Températures maximales en France au 21ème siècle

Comité de suivi de deuxième année - 2024

Encadrants: Philippe NAVEAU, Nathalie BERTRAND, Aurélien RIBES

Occitane Barbaux (occitane.barbaux@umr-cnrm.fr)

20 Juin 2024

ÉDUCATION

ENSAE | Ingénieur Master 2

Diplômée en Novembre 2020 | Paris, France

Diplômée de l'École Nationale de la Statistique et de l'Administration Économique (ENSAE), **voie Machine Learning, Statistiques et Apprentissages**.

EXPERIENCE

IRSN | Soutien statistique au Groupe de travail Vent et Neige | Intérim 18 mois

Nov 2020 – 22 Mai 2022 | BEHRIG | Fontenay-aux-roses, France

IRSN | Stagiaire Modélisation de valeurs extrêmes

Mai 2020 – Sep 2020 | BEHRIG | Fontenay-aux-roses, France

Stage de **modélisation des vents extrême**.

UNIVERSITY OF CAMBRIDGE | Stagiaire Biostatistiques

Jan 2019 – Jul 2019 | Cardiovascular Epidemiology Unit | Cambridge, UK

Stage de recherche sur les facteurs de risques cardio-vasculaires.

Formations suivies :

- Stage ”Changement Climatique” de Météo France. Octobre 2022.
- École d’été ”Traitement des Données Massives et Apprentissage: Applications en Géophysique, Écologie et SHS” - Juin 2023
- Semaine des Doctorants de l’ED 129. (Science Ouverte, Dévelopment Durable, Ethique) - Avril 2023
- MOOC ”Ethique de la recherche et intégrité scientifique” - Janvier 2023
- ”Why and how to develop your postdoc project 1” - Octobre 2023
- Formation au débat public sur des controverses scientifiques - Decembre 2023
- Formation ”Conducting a project : conduct, lead and manage” - Avril 2024

Présentation des travaux :

- Journées des doctorants IRSN (Poster) - Mars 2023
- Valpred 2023 (Poster) - Avril 2023
- Semaine des Doctorants de l'ED 129 (Poster) - Avril 2023
- Extreme Value Analysis 2023 (Poster) - Juin 2023
- Lancement PC4 TRACCS (Présentation) - Février 2024
- Journées des doctorants IRSN (Présentation) - Avril 2024
- 55ièmes Journées de Statistique (Présentation) - Mai 2024
- Journées des doctorants MF (Présentation) - Juin 2024
- International Meeting on Statistical Climatology (Présentation) - Juin 2024
- Advances in Extreme Value Analysis and Application to Natural Hazards 2024 (Poster) - Juillet 2024
- Enseignement :"Classification Non-supervisée" à l'École Nationale de la Météo pour les années scolaires 2022-2023 et 2023-2024.

Climate Change

- Increase in **frequency and intensity** of extremely hot events. [1]
- Increasing knowledge of the warming phenomena, using both **observations and climate models**.

Safety concerns

- Reliability of **safety-significant equipment**.
- Building Codes using stationary return levels which may **vary during the building's life**.

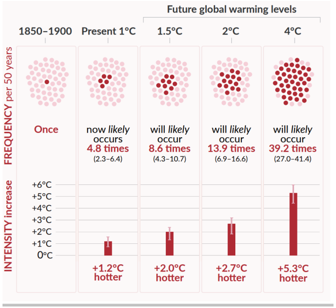


Figure: IPCC, 2021: Summary for Policymakers by Masson Delmotte, V et al.[6]

Our Goal: Defining the risk of excess and applying it to extreme temperature levels by 2100 at a local scale.

How?

- Adapting the stationary return level to a **non-stationary context**, considering the lifetime of the building.
- Estimating extreme temperature levels, integrating **information from climate models and local observations**, using tools based on Extreme Value Theory and a Bayesian framework.
- Providing a usable estimate taking **uncertainty** into account.
- Adapting the method to temperature in **various places of interest**, taking into account the inherent limitations of each zone.

Need:

- Separating the **period of interest** from the **return period** (annual probability = $\frac{1}{T}$).
- Account for **non-stationnarity,** $Y_{2023} \neq Y_{2050}$.
- Applied similarly with or without stationarity.

Our choice: Equivalent Reliability [7] [5]

For period $t_1, .., t_2$ and annual probability p , solution $\mathbf{z_p}$ of :

$$P[Max(Z_{t_1}, Z_{t_1+1}, ..., Z_{t_2}) \leq \mathbf{z}] = (1-p)^{\mathbf{t_2-t_1}+1}$$

Constraints:

- **Extreme Values** Analysis: Annual Maxima, use of **GEV** distribution.
- **Non-stationarity**: Use of mean European Temperature as covariate allows for a better time relationship and **scenario integration**.

$$Y \sim \mathbb{P}_t = GEV(\mu_t, \sigma_t, \xi)$$

$$\begin{cases} \mu(t) & = \mu_0 + \mu_1 X_t \\ \sigma(t) & = exp(\sigma_0 + \sigma_1 X_t) \\ \xi(t) & = \xi_0 \end{cases}$$

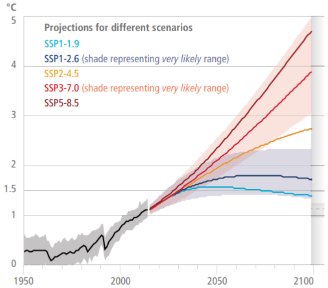
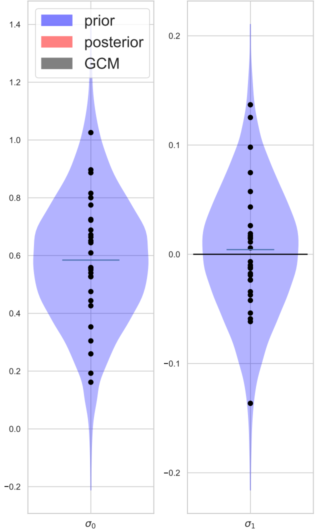


Figure: Global surface temperature changes for various scenarios. (IPCC, 2022)

Bayesian framework[9]

A-priori knowledge

- Include only information from **climate models**. (historical and scenario).



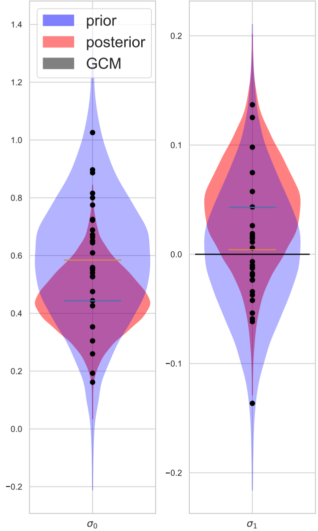
Bayesian framework[9]

A-priori knowledge

- Include only information from **climate models**. (historical and scenario).

Updated using observations

- Maxima constraint using **Markov chain Monte Carlo** (NUTS).
- Using past **local observations**.



Bayes Theorem:
Using $\psi = \{X_{1850} - X_{2100}, \theta_{GEV}\}$, y^0 for local observations and x^0 for covariate observations.

$$\mathbb{P}(\psi|y^0 \cap x^0) = \frac{\mathbb{P}[y^0|(\psi|x^0)]\mathbb{P}(\psi|x^0)}{\mathbb{P}(y^0)}$$

Bayesian constraint is done in **two steps**[9]:

- **Covariate constraint**: $\mathbb{P}(\psi|x^0)$ is calculated using a conjugate.
- **Local constraint** $\mathbb{P}(\psi|x^0, y^0)$ is calculated by constraining $\mathbb{P}(\psi|x^0)$ with a **MCMC** chain.

Based on the Metropolis-Hasting Algorithm[8][3] with a fixed transition.

Issues:

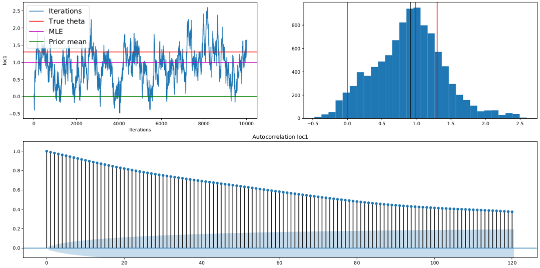
- Created and adjusted on a **unique example**.
- **No testing** done on chain quality or convergence.
- **Errors** when the scale of the distribution parameters differs.
- **Too slow** for application such as online constraint.

Aims:

- Evaluating the existing algorithm.
- Exploring and testing out variants.
- Choosing the most suitable version.
- Comparing to popular alternatives.

Original:

- Some chains do **not mix well.**
- Median full ESS is 291 (For 10000 iterations)

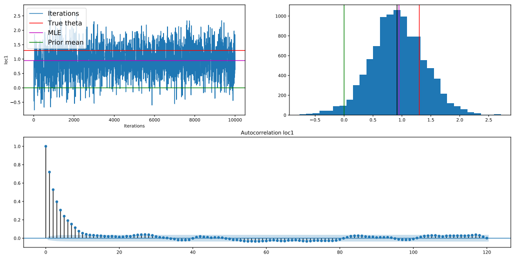


Original:

- Some chains do **not mix well**.
- Median full ESS is 291 (For 10000 iterations)

SCAM: Single Component Adaptive Metropolis[4]

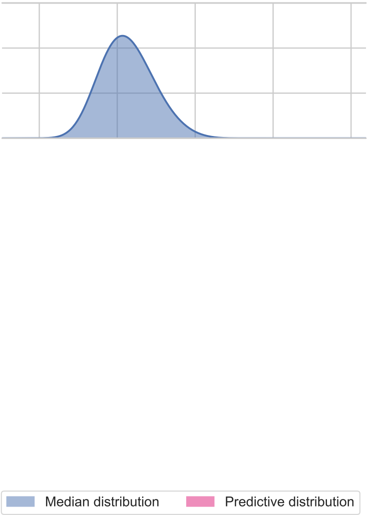
- Chains mix better but still have a **lot of autocorrelation**.
- Median full ESS is 1226 (For 10000 iterations)



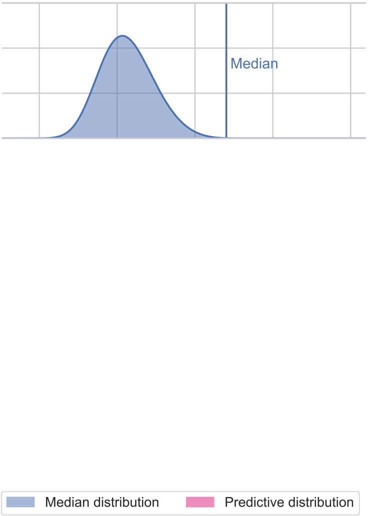
NUTS: the No-U-Turn-Sampler

- STAN implementation in C++. **Very fast** (up to 100x).
- **Adaptative MCMC** that uses the derivatives of the density function being sampled
- More **robust** when the correlation between parameters is high.
- Only keep samples after convergence.

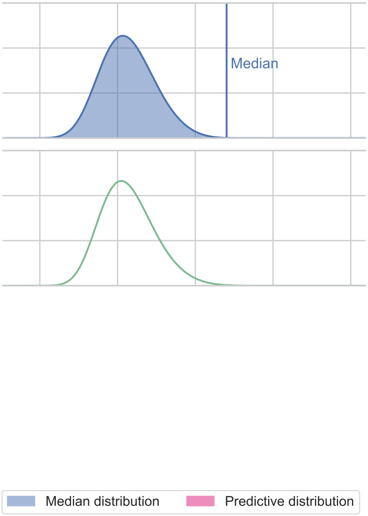
$\Longrightarrow$ **Best choice**. Being added to Yoann's BSAC + used as an addon for NSSEA by Saïd.
$\Longrightarrow$ Participation to **Yoann's future paper**.



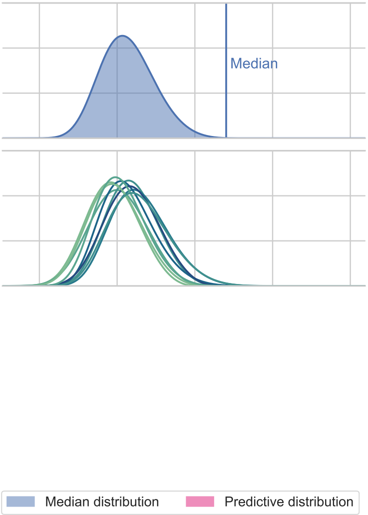
- Using all draws: for return levels, **median**



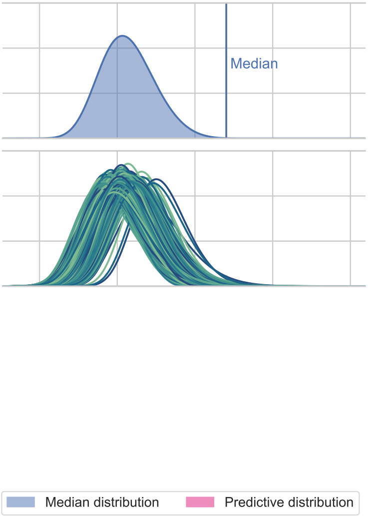
- Using all draws: for return levels, **median**



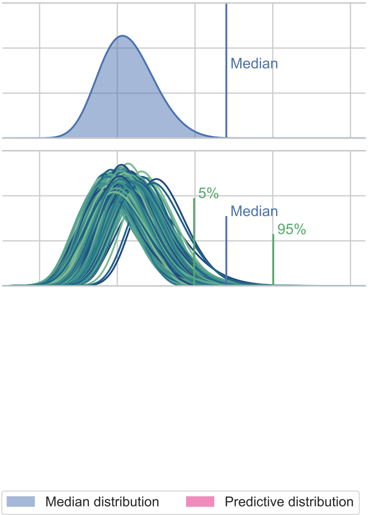
- Using all draws: for return levels, **median**



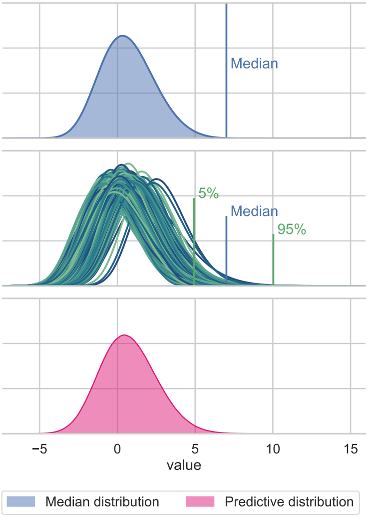
- Using all draws: for return levels, **median**



- Using all draws: for return levels, **median** and **confidence intervals**



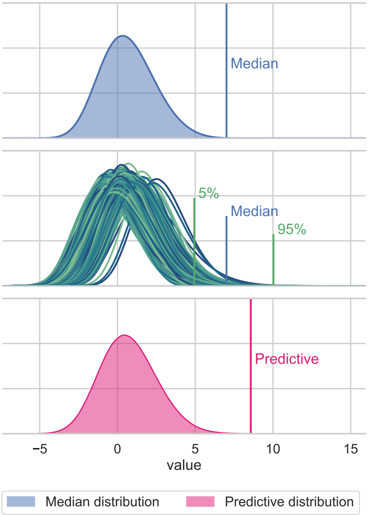
- Using all draws: for return levels, **median** and **confidence intervals**
- **Issue** : Confidence Level is **another parameter** to choose.



- Using all draws: for return levels, **median** and **confidence intervals**
- **Issue** : Confidence Level is **another parameter** to choose.
- **Predictive distribution[2]** : One distribution **averaged** over the distribution of the model parameters.

$$P(Z \leq z|z_0) = \int_\Theta P(Z \leq z|\theta)\pi(\theta|z_0)d\theta$$

$$\sim \frac{1}{N}\sum_{i=1}^{N} P(Z \leq z|\theta_i)$$

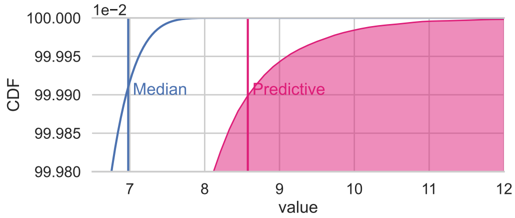


- Using all draws: for return levels, **median** and **confidence intervals**
- **Issue** : Confidence Level is **another parameter** to choose.
- **Predictive distribution[2]** : One distribution **averaged** over the distribution of the model parameters.

$$P(Z \leq z|z_0) = \int_\Theta P(Z \leq z|\theta)\pi(\theta|z_0)d\theta$$
$$\sim \frac{1}{N}\sum_{i=1}^{N} P(Z \leq z|\theta_i)$$

- One distribution **blending all draws.**

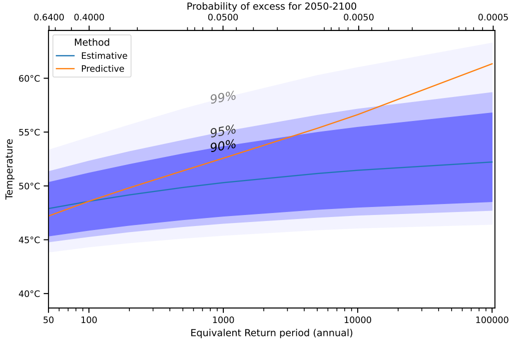
- Account for **estimation error** and **stochastic error**.
- Create an unbounded distribution (with our parameter distribution)
- Difference with the median is most visible **on the tail**.

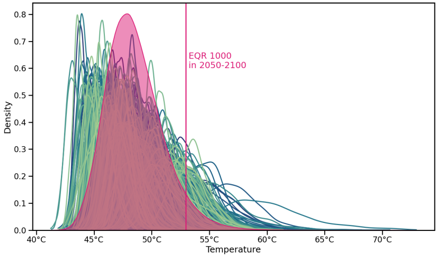


y_p solution of $\int_{\Theta} P(max(Y_{t_1},, Y_{t_1+1}, ..., Y_{t_2}) < y_p|\theta)p(\theta|y^{obs})d\theta = (1-p)^{t_2-t_1+1}$

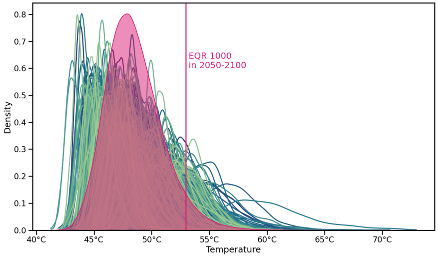
Combination of both methods:

- Account for **estimation error** and **stochastic error**.
- Cover the **design period** $t_1 : t_2$
- For more rare p, more important effect of parameter spread.

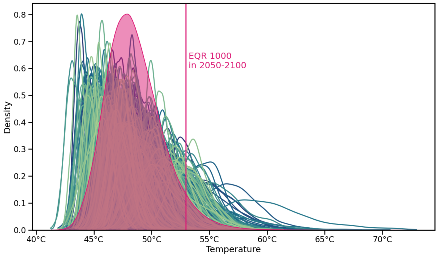




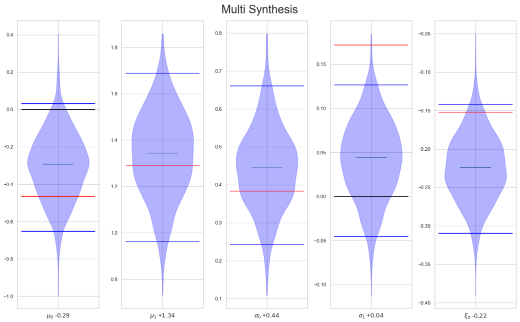
- Unbounded predictive distributions



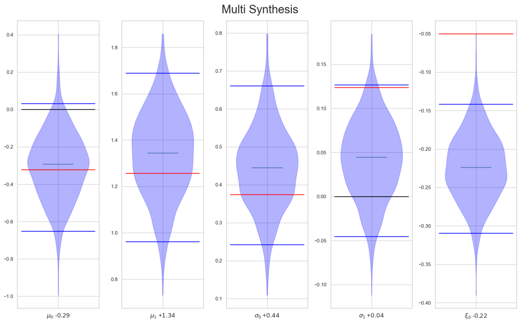
- Unbounded predictive distributions
- Can create **large, physically unrealistic** samples.



- Unbounded predictive distributions
- Can create **large, physically unrealistic** samples.
- **Unlikely but possible** theta θ and covariate as well as extreme event.

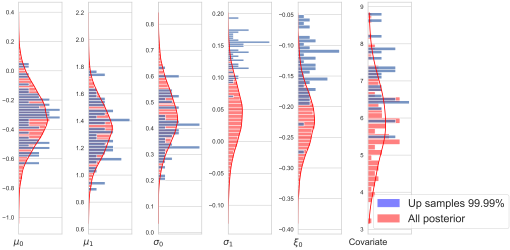


- Unbounded predictive distributions
- Can create **large, physically unrealistic** samples.
- **Unlikely but possible** theta θ and covariate as well as extreme event.
- for example : S083 has X=+8.2°C (mean is +6,1°C)



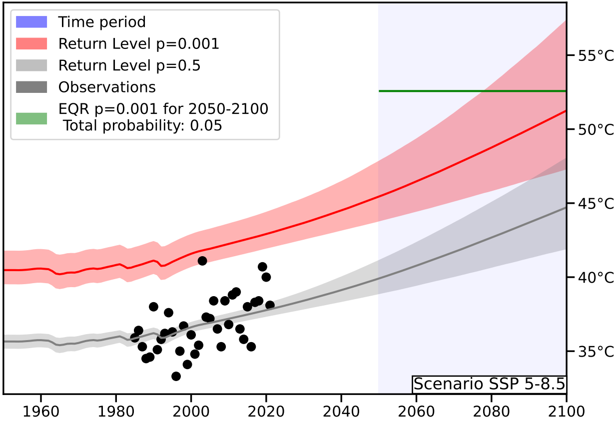
- Unbounded predictive distributions
- Can create **large, physically unrealistic** samples.
- **Unlikely but possible** theta θ and covariate as well as extreme event.

- for example : S070 has X=+6.1°C (mean is +6,1°C)



- Unbounded predictive distributions
- Can create **large, physically unrealistic** samples.
- **Unlikely but possible** theta θ and covariate as well as extreme event.
- Mostly explained by **high** σ_1 **and** ξ.

- **Prior adaptation**:
  - Specification: for eg, add a 0 upper bound on ξ, use something other than a Gaussian, etc.
  - Precision: Add 'expert opinion' weight or other information sources.
  - Hierarchical model: using sources of information like IA downscaling to refine the posterior in various steps
- **Model specification** :
  - Prior on the upper bound
  - Other parameter specification (eg: $\sigma_t = log(1 + exp(\sigma_o + \sigma_1 X_t)))$
- **Theoretical exploration**:
  - Define conditions necessary for a bounded predictive.



Period of interest: **2050-2100**.

For an equivalent return level of 1000 years:

- Predictive is 52.9°C
- Median is 50.3°C with 55.5°C for 95% upper bound.

Interpretation

53°C an annual probability of excess of $\frac{1}{1000}$ over 2050-2100. Similarly, 53°C has a 5% probability of excess over 2050-2100.

- <1a> Chose **Equivalent Reliability** as quantity of interest.
- <1a> Adapted Robin and Ribes' (2020) estimation method and applied it to Tricastin.
- MCMC Algorithm: Large improvement in **time and precision**, comparison with alternatives.
- **Predictive** estimation taking parameter **uncertainty** into account.
- **First paper in writing stage** + Collaboration with Y. Robin on the STAN add-on.
- Many potential **avenues of improvement** (Data, Prior specification, model specification, etc).

Step 1 (July - August -September): **Loose end**
Writing: ER and Predictive paper, Part of Yoann's paper on Stan.
Code: Clean, share, document, and add interesting options like online updating, multi scenario. Account for Yoann's BSAC.
$\Longrightarrow$ **Aim:** Stable and usable version of the method from data to final value.

Step 2 (Autumn and Winter) : **Application**
Application (Various scenarios and point of interest)
Testing various parameterizations and interventions on the prior.
How to evaluate a prior quality?
$\Longrightarrow$ **Aim:** Creating a method for specifying the prior.

Step 3 (Futur):
Depending on Step 2's results, include other sources of information such as IA downscaling.

[1] Chapter 11: Weather and Climate Extreme Events in a Changing Climate.

[2] S. Coles and J. Tawn.
Bayesian modelling of extreme surges on the UK east coast.
Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 363(1831):1387–1406, June 2005.
Publisher: Royal Society.

[3] W. K. Hastings.
Monte Carlo sampling methods using Markov chains and their applications.
Biometrika, 57(1):97–109, Apr. 1970.

[4] M. D. Homan and A. Gelman.
The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.
The Journal of Machine Learning Research, 15(1):1593–1623, Jan. 2014.

[5] Y. Hu, Z. Liang, V. P. Singh, X. Zhang, J. Wang, B. Li, and H. Wang.
Concept of Equivalent Reliability for Estimating the Design Flood under Non-stationary Conditions.
Water Resources Management, 32(3):997–1011, Feb. 2018.

[6] Intergovernmental Panel On Climate Change (Ipcc).
Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
Cambridge University Press, 1 edition, July 2023.

[7] Z. Liang, Y. Hu, H. Huang, J. Wang, and B. Li.
Study on the estimation of design value under non-stationary environment.
South-to-North Water Transfers Water Sci Tech, 14:50–53, 2016.

[8] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller.
Equation of State Calculations by Fast Computing Machines.
The Journal of Chemical Physics, 21(6):1087–1092, June 1953.

[9] Y. Robin and A. Ribes.
Nonstationary extreme value analysis for event attribution combining climate models and observations.
Advances in Statistical Climatology, Meteorology and Oceanography, 6(2):205–221, Nov. 2020.
Publisher: Copernicus GmbH.

Températures maximales en France au 21ème siècle

Merci d'avoir suivi cette présentation
Des questions?

Only need to be able to simulate from conditional distributions. (Maybe possible use of X_T)
Multivariate : $\psi = (\psi_1, \ldots, \psi_d)'$, full conditionnals are $\pi(\psi_i|\psi_{-i}) = \pi_i(\psi_i)$
Description of algorithm:

- Initialisation: k=1, initial state of chain $\psi^{(0)}$
- Boucle: For new value $\psi^{(k)}$:
  - $\psi_1^{(k)} \sim \pi(\psi_1|\psi_{-1}^{(k-1)})$
  - $\psi_2^{(k)} \sim \pi(\psi_2|\psi_{-1,2}^{(k-1)}, \psi_1^{(k)})$
  - ...
  - $\psi_d^{(k)} \sim \pi(\psi_d|\psi_{-d}^{(k)})$

$\pi(\psi)$ is still the density of interest. We now have a transition kernel $p(\psi_{i+1}, \psi_i)$, easy to simulate from, to get successive values.

- Initialisation : k=1,initial state of chain $\psi^{(0)}$
- Boucle: For new value $\psi^{(k)}$:
  - Generate new proposed value ψ' using the kernel transition function.
  - Calculate Acceptance Probability (ratio) $A(\psi^{(k-1)}, \psi')$ of the proposed change of value:

$$A(\psi^{(k)}, \psi') = min\{1, \frac{\pi(\psi')L(\psi'|x)p(\psi', \psi^{(k-1)})}{\pi(\psi^{(k-1)})L(\psi^{(k-1)}|x)p(\psi^{(k-1)}, \psi')}\}$$

  - Accept $\psi^{(k)} = \psi'$ with probability $A(\psi^{(k)}, \psi')$ and keep $\psi^{(k)} = \psi^{(k-1)}$ otherwise.

Based on Bayesian Modelling of Extreme Rainfall Data from Elizabeth Smith
Gibbs concept (each parameter is updated in turn) and conditionals are MH (do we accept the new value produced by the transition function?)
Avantage: Each parameter has his own trajectory (One may not move much and another a lot) + varying transition kernel (proportionnal) (not hard to do for simple MH too)
$\rightarrow$ Less dependance than normal MH?.
Description of algorithm:

- Initialisation : k=1,initial state of chain $\theta^{(0)}$
- Boucle: For new value $\theta^{(k)}$:
  - In turn, for each parameter $\theta_j^{(k)}$
    - $\theta_j^{'} = \theta_j^{(k-1)} + \varepsilon_j$
    - Accept or refuse using $A(\theta_j^{(k-1)}, \theta_j^{'})$ with $\theta_{-j}^{(k)}$ seen as known.

$\pi(\psi)$ is the density of interest and $p(\psi_{i+1}, \psi_i)$ a transition kernel, easy to simulate from, to get successive values.

- Initialisation : k=1,initial state of chain $\psi^{(0)}$
- Loop: For new value $\psi^{(k)}$:
  - Generate new proposed value ψ' using the kernel transition function.
  - Calculate Acceptance Probability (ratio) $A(\psi^{(k-1)}, \psi')$ of the proposed change of value:

$$A(\psi^{(k)}, \psi') = min\{1, \frac{\pi(\psi')L(\psi'|x)p(\psi', \psi^{(k-1)})}{\pi(\psi^{(k-1)})L(\psi^{(k-1)}|x)p(\psi^{(k-1)}, \psi')}\}$$

  - Accept $\psi^{(k)} = \psi'$ with probability $A(\psi^{(k)}, \psi')$ and keep $\psi^{(k)} = \psi^{(k-1)}$ otherwise.

Issue: No objective test for convergence of the chain.

Various criteria of chain quality:

- Acceptance Rate: Ratio of accepted samples. Information on the chain's proper functioning.
- Effective Sample size: Number of independent samples with the same information. Information on autocorrelation within the chain.
- Visual check: Information on mixing quality and visual convergence.

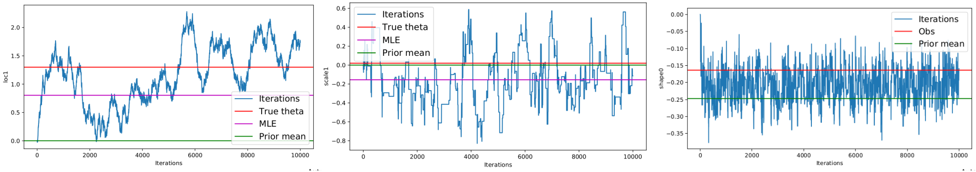


Figure: Examples of trace plots