

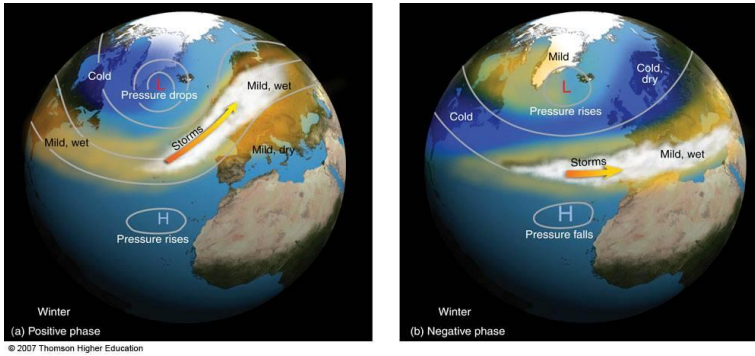


EXEMPLES D'APPROCHES STATISTIQUES APPLIQUÉES AUX VARIABLES CLIMATIQUES

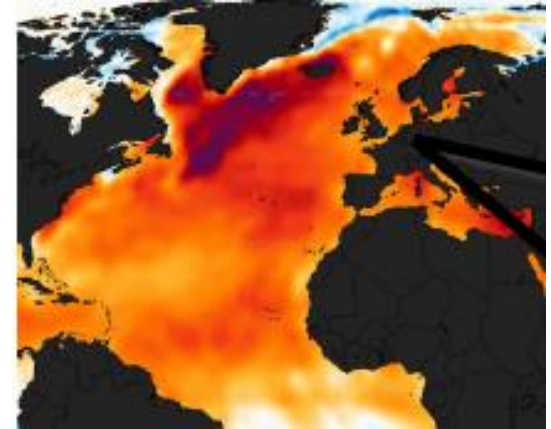


MODULÉE PAR DE GRANDES OSCILLATIONS

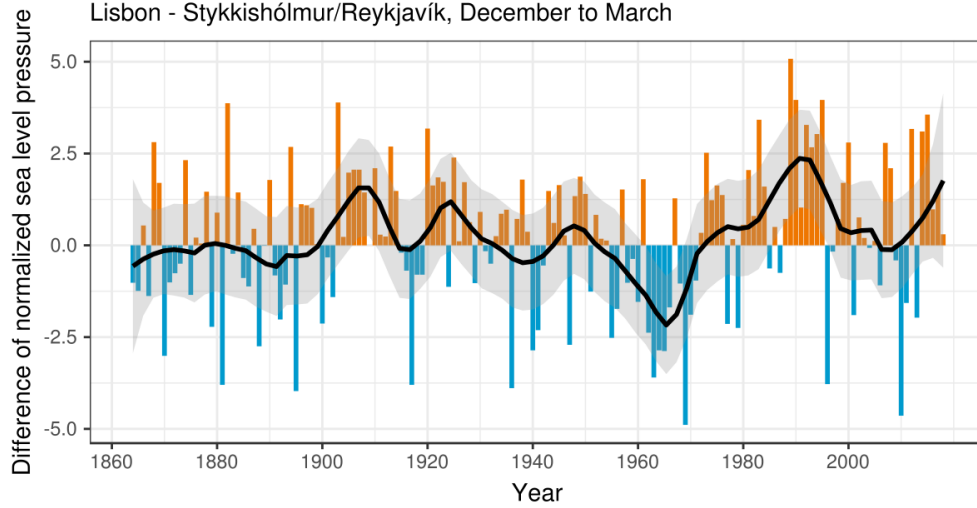
NAO



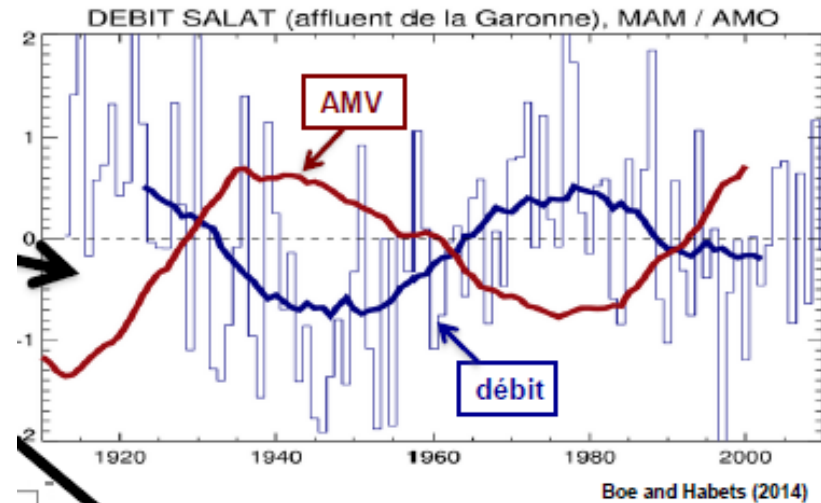
AMV



North Atlantic Oscillation (NAO) winter index
Lisbon - Stykkishólmur/Reykjavík, December to March

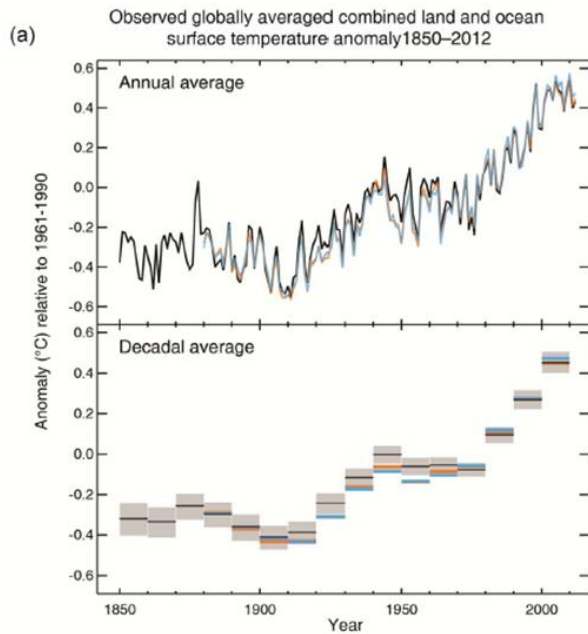


NAO Index Data provided by the Climate Analysis Section, NCAR, Boulder, USA, Hurrell (2003)
Updated regularly. Accessed 2018-10-21

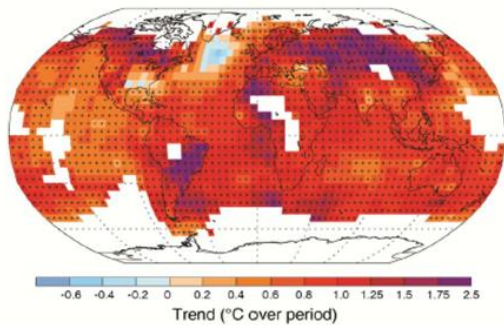


Boe and Habets (2014)

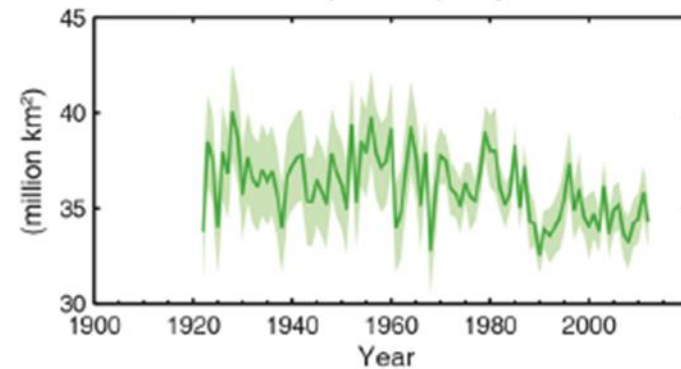
ET QUI ÉVOLUE



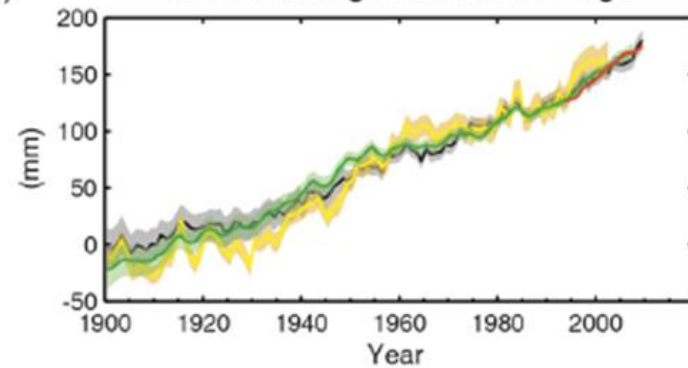
(b) Observed change in average surface temperature 1901–2012



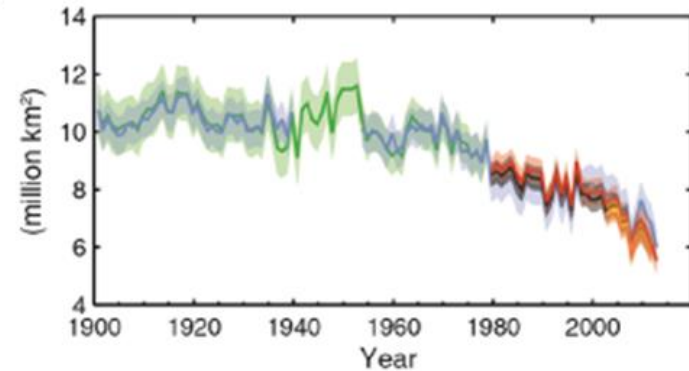
(a) Northern Hemisphere spring snow cover



(d) Global average sea level change



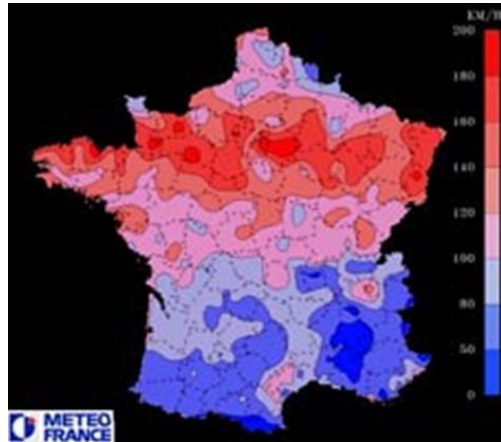
(b) Arctic summer sea ice extent



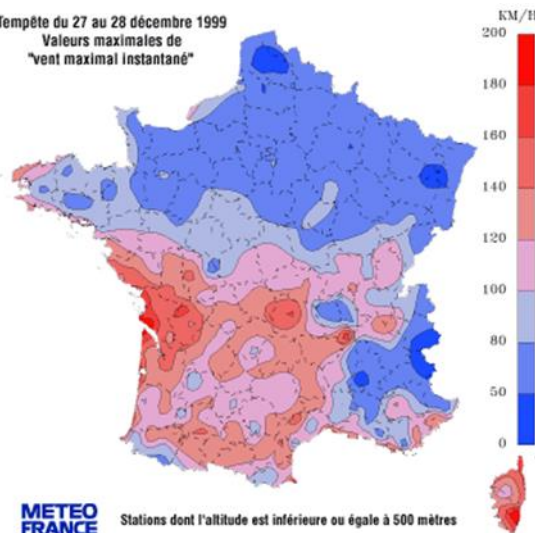
EDF EST ADAPTÉE AUX CONDITIONS HISTORIQUES, MAIS ...

Variabilité et changement climatique

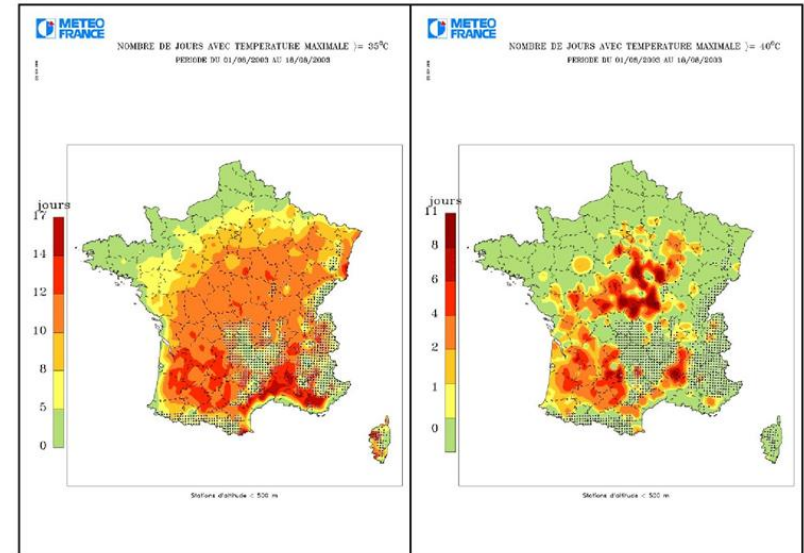
Décembre 1999



Tempête du 27 au 28 décembre 1999
Valeurs maximales de "vent maximal instantané"



Août 2003



	threshold 3% points	Records for period 1950-2002		Records for period 1950-2003	
BORDEAUX	31.99	08/09/98	38.30	08/04/03	40.01
LYON	31.96	07/22/83	39.30	08/13/03	40.49
METZ	29.48	08/11/98	37.86	08/07/03	38.79
MONTELIBAR	33.21	07/06/82	39.49	08/05/03	40.49
ORLEANS	29.98	09/04/52	37.85	08/06/03	39.30
POITIERS	30.69	07/22/90	37.09	08/05/03	39.10
REIMS	29.11	08/11/98	37.07	08/12/03	38.90

THÉORIE STATISTIQUE DES VALEURS EXTRÊMES

▪ Block maxima :

If X_1, \dots, X_n are n iid random variables, then, if there are constant real values a_n and $b_n > 0$ and a non degenerated law G such that :

$$\lim_{n \rightarrow +\infty} P\left(\frac{M_n - a_n}{b_n} \leq x\right) = G(x)$$

Then G belongs to one of the 3 following families:

• II - Fréchet : $G(x) = \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{if } x > 0 \end{cases}$
with $\alpha > 0$

• III - Weibull : $G(x) = \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$
with $\alpha > 0$

• I - Gumbel : $G(x) = \Lambda(x) = \exp(-\exp(-x))$,
with $x \in \mathfrak{R}$

▪ Threshold excesses:

An equivalent theorem as for maxima states that when u tends to infinity, then the survival distribution of the exceedances tends to a Generalized Pareto Distribution:

$$\bar{H}(y) = \begin{cases} (1 + \xi \frac{y}{\tilde{\sigma}})^{-1/\xi} & \text{si } \xi \neq 0 \\ \exp(-\frac{y}{\tilde{\sigma}}) & \text{si } \xi = 0 \end{cases}$$

defined on \mathfrak{R}^+ if $\xi \geq 0$ or on $[0; -\tilde{\sigma}/\xi[$ if $\xi < 0$.

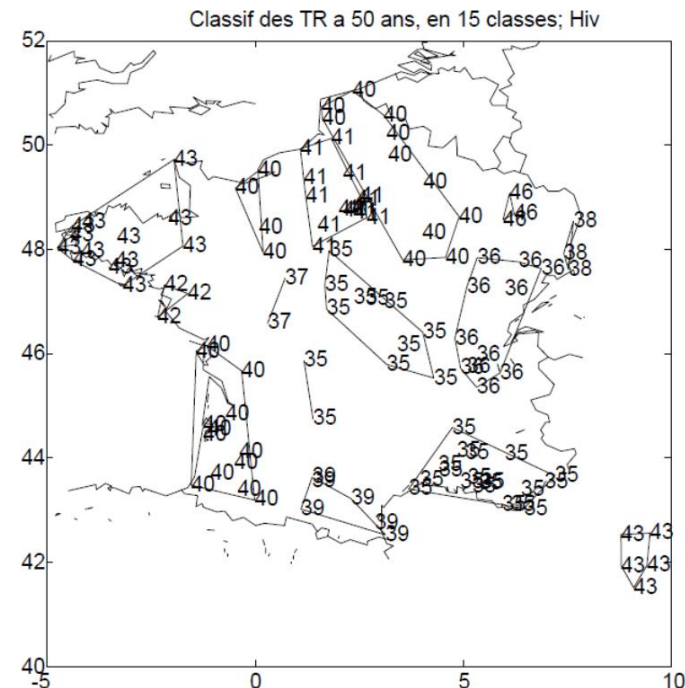
▪ It is a 2 parameter function :

- $\tilde{\sigma} > 0$ is the scale parameter.
- $\xi \in \mathfrak{R}$ is the shape parameter.

◆ Remark: the exceedance dates are uniformly distributed

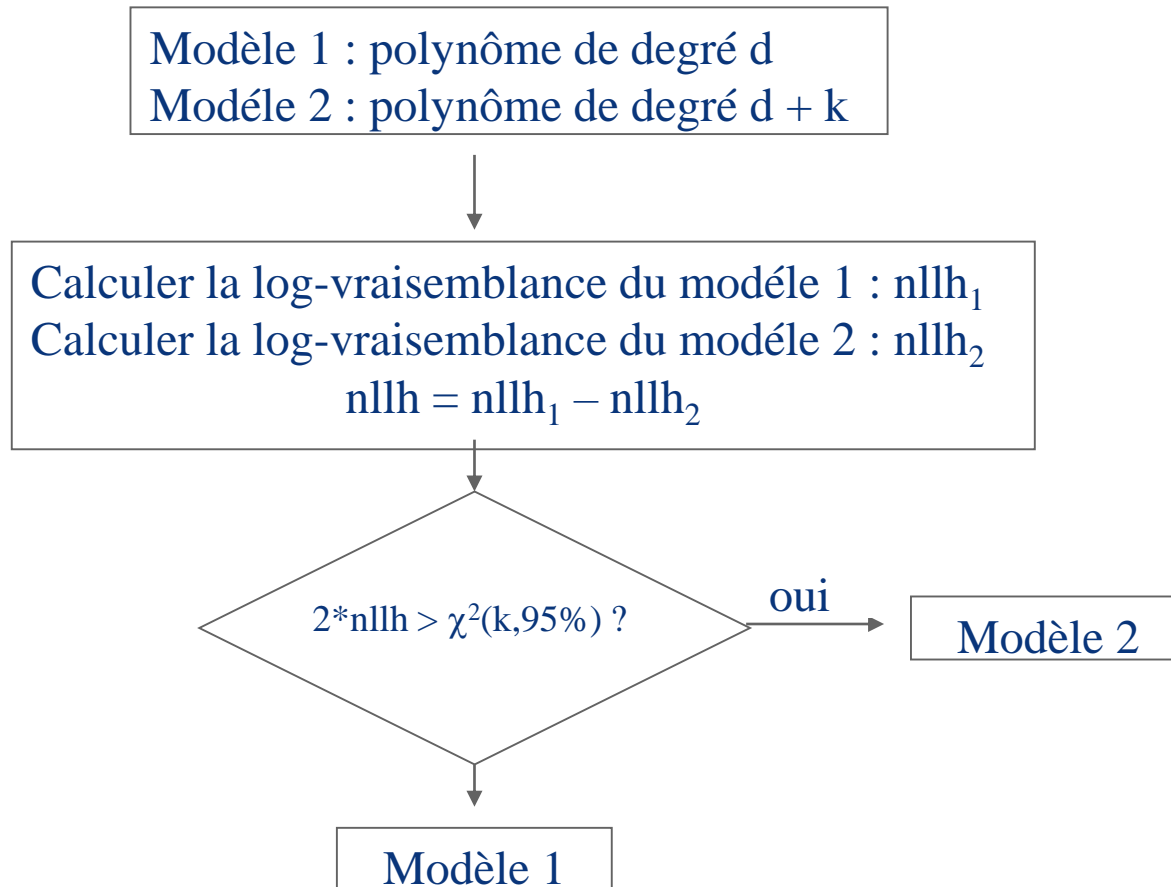
APPLICATION AUX VITESSES DE VENT

- Données quotidiennes de vitesse de vent instantanée maximale (en m/s) mesurées avec une constante d'intégration de 0,5 s.
 - 120 stations du réseau synoptique de Météo-France
 - 1er janvier 1981 au 31 décembre 2000 (7305 jours)
- Préparation des données :
 - jitterisation des données brutes par ajout d'un bruit uniforme sur $[-0,5 \text{ m/s}; 0,5 \text{ m/s}]$
 - déclusterisation (un cluster est une succession de jours de grands vents) selon une règle à 1 jour : on remplace tout cluster par le jour du VMI conservant, pour ce jour, sa vitesse
 - estimation par saison pour garder une équidistribution des variables
- Méthodes :
 - méthode de valeurs extrêmes GEV (Generalized Extreme Value)
 - méthode de dépassement de seuil POT (Peak over Thresholds)



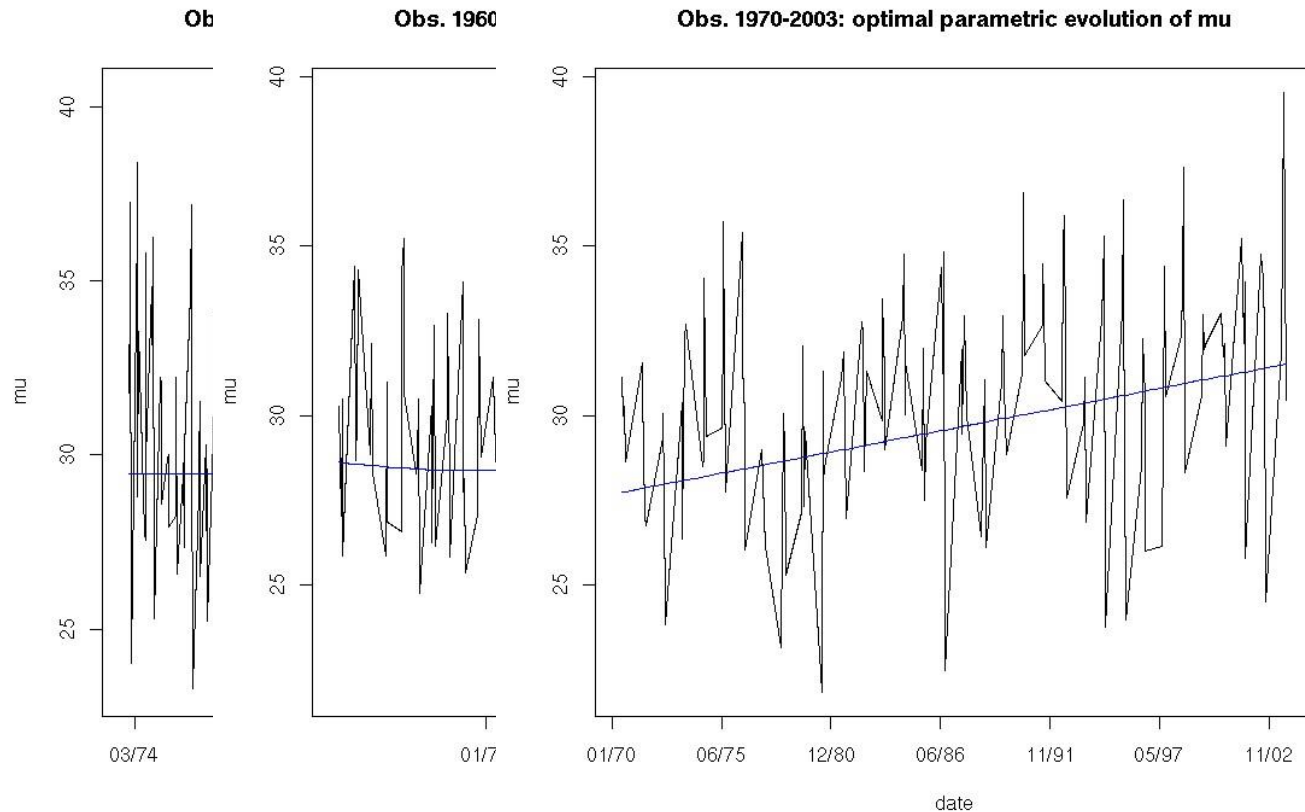
ET S'IL Y A UNE TENDANCE?

- 1ères réflexions en 2002 avec D. Dacunha-Castelle: identification et extrapolation de tendances des paramètres des lois



S. Parey, F. Malek, C. Laurent, D. Dacunha-Castelle: Trends and climatic evolution: Statistical approach for very high temperatures in France, Climatic Change (2007) 81:331 - 352

MAIS LA TENDANCE DÉPEND DE LA PÉRIODE



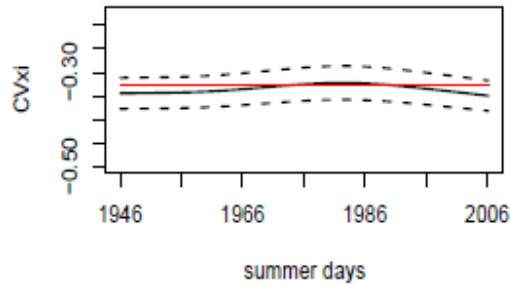
GEV

POT

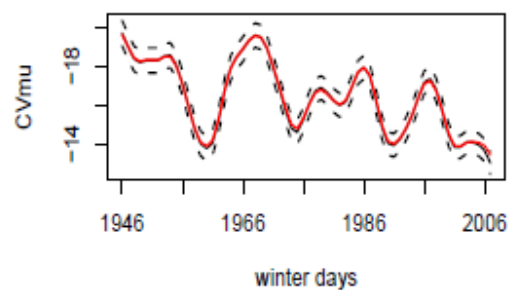
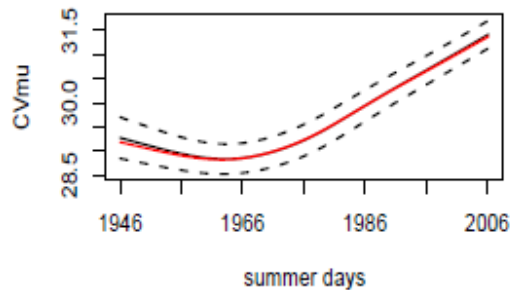
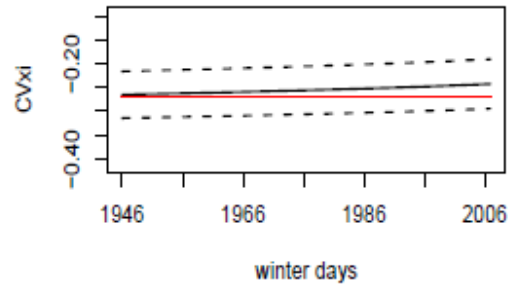
Period	P10	P21	P22	P31	P32	P33
$l(t)$ degree	5	0	2	0	0	2
$\sigma(t)$ degree	1	0	0	0	4	0

QUELS PARAMÈTRES ÉVOLUENT?

La Rochelle_Summer

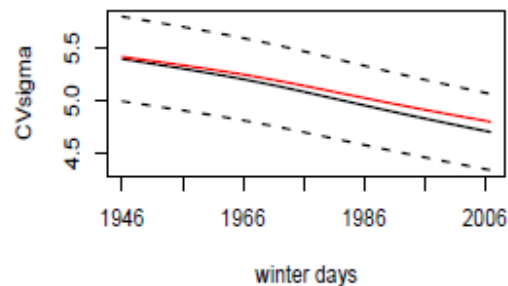
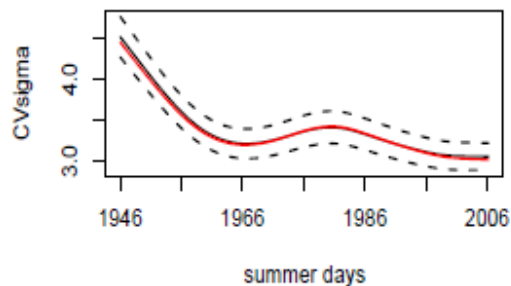


Uman_Winter



Noir: paramètre de forme variable

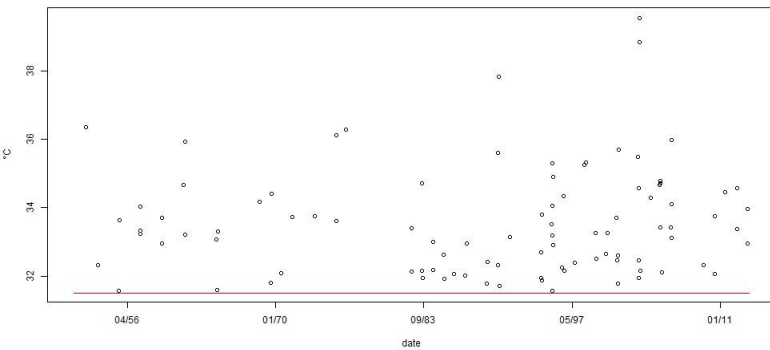
Rouge: paramètre de forme constant



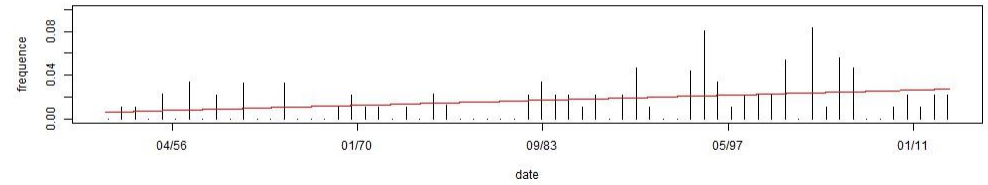
QUELS PARAMÈTRES ÉVOLUENT?

Seuil constant

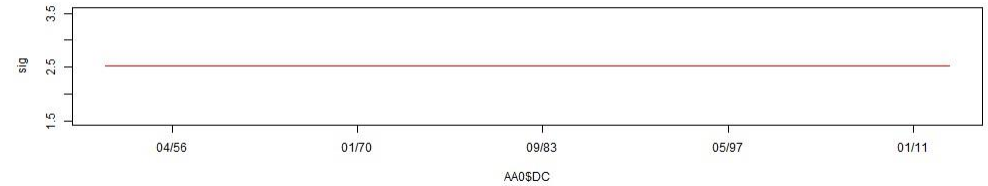
Seuil constant



Tendance Poisson

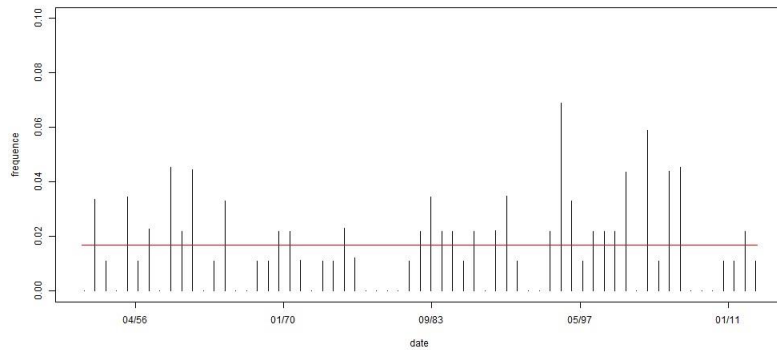


Tendance Pareto

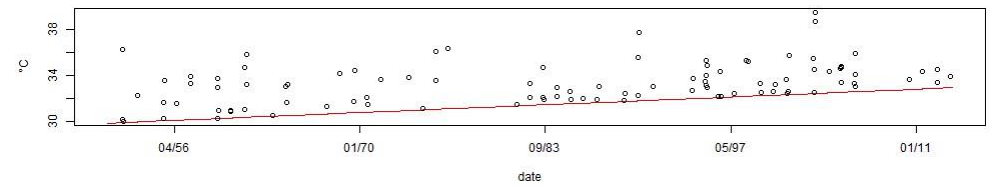


Seuil variable

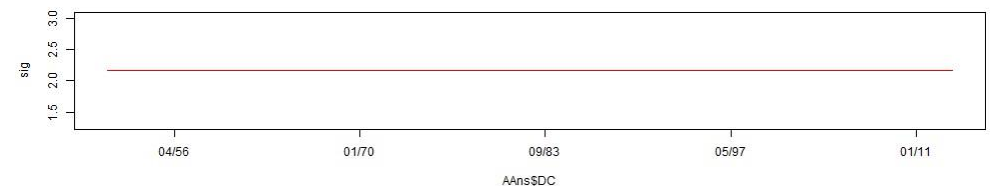
Tendance Poisson



Tendance Seuil



Tendance Pareto



ET IL FAUT REDÉFINIR UN NIVEAU DE RETOUR

- Valeur atteinte ou dépassée en moyenne 1 fois tous les N ans => **valeur dont l'espérance d'être atteinte ou dépassée dans les N prochaines années est 1**

$$\frac{1}{nb} \sum_{t=t_0}^{t_0+365a} \left\{ 1 - \exp \left[- \left(1 + \frac{\xi}{\sigma(t)} (z_a - \mu(t)) \right)^{-1/\xi} \right] \right\} = 1 \quad \text{GEV}$$

$$\text{POT} \quad \sum_{t \in D(t_0, a)} \left(1 + \frac{\xi}{\sigma(t)} (z_a - u) \right)^{-1/\xi} I(t) = 1$$

- Il dépend du choix de la tendance
- Intervalle de confiance: prise en compte de l'incertitude sur la tendance, choix d'une méthode de bootstrap
- D'autres définitions depuis:

- ENE:
$$\sum_{t=T_1}^{T_1+m-1} \{ 1 - G_Z(z^{ENE}(m) | \theta_t) \} = 1$$

- RE:
$$RE_{T_1-T_2}^{NE} = \prod_{t=T_1}^{T_2} (1 - p_t) = \prod_{t=T_1}^{T_2} G_Z(z_q | \theta_t)$$

- DLL:
$$z_{T_1-T_2}^{DLL}(m) = F_{T_1-T_2}^{-1}(1 - 1/m) \quad \text{with} \quad F_{T_1-T_2}(z) = \prod_{t=T_1}^{T_2} G_{Z,t}(z) = \prod_{t=T_1}^{T_2} G_Z(z | \theta_t)$$

- ADLL:
$$\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} G_Z(z_{T_1-T_2}^{ADLL}(m) | \theta_t) = 1 - 1/m$$

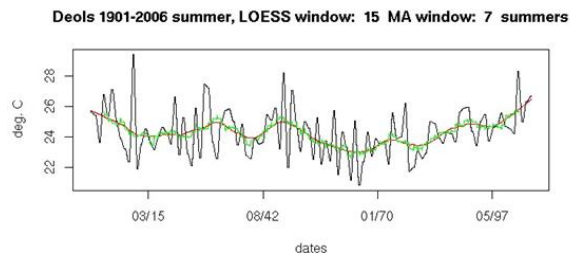
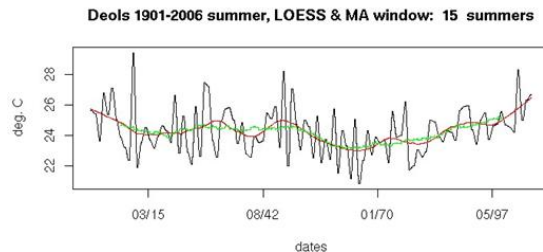
LIMITES DE L'EXTRAPOLATION \Rightarrow NOUVELLE APPROCHE

- Etude fine des rapports entre moyenne, variance et extrêmes: thèse de Thi Thu Huong Hoang (2010)

- Analyse des évolutions de moyenne, variance et extrêmes: méthodes non paramétriques

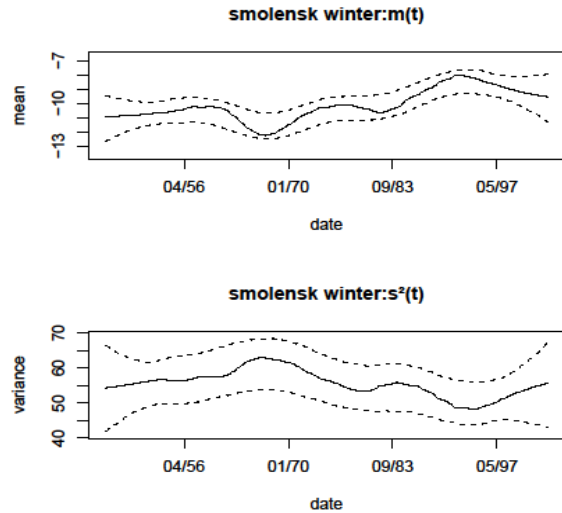
- splines cubiques $l - \frac{1}{2} \lambda \int (f''(t))^2 dt$

- Loess: régression locale: la taille de la fenêtre est contrôlée par le paramètre de lissage $\lambda < 1$; elle contient la proportion λ des points de la série totale et applique une pondération tri cubique proportionnelle à $(1 - (\text{dist}/\text{maxdist}))^3$

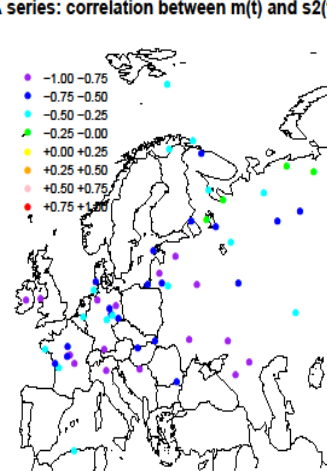


ÉVOLUTIONS DE MOYENNE ET DE VARIANCE

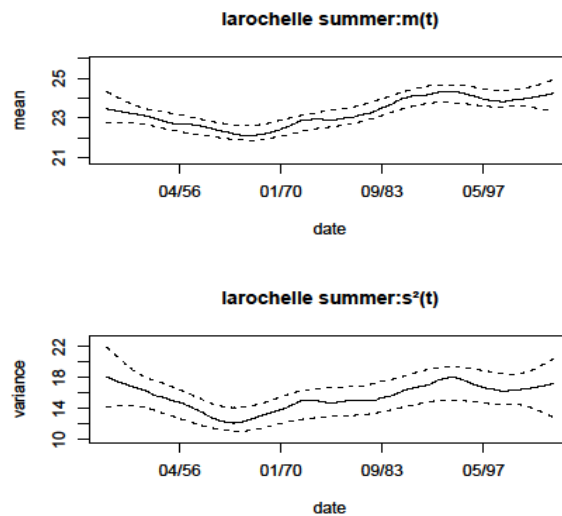
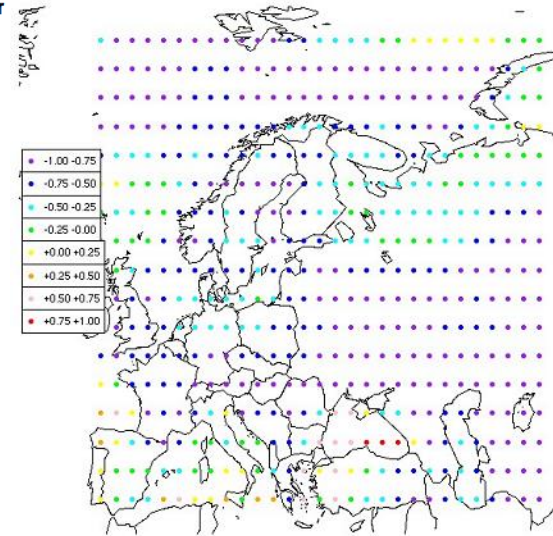
HIVER



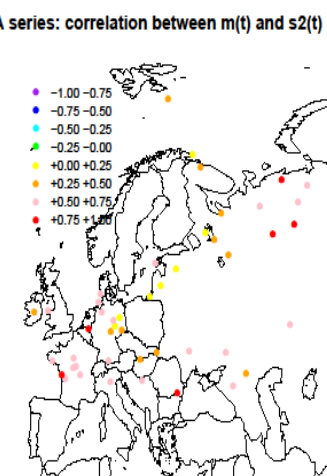
55 ECA series: correlation between m(t) and s²(t) in winter



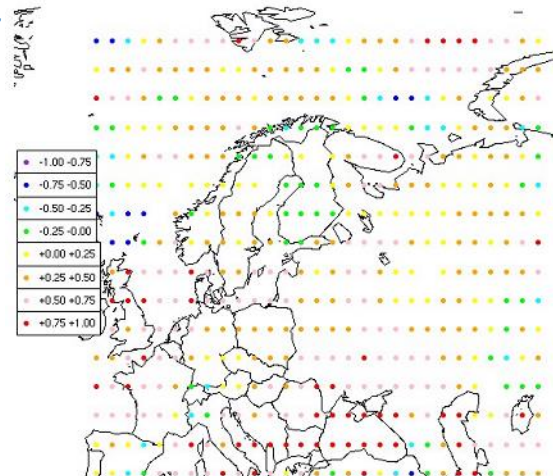
ERA40: correlations in winter



55 ECA series: correlation between m(t) and s²(t) in summer

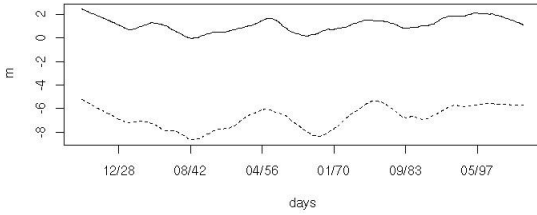


ERA40: correlations in summer

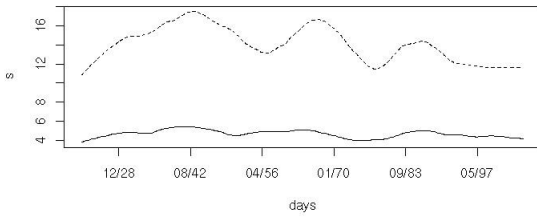


LIEN AVEC LES EXTRÊMES

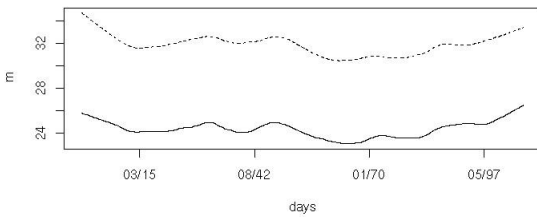
mean function cold temperature in deols



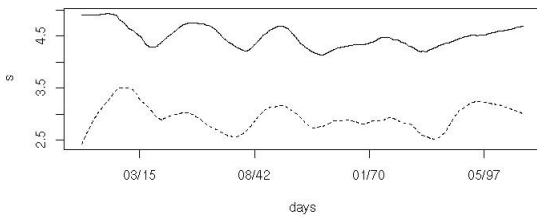
scale function cold temperature in deols



mean function hot temperature in deols

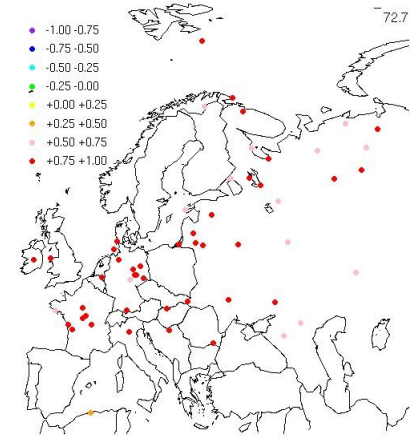


scale function hot temperature in deols

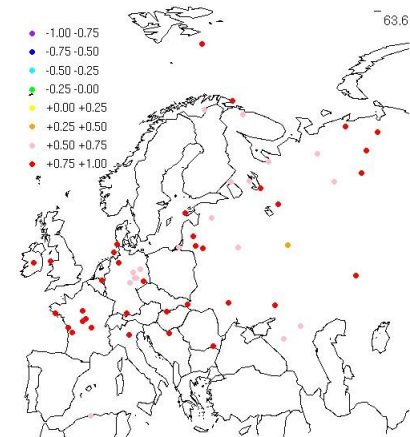


HIVER

55 ECA series: correlation between $m(t)$ and $mx(t)$ in winter



55 ECA series: correlation between $m(t)$ and $mx(t)$ in summer



ETE

STATIONNARISATION DES EXTRÊMES?

Hypothèse :

$$Y_t = \frac{X_t - m_t}{s_t}$$

les extrêmes de Y_t sont stationnaires

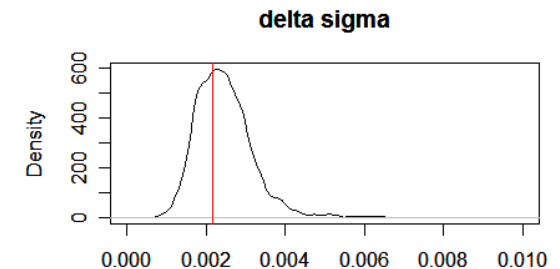
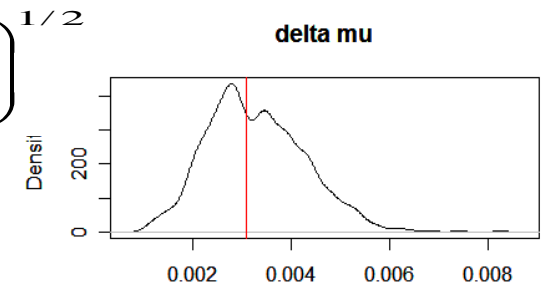
Test

1) définir une distance entre 2 fonctions du

temps: $\Delta(f, g) = \frac{1}{T} \left(\int (f(t) - g(t))^2 dt \right)^{1/2}$

2) calculer une table statistique des distances si l'hypothèse est vraie

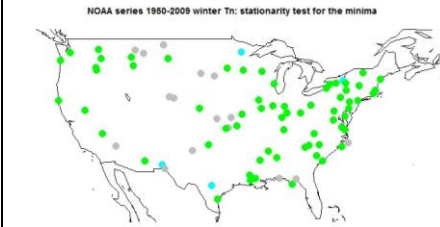
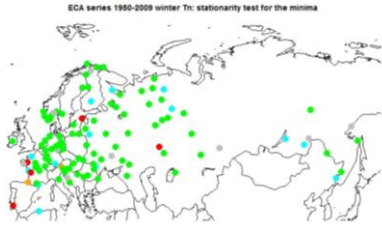
- 1000 échantillons d'une GEV stationnaire (ξ_Y, μ_Y, σ_Y)
- Estimer les paramètres:
 - 1) constants
 - 2) dépendant du temps
- Calculer Δ



RÉSULTATS POUR DIFFÉRENTES LOCALISATIONS

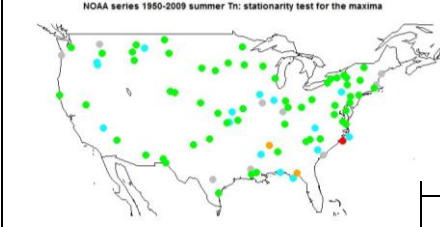
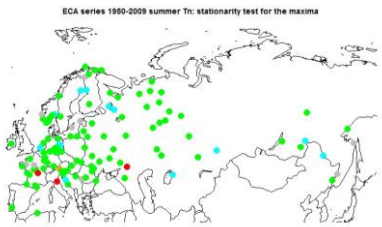
Minimum Winter TN

a)

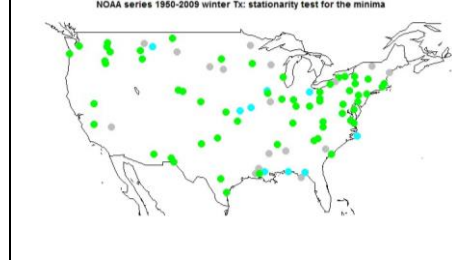
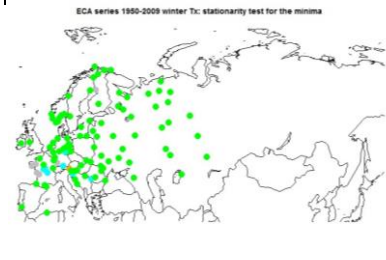


Maximum Summer TN

b)

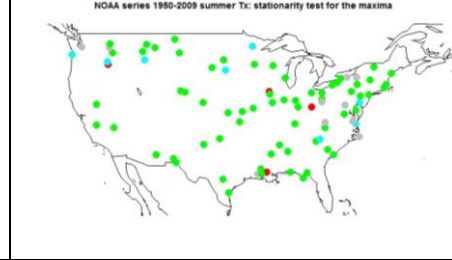
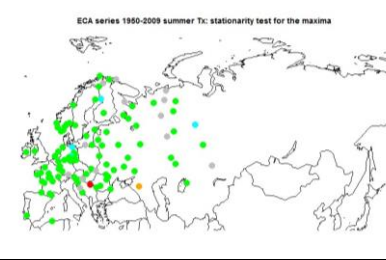


c)



Maximum Summer TX

d)



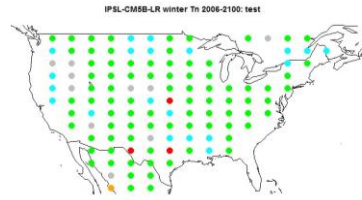
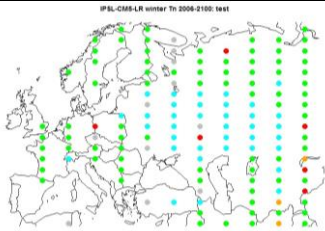
● TRUE for μ & σ ● TRUE for μ only ● TRUE for σ only ● FALSE ● non convergence



VALIDITÉ POUR LE CLIMAT FUTUR

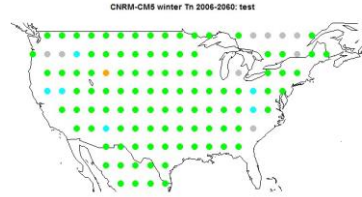
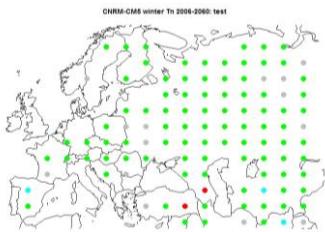
Minimum Winter TN: IPSL-CM5-LR

a)



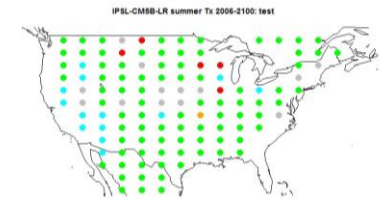
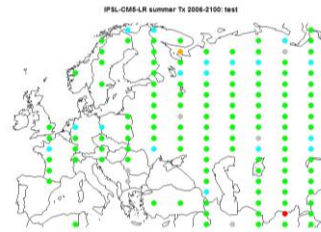
Minimum Winter TN: CNRM-CM5

b)



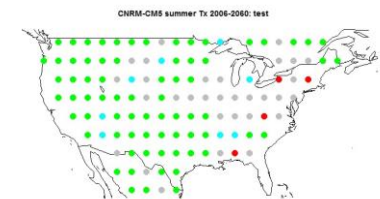
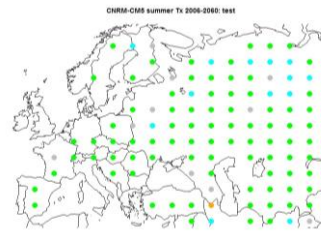
Maximum summer TX: IPSL-CM5-LR

c)



Maximum summer TX: CNRM-CM5

d)



• TRUE for μ & σ • TRUE for μ only • TRUE for σ only • FALSE • non convergence

MODÈLE STOCHASTIQUE DE TEMPÉRATURE

Hoang, T. T. H. (2010), Modélisation de séries chronologiques non stationnaires, non linéaires: application à la définition des tendances sur la moyenne, la variabilité et les extrêmes de la température de l'air en Europe <http://www.tel.archivesouvertes.fr/tel-00531549/fr/>

Dacunha-Castelle D., Hoang T.T.H., Parey S.: Modeling of air temperatures: preprocessing and trends, reduced stationary process, extremes, simulation, Journal de la Société Française de Statistique, 2013

Parey S., Hoang T.T.H., Dacunha-Castelle D. (2013): Validation of a stochastic temperature generator focusing on extremes and an example of use for climate change, Climate Research

PRE-PROCESSING

- Le but est d'enlever les parties déterministes (tendances et saisonnalités) pour obtenir un processus stationnaire
- Le traitement est basé sur des méthodes paramétriques et non paramétriques

$$X(t) = m(t) + S(t) + s(t)S_v(t)Z(t)$$

– $m(t), s(t)$: tendances de moyenne et écart-type, $S(t), S_v(t)$: saisonnalités

– **Estimation :**

» estimer $m(t)$ par loess, $S(t)$ par une fonction trigonométrique à partir de la série $X(t)$, puis $s(t)$ par loess et $S_v^2(t)$ par une fonction trigonométrique à partir de $[X(t) - \hat{m}(t) - \hat{S}(t)]^2$

» Pour $m(t), s(t)$, la cross-validation partitionnée modifiée ⁽¹⁾ est utilisée, pour $S(t), S_v^2(t)$ le critère Akaike

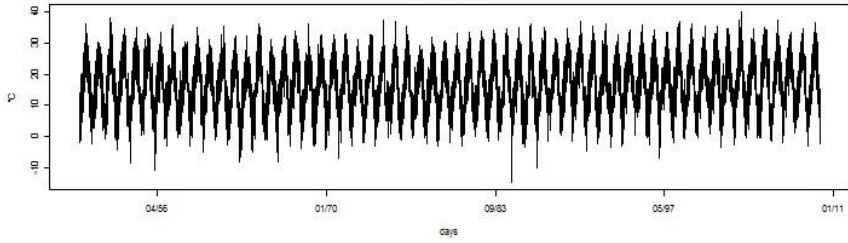
» La série réduite:

$$Z_t = (X_t - \hat{m}_t - \hat{S}_t) / (\hat{s}_t \hat{S}_{v_t})$$

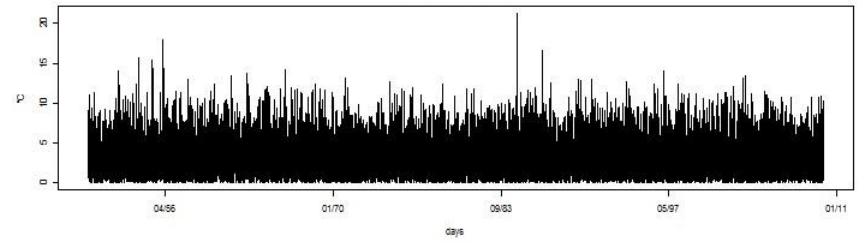
⁽¹⁾ *Modified partitioned CV: new algorithm for correlated data (thesis of Hoang, 2010)*

ILLUSTRATION

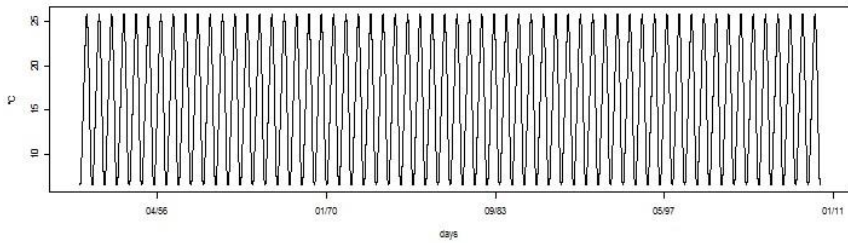
DEOLS: X



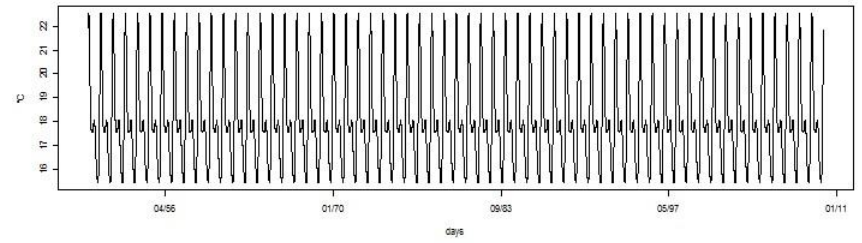
DEOLS: X-Sm-m



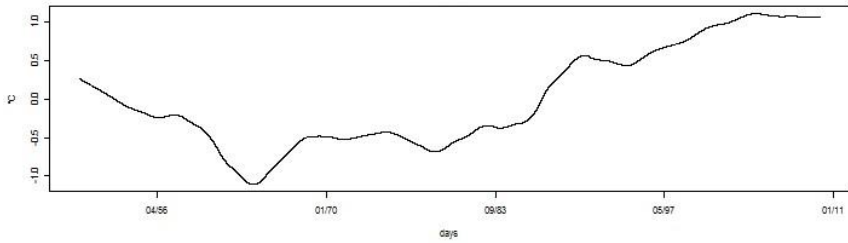
DEOLS: Sx



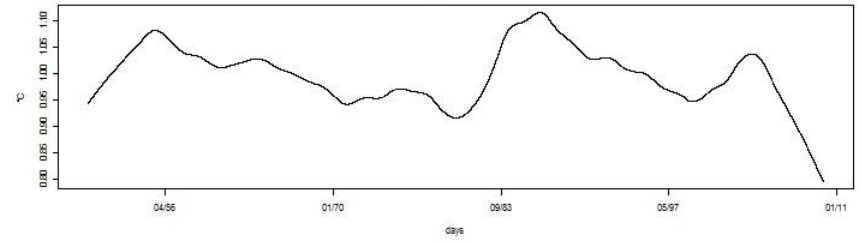
DEOLS: Sv2



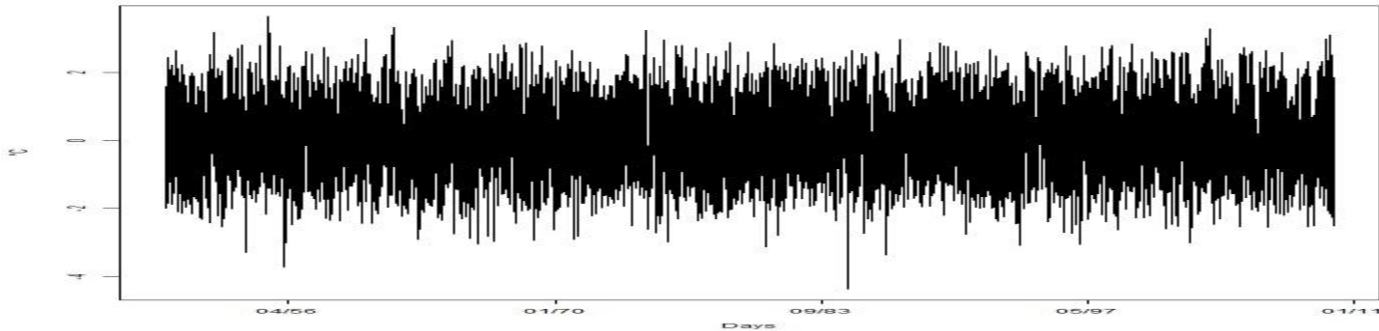
DEOLS: m



DEOLS: s2

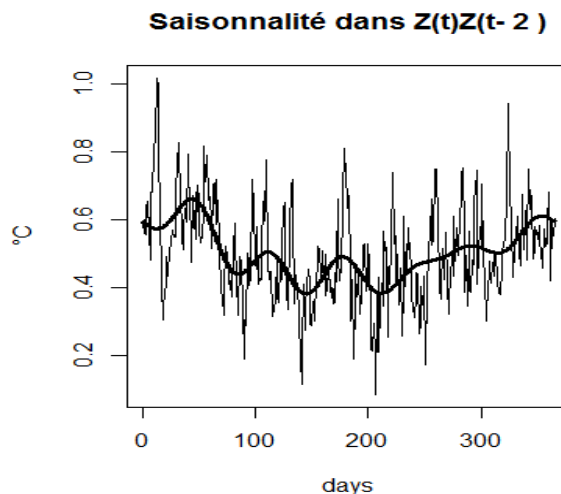
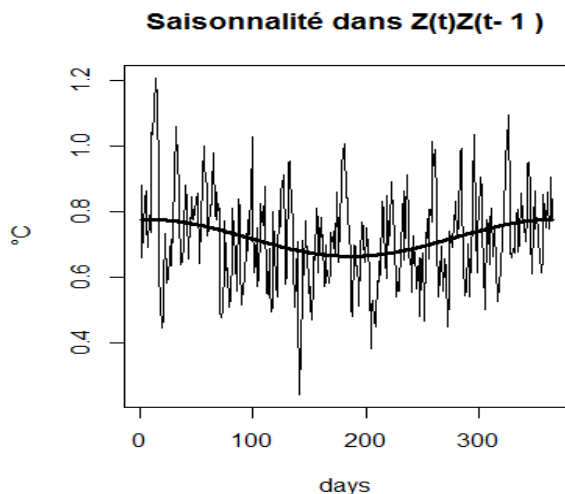


DEOLS: Z



CARACTÉRISTIQUES DE LA SÉRIE RÉDUITE

- courte mémoire
- il reste de la saisonnalité dans la corrélation et la volatilité
- cyclo-stationnaire
- borné dans les queues
- non linéaire
- La volatilité dépend de l'état
- Etudes:
 - tests de tendances⁽²⁾ sur les statistiques de base de Z_t (moyenne, variance, skewness, kurtosis)
 - test de cyclo-stationnarité⁽³⁾ des extrêmes
 - Analyse des tendances et saisonnalités des séries $Z_t Z_{t-k}$

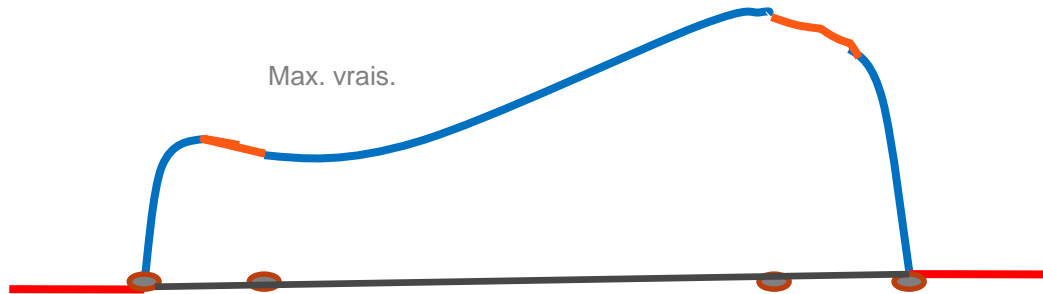


PRINCIPE DES TESTS DE STATIONNARITÉ

- Modèle considéré: $X(t) = \theta(t) + \varepsilon(t)$, la distribution de ε est connue ou non
- **Hypothèse du test:** θ est constant / θ n'est pas constant
- Si:
 - \hat{c}_n est un estimateur constant de θ (par m.l.e si la loi de ε est connue, par moindres carrés sinon)
 - $\hat{\theta}_n$ est un estimateur non paramétrique de θ (par splines si la loi de ε est connue, par loess sinon)
- **but** : comparer ces 2 estimateurs via une distance L^2 :
$$\Delta = \left\| \hat{\theta}_n - \hat{c}_n \right\|$$
- *En pratique* : test sur Δ (construire une **distribution empirique de Δ sous H_0** par simulation si la distribution de θ est connue ou par permutation (ou bootstrap par blocs) sinon)
- **tests de tendances** dans les **moments**: moyenne, variance, skewness, kurtosis, correlation, et les **extremes**

TYPE DE MODÈLE

Forme de la variance conditionnelle



Modèle FARCH (Functional AutoRegressive conditional Heteroscedastic)

Approximation (schéma d'Euler d'ordre 1) de la chaîne de Markov discrète donnée par les observations discrètes d'une diffusion continue avec les mêmes coefficients

LE MODÈLE SFHAR (SEASONAL FUNCTIONAL HETEROSCEDASTIC AUTOREGRESSIVE)

➤ $Z_t = b(Z_{t-1}) + a(Z_{t-1})\varepsilon_t, \quad \varepsilon_t \propto N(0,1)$

➤ Extension: modèle SFHAR

$$Z(t) = \left[\theta_{0,k} + \sum_{j=1}^{p_1} \left(\theta_{1,k}^j \cos \frac{2j\pi t}{365} + \theta_{2,k}^j \sin \frac{2j\pi t}{365} \right) \right] Z(t-1) + a(t, Z_{t-1})\varepsilon_t$$

$\varepsilon_t \propto N(0,1)$

➤ Estimer $a^2(t, Z_{t-1})$ avec contraintes:

- Zero hors des frontières
- positif
- contraintes C sur les premières dérivées du processus en temps continu

$$(a^2)'(r_1) = \frac{2b(r_1, t)}{1-1/\xi_1} \quad \text{et} \quad (a^2)'(r_2) = \frac{2b(r_2, t)}{1-1/\xi_2}$$

- obtention de \hat{a}

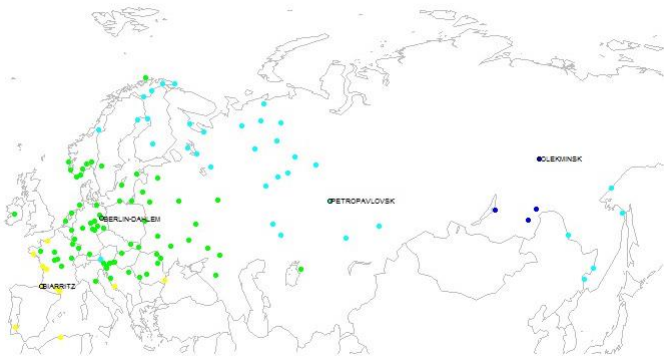
$$\left\{ \begin{array}{l} \hat{a}^2(t, Z_{t-1}) = (\hat{r}_2 - t)(t - \hat{r}_1) \sum_{k=0}^5 \sum_{j=1}^{p_2} \left(\alpha_{1,k}^j \cos \frac{2j\pi t}{365} + \alpha_{2,k}^j \sin \frac{2j\pi t}{365} \right) Z_{t-1}^k \\ C(\hat{r}_1, t), C(\hat{r}_2, t) \\ \hat{a}^2(t) > 0 \quad \forall t \end{array} \right.$$

PROCÉDURE D'ESTIMATION

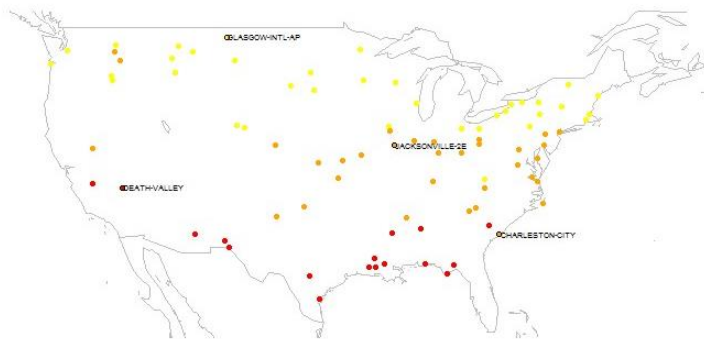
- Estimation de la partie auto-régressive (AR(1))
- Choix du nombre de sinus et cosinus par un critère Akaike
- Estimation de la volatilité par maximum de vraisemblance avec contraintes
 - Valeurs initiales obtenues par estimation basée sur les moindres carrés
 - Maximum de vraisemblance: estimation sous contraintes
- Simulation de $Z(t)$
- Obtention de $X(t)$ par ré-introduction des tendances et saisonnalités

VALIDATION: CHOIX DE DIFFÉRENTS CLIMATS

1950-2009 ECA series TN: annual mean temperature



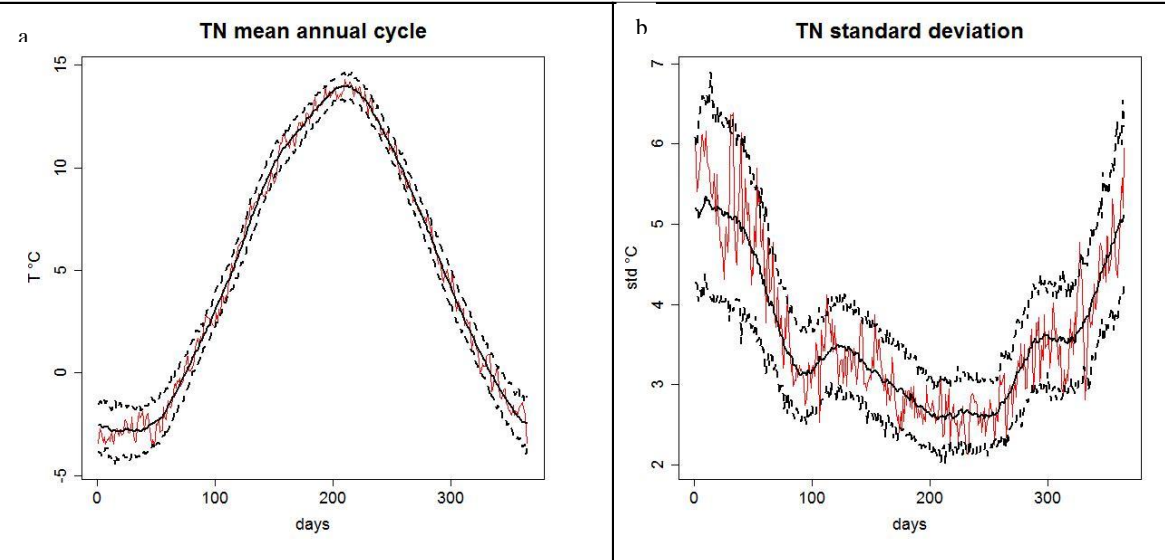
NOAA series TX annual mean temperature



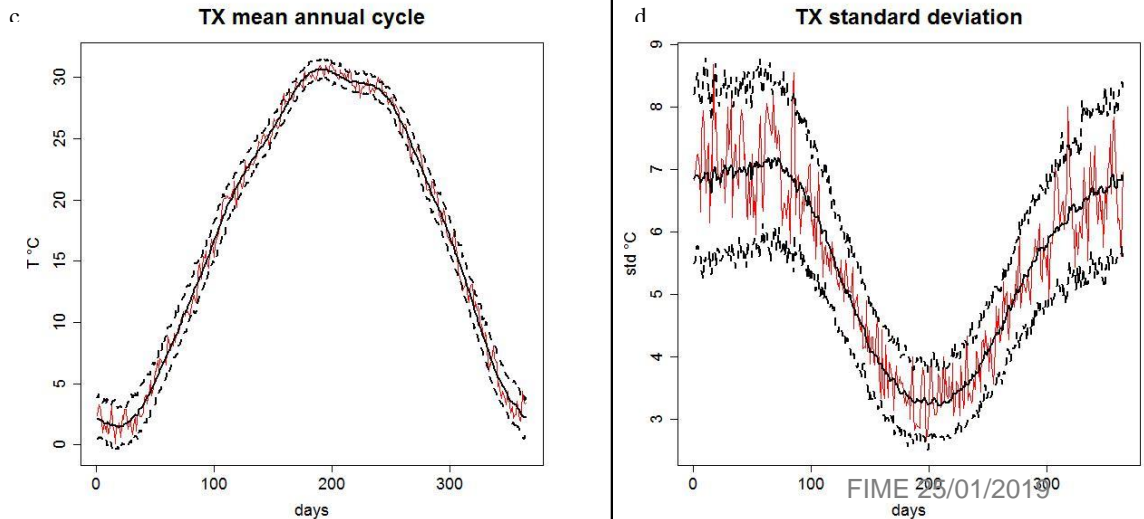
Weather station	Daily minimum temperature TN		Daily maximum temperature TX	
	period	Mean annual mean (°C)	period	Mean annual mean (°C)
Biarritz	1956-2009	10.1	1956-2009	17.7
Berlin	1950-2009	5.1	1950-2009	13.4
Petropavlovsk	1950-2009	-3.3	1950-2009	6.9
Olekminsk	1950-2009	-11.3	-	-
Death Valley	1962-2009	17.0	1962-2009	32.8
Charleston	1950-2009	15.4	1950-2009	23.0
Jacksonville	1950-2009	5.2	1950-2009	17.5
Glasgow	1950-2009	-0.7	1950-2009	12.5

CYCLES ANNUELS MOYENS

BERLIN: daily minimum temperature TN



JACKSONVILLE: daily maximum temperature TX



EXTRÊMES

	TX		TN	
	observations	simulations	observations	simulations
Berlin	38.2 [37.1;39.2]	39.8 [38.8;41.0]	-23.4 [-25.5;-21.0]	-26.5 [-31.5;-22.9]
Biarritz	39.6 [38.8;40.4]	41.0 [39.0;43.5]	-9.4 [-12.2;-6.6]	-11.0 [-12.6;-9.7]
Petropavlovsk	38.5 [37.6;39.5]	41.5 [39.3;44.8]	-43.7 [-45.2;-42.1]	-48.7 [-52.5;-45.3]
Olekminsk	-	-	-56.3 [-57.8;-54.8]	-58.8 [-61.4;-56.2]
Death Valley	54.3 [53.5;55.1]	55.2 [54.3;56.1]	-6.4 [-7.5;-5.3]	-7.4 [-8.8;-6.0]
Jacksonville	41.8 [40.3;43.3]	43.1 [41.5;44.5]	-29.5 [-31.3;-27.7]	-33.8 [-38.5;-30.6]
Glasgow	42.0 [41.1;42.8]	45.5 [44.3;46.9]	-42.9 [-44.4;-41.4]	-46.9 [-50.4;-44.0]
Charleston	39.5 [38.6;40.4]	40.3 [39.5;41.2]	-11.3 [-13.7;-9.0]	-8.8 [-10.0;-7.5]

APPLICATION EN CONTEXTE DE CHANGEMENT CLIMATIQUE

Choix de 2 stations présentant une rupture en milieu de période

	m_1 (°C)	m_2 (°C)	s_1 (°C)	s_2 (°C)
TN Berlin	4.7	5.5	7.0	6.9
TX Death Valley	32.3	33.2	10.4	10.7

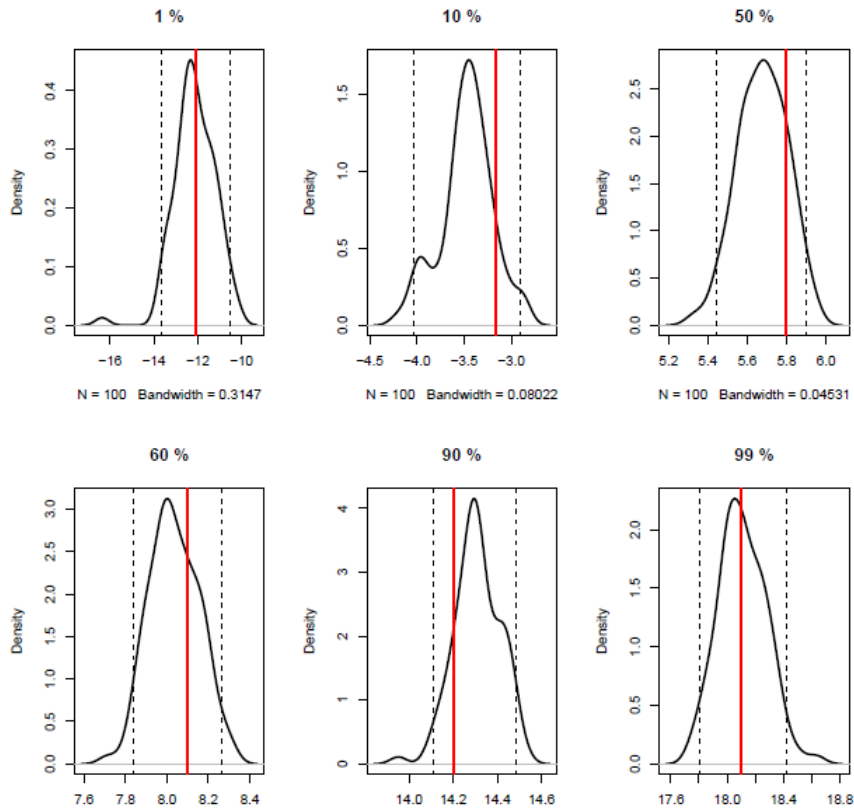
Estimation des paramètres du modèle sur la 1^{ère} période

Reconstruction sur la 2^{ème} période

$$X_2(t) = \hat{S}_{m_1}(t) + \hat{m}_1(t) + (m_2 - m_1) + \hat{S}_{v_1}(t) * \hat{s}_1(t) * \frac{s_2}{s_1} * Z(t)$$

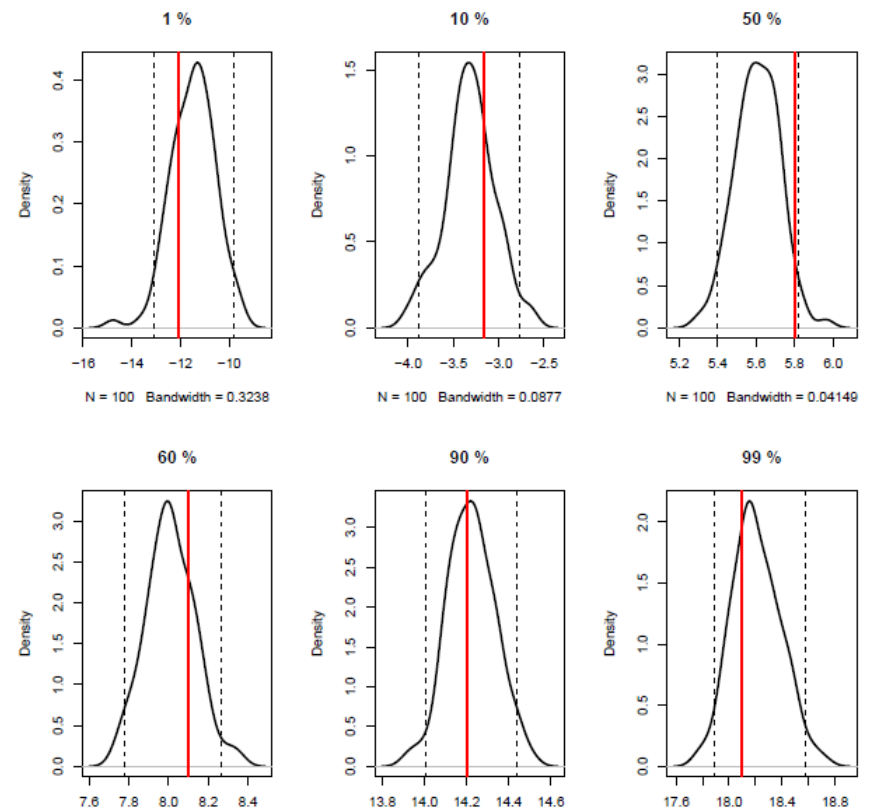
$$X_2(t) = \hat{S}_{m_2}(t) + \hat{m}_2(t) + \hat{S}_{v_2}(t) * \hat{s}_2(t) * Z(t)$$

DISTRIBUTION



1^{ère} méthode

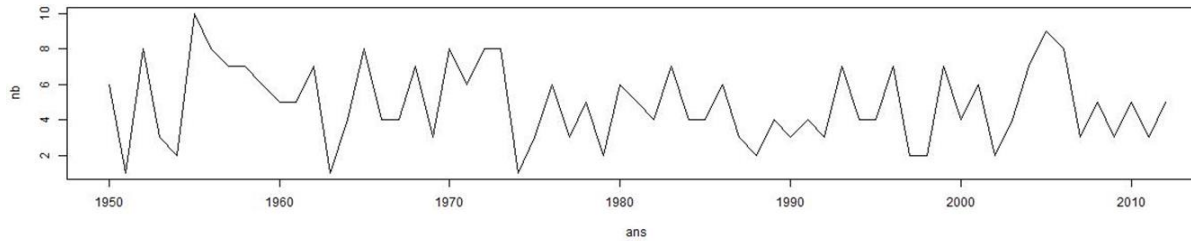
BERLIN



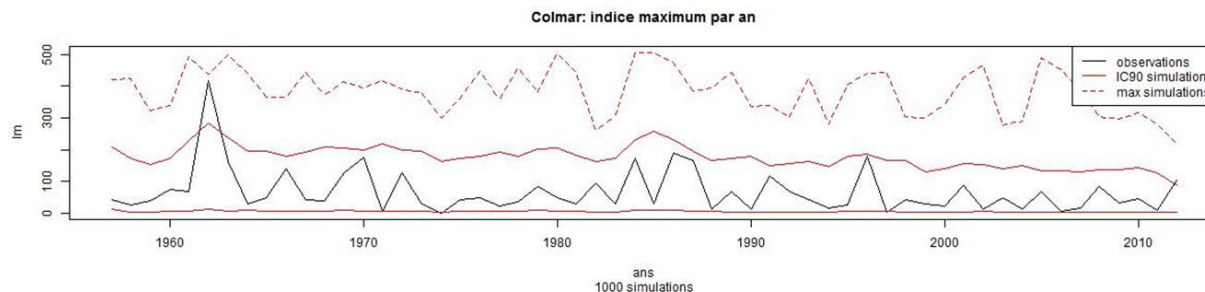
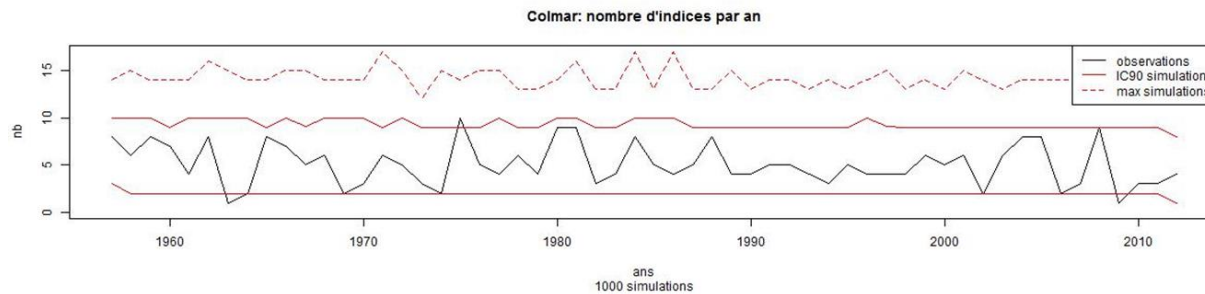
2^{ème} méthode

EXEMPLE D'UTILISATION: INDICES DE GEL

- Indice de gel = cumul des températures journalières $< 0^{\circ}\text{C}$ \Rightarrow peu d'épisodes chaque année



- Simulation d'un grand nombre de séries de température équivalentes à la série observée



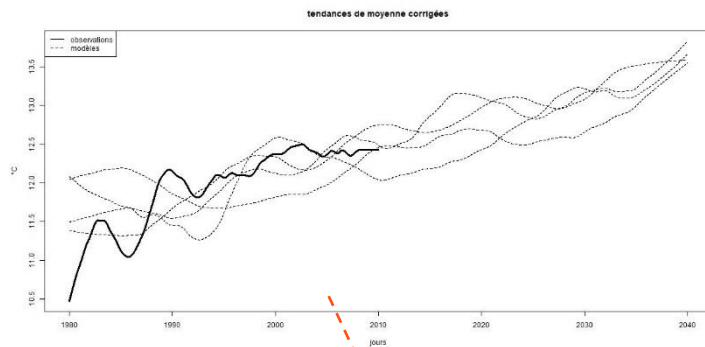
nb < nb5%: 2 nb > nb95%: 1
Im < Im5%: 3 IM > IM95%: 2
5% 56 years : 2,8

CRÉATION DE SÉRIES PASSÉ RÉCENT / FUTUR PROCHE

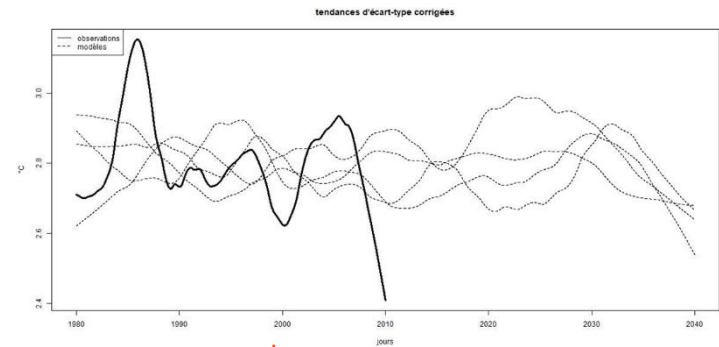
On utilise:

- Les tendances des modèles climatiques corrigées

moyenne

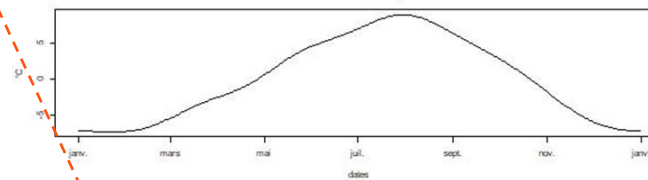


variance

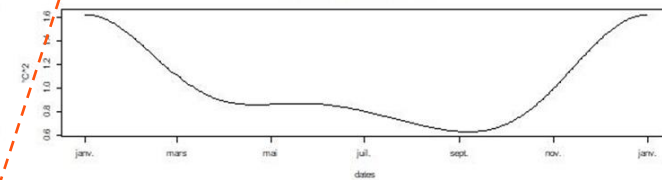


- Les saisonnalités observées

saisonnalité de la moyenne



saisonnalité de la variance



- Des résidus simulés avec le modèle

$$X(t) = m(t) + S(t) + s(t)Sv(t)Z(t)$$

Merci