sigma_Y^{-1}(Y - X^*\beta) + (X - X^*)'\Sigma_X^{-1}(X - X^*). \\ \ell^c_{\mathrm{TLS}}(\beta) &= \frac{(Y - X\beta)'\Sigma^{-1}(Y - X\beta)}{1 + \beta^2} \qquad \text{(assuming } \lambda = 1\text{)}. \end{aligned}$$

Inference

- Estimation : MLE in closed form (SVD of $[Y, X]$),
- CI : asymptotic - too low converage probability.

TLS: Likelihood

$$\ell^c_{\text{TLS}}(\beta) = \frac{(Y - X\beta)'\Sigma^{-1}(Y - X\beta)}{1 + \beta^2}. \quad (3)$$



If you assume the wrong model

- OLS is "the truth": $\widehat{\beta}_{TLS}$ is not optimal (too much variance),
- TLS is "the truth": $\widehat{\beta}_{OLS}$ is biased toward 0.

1. Introduction - Problématique
2. **Modèles et inférence statistique**
   - OLS
   - TLS
   - **EIV**
   - Alternative
3. Problèmes et défis
   - Grande dimension
   - Modélisation de la variabilité interne
   - Estimation de grandes matrices de covariance
   - Estimer l'incertitude de modélisation
4. Conclusion

EIV model (Huntingford et al., 2006; Hannart et al., 2014)

Philosophy: Taking into account the modeling uncertainty (ie different climate models provide different patterns X or X^*).

Regression equation $$Y^* = X^*\beta$$

One observes $$\begin{cases} Y = Y^* + \underbrace{\varepsilon_{Y,IV}}_{\varepsilon_Y} \\ X = X^* + \underbrace{\varepsilon_{X,IV}}_{\varepsilon_X} \end{cases}$$

With $$\begin{cases} \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y = \Sigma_{IV}, & \text{(I.V.)} \\ \mathrm{Cov}(\varepsilon_X) = \Sigma_X = \Sigma_{IV}, & \text{(I.V.)} \end{cases}$$

EIV model (Huntingford et al., 2006; Hannart et al., 2014)

Philosophy: Taking into account the modeling uncertainty (ie different climate models provide different patterns X or X^*).

Regression equation $$Y^* = X^* \beta$$

One observes $$\begin{cases} Y = Y^* + \underbrace{\varepsilon_{Y,IV}}_{\varepsilon_Y} \\ X = X^* + \underbrace{\varepsilon_{X,IV} + \varepsilon_{Mod}}_{\varepsilon_X} \end{cases}$$

With $$\begin{cases} \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y = \Sigma_{IV}, & \text{(I.V.)} \\ \mathrm{Cov}(\varepsilon_X) = \Sigma_X = \Sigma_{IV} + \Sigma_{Mod}, & \text{(I.V. + Mod. Uncert.)} \end{cases}$$

EIV model (Huntingford et al., 2006; Hannart et al., 2014)

Philosophy: Taking into account the modeling uncertainty (ie different climate models provide different patterns X or X^*).

Regression equation $$Y^* = X^* \beta$$

One observes $$\begin{cases} Y = Y^* + \underbrace{\varepsilon_{Y,IV} + \varepsilon_{Obs}}_{\varepsilon_Y} \\ X = X^* + \underbrace{\varepsilon_{X,IV} + \varepsilon_{Mod}}_{\varepsilon_X} \end{cases}$$

With $$\begin{cases} \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y = \Sigma_{IV} + \Sigma_{Obs}, & \text{(I.V. + Obs. Uncert.)} \\ \mathrm{Cov}(\varepsilon_X) = \Sigma_X = \Sigma_{IV} + \Sigma_{Mod}, & \text{(I.V. + Mod. Uncert.)} \end{cases}$$

EIV model

Regression equation $$Y^* = X^*\beta$$

Observations $$\begin{cases} Y = Y^* + \varepsilon_Y, & \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y, \\ X = X^* + \varepsilon_X, & \mathrm{Cov}(\varepsilon_X) = \Sigma_X. \end{cases}$$

Σ_Y and Σ_X have no relationship.

EIV: Likelihood

Regression equation $$Y^* = X^*\beta$$

Observations $$\begin{cases} Y = Y^* + \varepsilon_Y, & \mathrm{Cov}(\varepsilon_Y) = \Sigma_Y, \\ X_i = X_i^* + \varepsilon_{X,i}, & \mathrm{Cov}(\varepsilon_{X,i}) = \Sigma_{X,i}. \end{cases}$$

Assuming $k = 1$,

$$\begin{aligned} \ell(\beta, X^*) &= cte + (Y - X^*\beta)'\Sigma_Y^{-1}(Y - X^*\beta) + (X - X^*)'\Sigma_X^{-1}(X - X^*). \\ \ell_c(\beta) &= cte + (Y - X\beta)'(\beta^2\Sigma_X + \Sigma_Y)^{-1}(Y - X\beta). \end{aligned}$$

No explicit maximum (i.e. MLE) !

Estimation (Hannart et al., 2014)

Philosophy: maximize $\ell(\beta, X^*)$ with a numerical algorithm (MLE).

Estimation: algorithm

- 1. Fix $\widehat{\beta}^0$,
- 2i. Compute $\widehat{X}_i^* = \text{Argmax}_{X^*} \ell(\widehat{\beta}^{i-1}, X^*)$,
- 3i. Compute $\widehat{\beta}_i = \text{Argmax}_{\beta} \ell(\beta, \widehat{X}_i^*)$,
- 4. When convergence occurs, you have $(\widehat{\beta}, \widehat{X^*})$.

Issue(s): may converge to a critical point or local maximum (not necessarily the global maximum, i.e. MLE).

Confidence intervals

Use asymptotic property of MLE.

Issue(s): too low coverage probability (i.e.: too small CI).

Motivation: Is linear regression suitable?

$$Y = \sum_{i=1}^{k} \beta_i X_i + \varepsilon$$

- Assumes that
  - models are able to simulate response patterns,
  - response magnitudes are unknown.

Motivation: Is linear regression suitable?

$$Y = \sum_{i=1}^{k} \beta_i X_i + \varepsilon$$

- Assumes that
  - models are able to simulate response patterns,
  - response magnitudes are unknown.
- The reality is probably more balanced

A new possible approach (Ribes et al., 2016)

$$\begin{cases} Y = Y^* + \varepsilon_Y, & \varepsilon_Y \sim N(0, \Sigma_Y), \\ X_i = X_i^* + \varepsilon_{X_i}, & \varepsilon_{X_i} \sim N(0, \Sigma_{X_i}), \quad i = 1, \ldots, k, \end{cases}$$

A new possible approach (Ribes et al., 2016)

$$Y^* = \sum_{i=1}^{k} X_i^*,$$

$$\begin{cases} Y = Y^* + \varepsilon_Y, & \varepsilon_Y \sim N(0, \Sigma_Y), \\ X_i = X_i^* + \varepsilon_{X_i}, & \varepsilon_{X_i} \sim N(0, \Sigma_{X_i}), \quad i = 1, \dots, k, \end{cases}$$

A new possible approach (Ribes et al., 2016)

$$Y^* = \sum_{i=1}^{k} X_i^*,$$

$$\begin{cases} Y = Y^* + \varepsilon_Y, & \varepsilon_Y \sim N(0, \Sigma_Y), \\ X_i = X_i^* + \varepsilon_{X_i}, & \varepsilon_{X_i} \sim N(0, \Sigma_{X_i}), \quad i = 1, \ldots, k, \end{cases}$$

- Interpretation: models give information on each term X_i^*, then an additional constraint on their sum comes from observations.
- Bayesian perspective:
  - Prior on each X_i^*, derived from models,
  - Posterior after constraining the sum by observations.
- All inference can be made with maximum likelihood

$$\widehat{X}_i^* = X_i + \Sigma_{X_i}(\Sigma_Y + \Sigma_X)^{-1}(Y - X) \sim N(X_i, \Sigma_{\widehat{X}_i^*}).$$

1 Introduction - Problématique

2 Modèles et inférence statistique

3 Problèmes et défis
- Grande dimension
- Modélisation de la variabilité interne
- Estimation de grandes matrices de covariance
- Estimer l'incertitude de modélisation

4 Conclusion

Grande dimension des données

Jeu de donnée climatique typique (e.g. temperature de surface)

- Dimension spatiale : $5^o \times 5^o \sim 2600$ points de grille,
- Dimension temporelle : 50 - 100 ans (periode instrumentale),
- Dimension de $Y \sim 10^5$.
- Autre possibilités : cycle annuel, cycle diurne, autres variables...

Grande dimension des données

Jeu de donnée climatique typique (e.g. temperature de surface)

- Dimension spatiale : $5^o \times 5^o \sim 2600$ points de grille,
- Dimension temporelle : 50 - 100 ans (periode instrumentale),
- Dimension de $Y \sim 10^5$.
- Autre possibilités : cycle annuel, cycle diurne, autres variables...

Cycle annuel, cycle diurne, autres variables...

Grande dimension des données

Jeu de donnée climatique typique (e.g. temperature de surface)

- Dimension spatiale : $5^o \times 5^o \sim 2600$ points de grille,
- Dimension temporelle : 50 - 100 ans (periode instrumentale),
- Dimension de $Y \sim 10^5$.
- Autre possibilités : cycle annuel, cycle diurne, autres variables...

Cycle annuel, cycle diurne, autres variables...

Question : quelle(s) variable(s) regarder, quelle période / région, etc ?

Decreasing the dimension (or pre-processing)

Statistical investigation of climate at the global scale requires to reduce the spatio-temporal dimension of datasets.

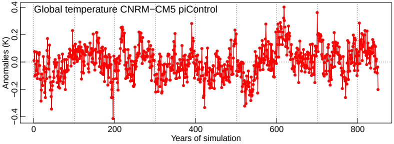
- Decadal means,
- Projection on spherical harmonics (e.g. truncation T4, $\sim$ spatial scales $>$ 5000 kms),
- Use of simple climate indices (global mean, land-sea contrast, inter-hemispheric contrast, annual cycle, etc).
- Projection on PCs,

This treatment is quite arbitrary and non optimal.

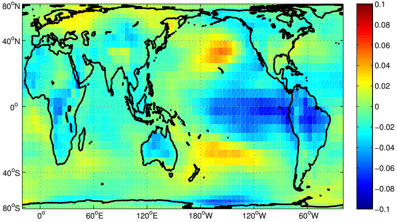
Comment modéliser la variabilité interne ?

Quelques illustrations de la variabilité interne :

Temporel



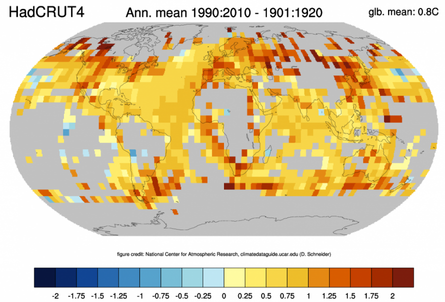
Spatial



Comment modéliser la variabilité interne ?

Quelques illustrations de la variabilité interne :

Variations du réseau d'observation



Observational coverage in the early 20th century.

Pas de modélisation statistique spatio-temporelle simple !

1. Introduction - Problématique
2. Modèles et inférence statistique
   - OLS
   - TLS
   - EIV
   - Alternative
3. **Problèmes et défis**
   - Grande dimension
   - Modélisation de la variabilité interne
   - **Estimation de grandes matrices de covariance**
   - Estimer l'incertitude de modélisation
4. Conclusion

Problem statement

- Most inference methods assume that Σ is known.
(and the full distribution of the internal variability ε).
- Usually, climate models are used to derive a few realisations of ε, say $\varepsilon_1, \ldots, \varepsilon_p$.
- The distribution, or at least, $\Sigma = \mathrm{Cov}(\varepsilon)$ is estimated from these.
- Optimal statistics requires to estimate Σ^{-1}
(eg $\widehat{\beta}_{OLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$).

This is very challenging in high dimension.

Estimation of Σ in large dimension

Let us assume that $\varepsilon_1, \ldots, \varepsilon_p \sim N(0, \Sigma)$ are available for estimating Σ ($n \times n$).

Estimation of Σ in large dimension

Let us assume that $\varepsilon_1, \ldots, \varepsilon_p \sim N(0, \Sigma)$ are available for estimating Σ $(n \times n)$.

What about $\widehat{\Sigma}$?

The sample estimate $\widehat{\Sigma}$ is a poor estimator of Σ in large dimension (n close to p).

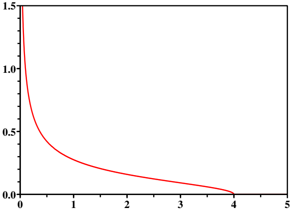
Estimation of Σ in large dimension

Let us assume that $\varepsilon_1, \ldots, \varepsilon_p \sim N(0, \Sigma)$ are available for estimating Σ ($n \times n$).

What about $\widehat{\Sigma}$?

The sample estimate $\widehat{\Sigma}$ is a poor estimator of Σ in large dimension (n close to p).

Illustration : case $\Sigma = I$, distribution of the eigenvalues of $\widehat{\Sigma}$ when $n, p \to \infty$ (Marčenko-Pastur distribution).



Regularising Σ (1)
Principle

Principle

We use an estimator of Σ such as

$$\widetilde{\Sigma} = \gamma\widehat{\Sigma} + \rho I.$$

Regularising Σ (1)

Principle

Principle

We use an estimator of Σ such as

$$\widetilde{\Sigma} = \gamma \widehat{\Sigma} + \rho I.$$

LW estimate (Ledoit & Wolf, 2004)

- Introduction of estimators $\hat{\gamma}, \hat{\rho}$ of γ, ρ to minimise the mean square error
$$\mathrm{E}\left(\|\widetilde{\Sigma} - \Sigma\|^2_{\mathcal{M}} \right).$$
- $$\widehat{\Sigma}_I = \hat{\gamma} \hat{\Sigma} + \hat{\rho} I.$$

Regularising Σ (1)

Principle

Principle

We use an estimator of Σ such as

$$\widetilde{\Sigma} = \gamma \widehat{\Sigma} + \rho I.$$

LW estimate (Ledoit & Wolf, 2004)

- Introduction of estimators $\hat{\gamma}, \hat{\rho}$ of γ, ρ to minimise the mean square error

$$\mathrm{E}\left(\|\widetilde{\Sigma} - \Sigma\|^2_{\mathcal{M}} \right).$$

- $$\widehat{\Sigma}_I = \hat{\gamma}\hat{\Sigma} + \hat{\rho} I.$$

New estimator (Ribes et al., 2009)

$$\widehat{\beta}_I = (X' \widehat{\Sigma}_I^{-1} X)^{-1} X' \widehat{\Sigma}_I^{-1} Y.$$

Integrated Optimal Fingerprinting approach

- Regularization with a target $\Delta \neq I$ (Hannart et Naveau, 2014, JMVA).

  Use of a Bayesian prior: $\Sigma \sim \mathcal{W}^{-1}(\Delta, \alpha)$ (centered on Δ),

  Derive estimators $\widehat{\rho}_1, \widehat{\rho}_2$ leading to $\widetilde{\Sigma}_\Delta = \widehat{\rho}_1 \widehat{\Sigma} + \widehat{\rho}_2 \Delta$.

- Estimation of Σ (and therefore Σ^{-1}) and β in a joint statistical framework (Hannart, 2016, JClim).

  Uncertainty on Σ is partly taken into account in the estimation and CI on β.

$$\widehat{\beta}_\Delta = (X' \widehat{\Sigma}_\Delta^{-1} X)^{-1} X' \widehat{\Sigma}_\Delta^{-1} Y.$$

1. Introduction - Problématique
2. Modèles et inférence statistique
   - OLS
   - TLS
   - EIV
   - Alternative
3. Problèmes et défis
   - Grande dimension
   - Modélisation de la variabilité interne
   - Estimation de grandes matrices de covariance
   - Estimer l'incertitude de modélisation
4. Conclusion

Estimer l'incertitude de modélisation

- Ensemble de modèles de taille limitée (~ 30, contrainte sur la dimension),
- Modèles non-indépendants (développements communs), plutôt *Ensemble d'opportunité*,
- Modèles trop couteux pour utiliser les techniques classiques de quantification de l'incertitude,
- Approche la plus répandue : considérer la dispersion des modèles,

"models (m_i) are stat. indistinguishable from the truth (m^*)", i.e.

$$(m_i - m_j) \sim N(0, 2\Sigma_m), \quad (m_i - m^*) \sim N(0, 2\Sigma_m),$$

Estimer l'incertitude de modélisation

- Ensemble de modèles de taille limitée (~ 30, contrainte sur la dimension),
- Modèles non-indépendants (développements communs), plutôt *Ensemble d'opportunité*,
- Modèles trop couteux pour utiliser les techniques classiques de quantification de l'incertitude,
- Approche la plus répandue : considérer la dispersion des modèles,

"models (m_i) are stat. indistinguishable from the truth (m^*)", i.e.

$$(m_i - m_j) \sim N(0, 2\Sigma_m), \quad (m_i - m^*) \sim N(0, 2\Sigma_m),$$

ou écrit différemment (μ: espérance de la population des modèles)

$$(m_i - \mu) \sim N(0, \Sigma_m), \quad (\mu - m^*) \sim N(0, \Sigma_m)$$

Conclusions

- Un grand nombre de modèles et méthodes statistiques utilisées en sciences du climat.
- De nombreux défis :

climat : estimation du réchauffement induit par les GES dans le passé,
stats : gestion de la grande dimension, modélisation spatio-temporelle, quantification d'incertitude / design.

- Cela ne concerne pas que la température moyenne globale.
- Autres interactions entre climat et statistiques, notamment théorie des valeurs extrêmes, attribution d'événements singuliers.