Modèles d'extrêmes et prédicteurs

Olivier Mestre [1,2]

(1) Météo-France, Ecole Nationale de la Météorologie, Toulouse, France

(2) Université Paul Sabatier, LSP, Toulouse, France

Basé sur des études réalisées en collaboration avec :

Stéphane Hallegatte (CIRED, Météo-France)

Sébastien Denvil (IPSL/LMD)

Fabrice Chauvin (CNRM/GMGEC)

ÉCOLE NATIONALE de la MÉTÉOROLOGIE ENM

METEO FRANCE Toujours un temps d'avance

Extrêmes

- **Une réponse à des problèmes concrets**
  - **dimensionnement d'ouvrages de BTP**
  - **hydrologie, crues, inondations**
  - **assurances**
  - **pollution…**
- **Paramètres : précipitations, vitesse du vent, température, hauteurs de neige…**
- **Un problème délicat à traiter :**
  - **estimation d'événements très rares ou… qui n'ont jamais eu lieu !**
  - **estimation de quantiles d'ordre très élevé, extrapolation**

Risque

- **RISQUE**

X : variable étudiée, de loi
x **: seuil de dimensionnement**
r : Risque=P[X>*x*]

Commanditaire : fixe le risque r

Météo-France : estime *x*

x **= Quantile d'ordre 1-r de la loi de la variable X**

Durée de retour

- **DUREE DE RETOUR**

  – **Usuellement, le risque r est exprimé comme comme une durée moyenne entre deux événements {X>*x*}**

  – **Echantillon de données quotidiennes**

Seuil de durée de retour 20 ans

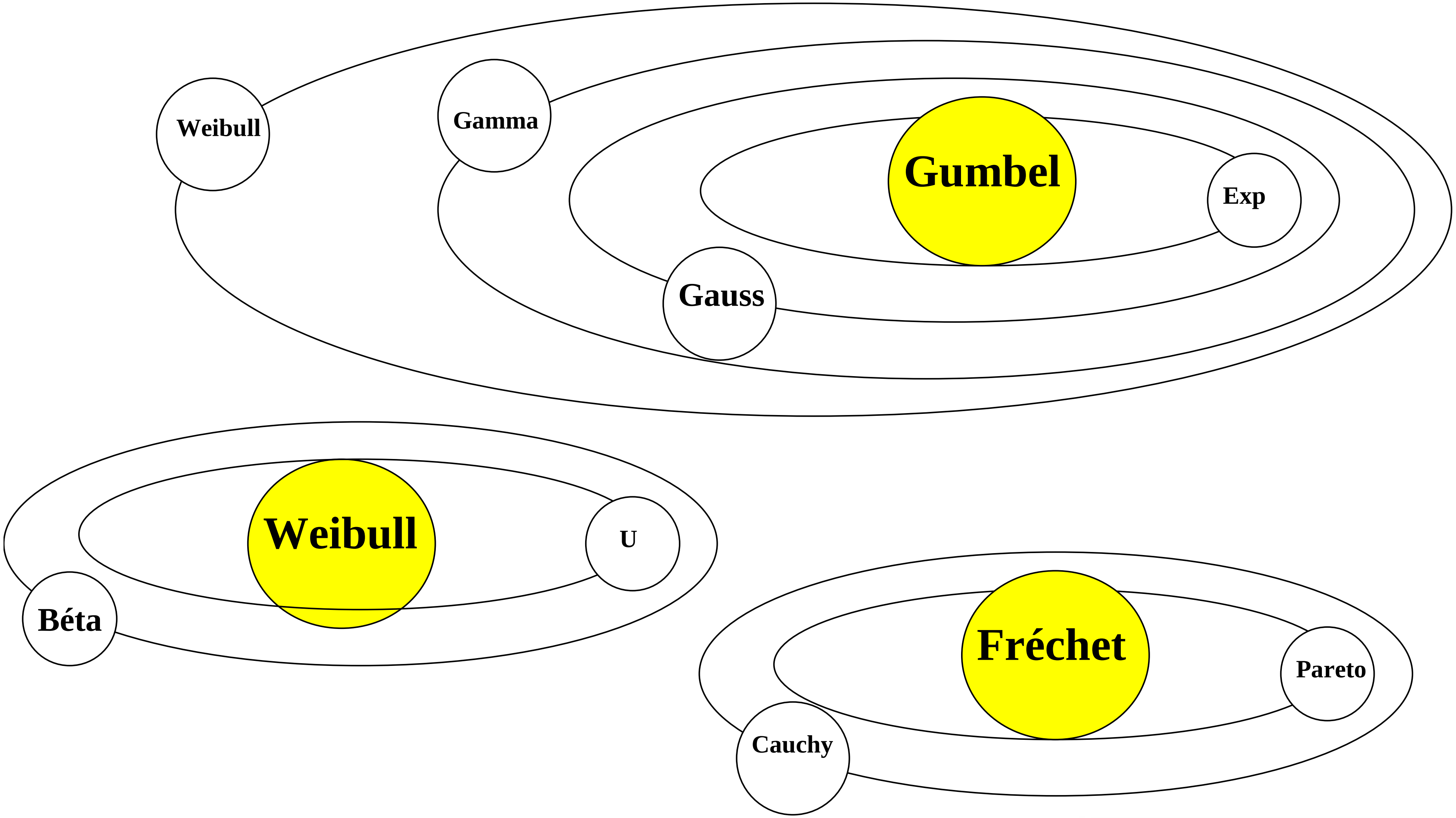
=

Quantile d’ordre 0,99983 de la distribution de X

Théorème des valeurs extrêmes

- **Illustration par l'exemple…**

Théorème des valeurs extrêmes



ATTENTION

- **Il existe des contre-exemples.**

- **En particulier :**

LE THEORÊME DES VALEURS EXTRÊMES NE S'APPLIQUE PAS AUX LOIS DISCRETES

- **Les paramètres géophysiques ne sont pas distribués comme des lois pures**

Vérifier l'adéquation du modèle basé sur les lois d'extrêmes !

Loi GEV

- **PRESENTATION UNIFIEE : GEV(μ,σ,ξ)**

On introduit le paramètre de forme ξ

ξ=0 Gumbel lois non bornées
queues légères (↓ exponentielle)

ξ>0 Fréchet
lois non bornées, queues lourdes (↓ puissance)

ξ<0 Weibull
lois bornées

La GEV par l'exemple

- **Petite manip!**

Fonction de répartition

- **Fonction de répartition G**

$$\mathbf{P}\left[\mathbf{M}_{\mathrm{N}} < x\right] = \mathbf{G}(x) = \mathbf{exp}\left[-\left(1+\xi\cdot\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right]$$

- **Quantile associé à la durée de retour T**

$$x_{\mathrm{T}} = \mu - \frac{\sigma}{\xi}\left[1-\left(-\ln\left(1-\frac{1}{\mathrm{T}}\right)\right)^{-\xi}\right]$$

- **Avantages**
  - **Max annuels: seuil de durée de retour 20 ans = quantile 0,95=1-1/20**
  - **La loi est théoriquement connue**

Estimons un seuil de durée de retour

- **Librairie « extRemes » sous R**

- **Précipitations au pas de temps quotidien à Toulouse**

Alternatives

- METHODES A SEUIL

  Loi des dépassement d’un seuil (seuil élevé)

- K plus grandes valeurs

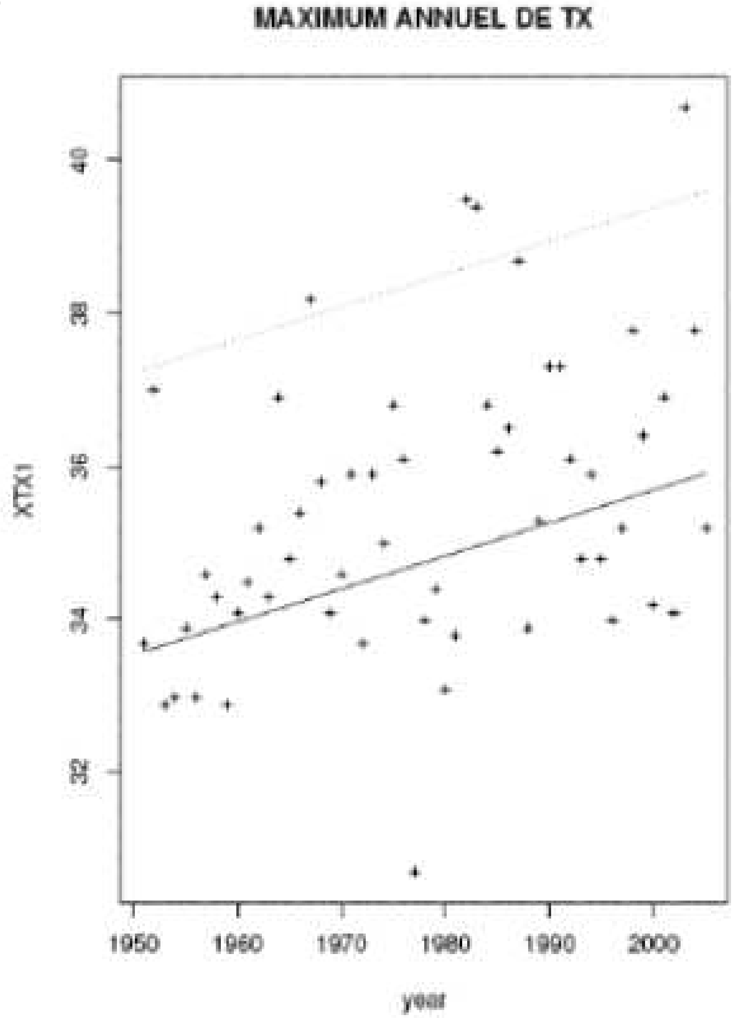
- Processus de points

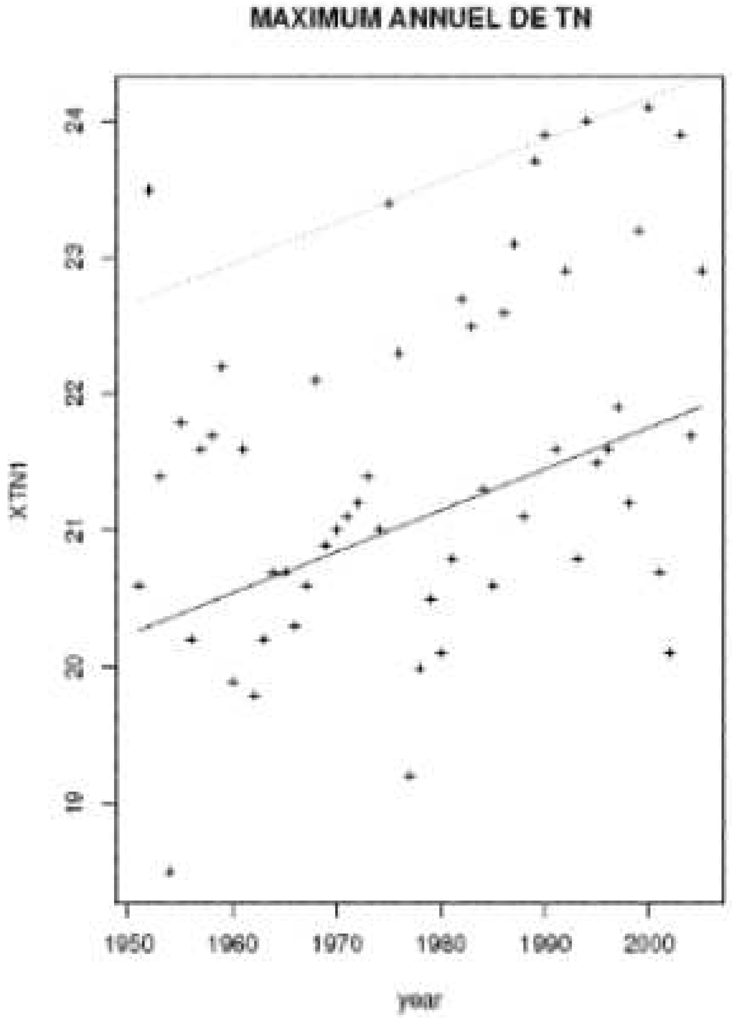
Extrêmes et covariables

- **BUT**
  - **Etudier l'influence de covariables sur les extrêmes**
  - **Etudier l'évolution des extrêmes (covariable = temps)**
- **Principe**
  - **Modéliser les paramètres en fonction des covariables**
  - **Giacomo Parsimoni (1725-1755)**

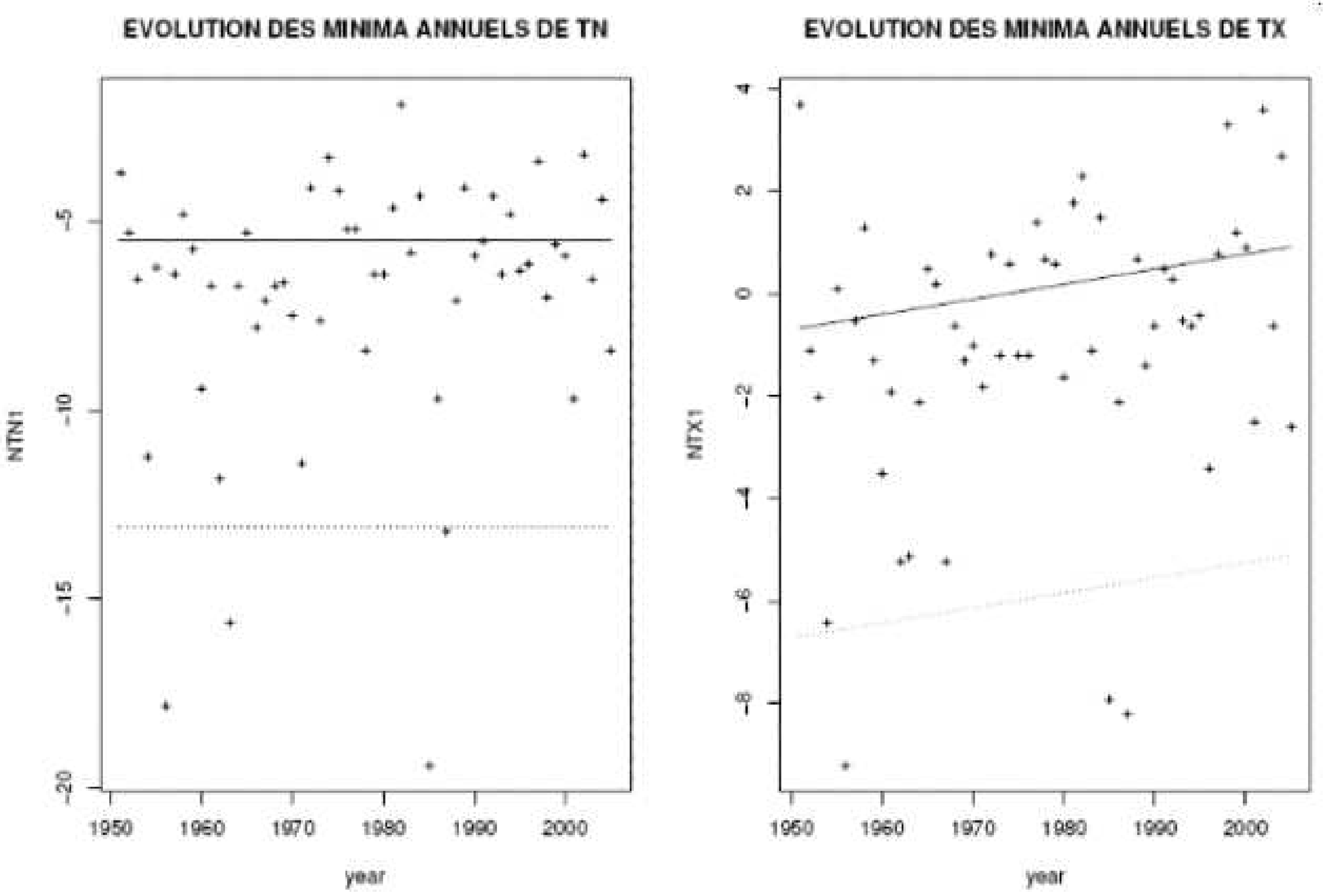
« De deux modèles d'égale beauté, choisir le moins compliqué »

Maxima annuels de températures à Toulouse





Exemple : Minima annuels de températures à Toulouse



Problème…

- **Variations conjointes de paramètres non indépendants**

modèles paramétriques difficiles à formuler *a priori*

Modèles additifs

- **Philosophie**

Approche basée sur les données
Plutôt que basée sur un modèle

Technique exploratoire

- **Inférence : approximations uniquement**

Modèles additifs

- **Modèles additifs:**

  Une manière efficace de faire de la régression non-linéaire

- **ATTENTION!**

ADAPTE A UN FAIBLE NOMBRE DE PREDICTEURS

2, 3 au maximum

Modèles linéaires et additifs

- Modèle linéaire gaussien : $IE[Y]=\beta_o+\beta_1X_1+\beta_2X_2$
- Modèle additif gaussien : $IE[Y]=S_1(X_1)+S_2(X_2)$

S_1, S_2 fonctions « lisses » des prédicteurs X_1, X_2, (LOESS, SPLINE)

Estimation de S_1, S_2 : « Backfitting »

- PRINCIPE DU BACKFITTING

$Y=S_1(X_1)+e \rightarrow$ estimation S_1^*

$Y-S_1^*(X_1)=S_2(X_2)+e \rightarrow$ estimation S_2^*

$Y-S_2^*(X_2)=S_1(X_1)+e \rightarrow$ estimation S_1^{**}

$Y-S_1^{**}(X_1)=S_2(X_2)+e \rightarrow$ estimation S_2^{**}

$Y-S_2^{**}(X_2)=S_1(X_1)+e \rightarrow$ estimation S_1^{***}

Etc… jusqu'à convergence

Modèles Additifs Généralisés (GAM)

- **Extension à des variables non gaussiennes**

- **Modèles additifs généralises (GAM) : modèles additifs du paramètre naturel de lois de la famille exponentielle (Poisson, Binomial, Gamma, Gauss…).**

$$g[\mu]=\theta=S_1(X_1)+S_2(X_2)$$

- **Modèles « Vecteurs Additifs Généralisés » (VGAM): généralisation du concept de régression…**

Modélisation VGAM de la GEV

- **Comment?**

  Modélisation VGAM : Yee & Wild, 1996

  Implémentation : package VGAM sous R (Yee, 2006)

  Algorithme *vector backfitting* et *vector spline*.

- **Attention**
  - **Peu de prédicteurs**
  - **Intervalles de confiance ponctuels : pas de matrice de covariance des estimateurs**
  - **Convergence parfois difficile à obtenir. Utiliser des fonctions de lien: log(σ)**

Evolution des maxima de températures dans un GCM

Avec Sébastien Denvil, LMD

Données

- Maxima annuels de température

- Période 1860-2100

- Modèle IPSL GCM (5ème IPCC)

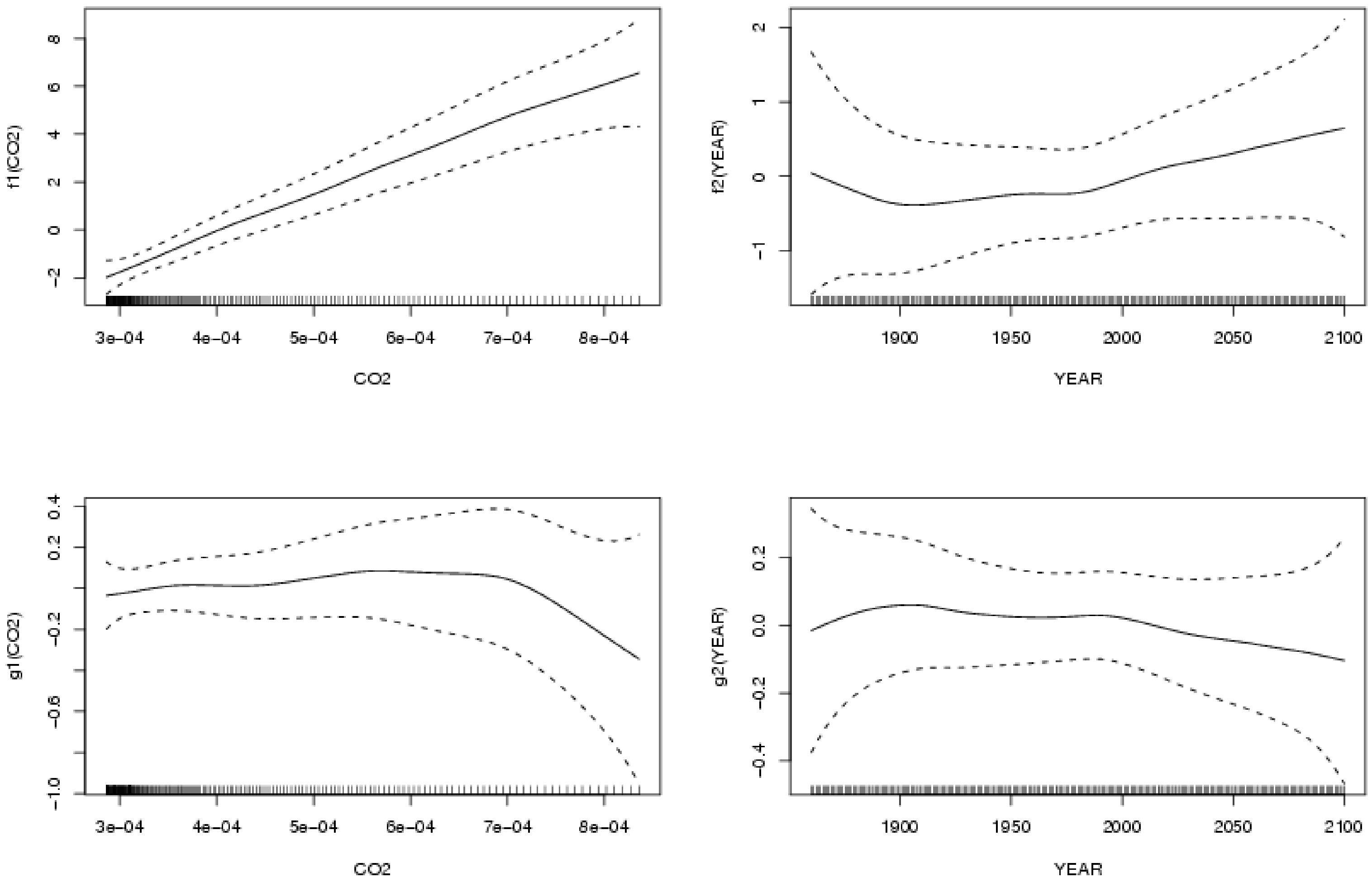
- Concentration des gaz à effet de serre et des aérosols

  avant 2000 : observations

  2000-2100 : SRES-A2 IPCC

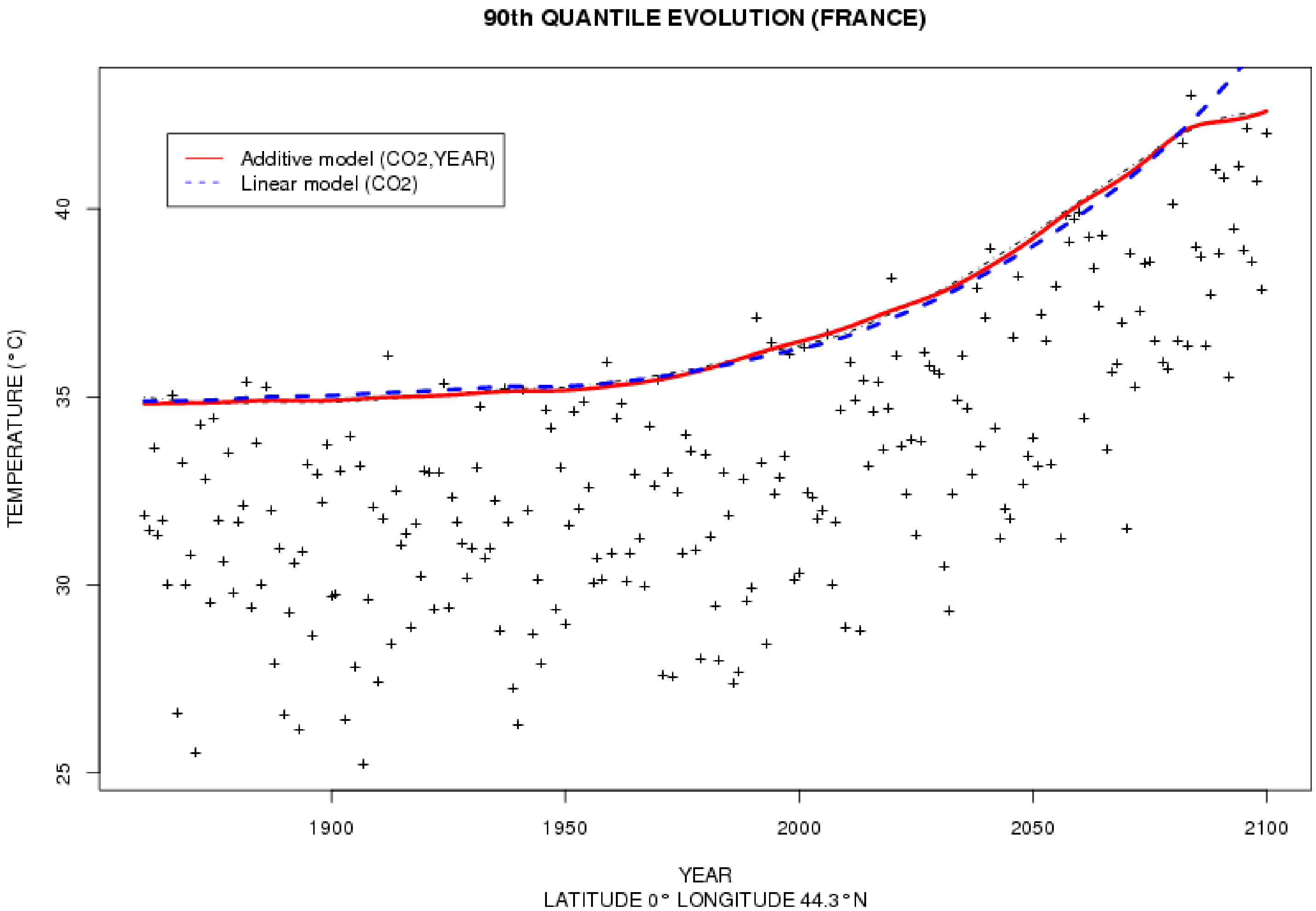
Exemple : point de grille sur la France

- **Rôle majeur joué par le CO_2. Légère modulation au cours du temps.**
- $\mu=f1(CO2)+f2(YEAR)$ $\sigma=g1(CO2)+g2(YEAR)$ $\xi=constant$

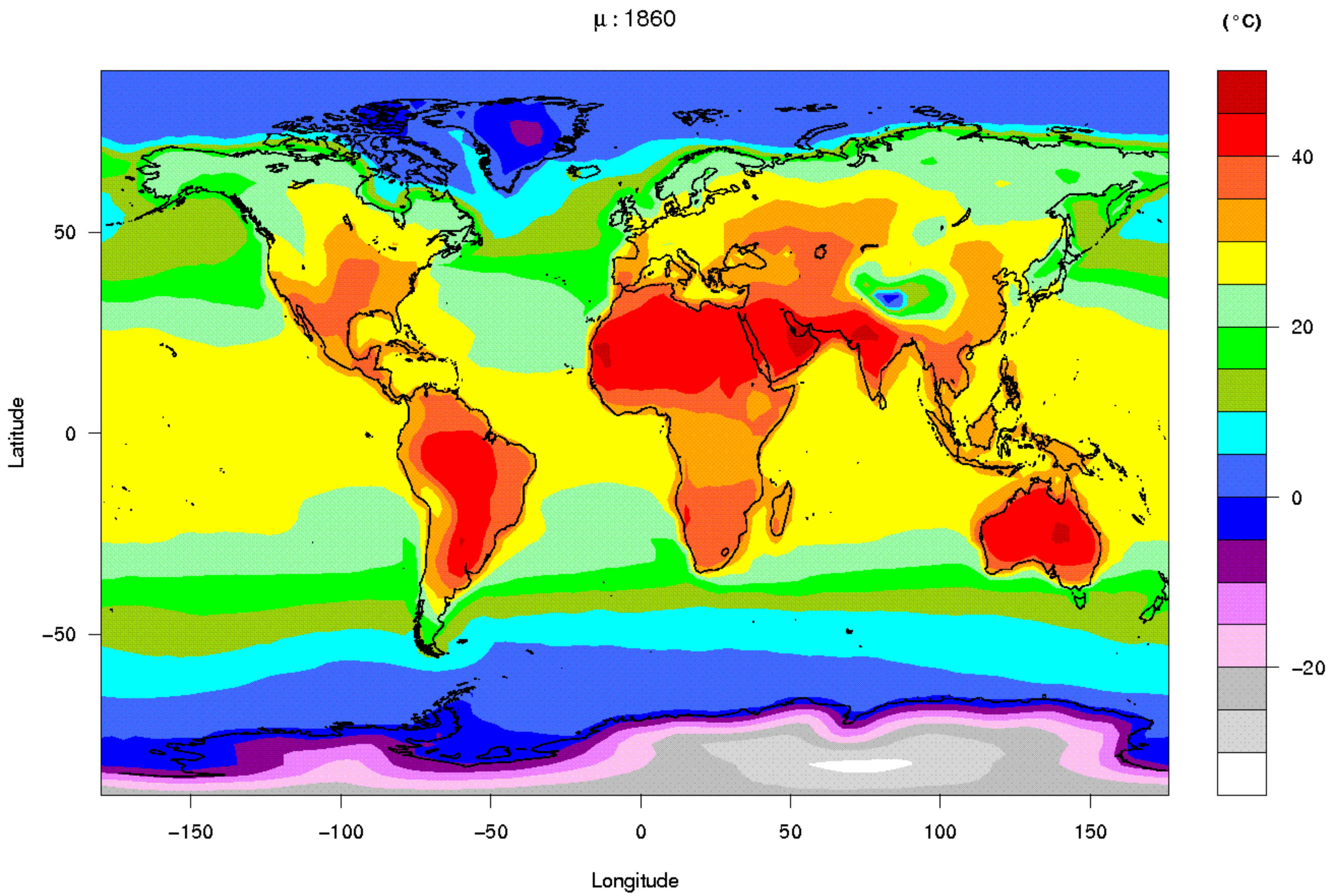


Exemple : point de grille sur la France

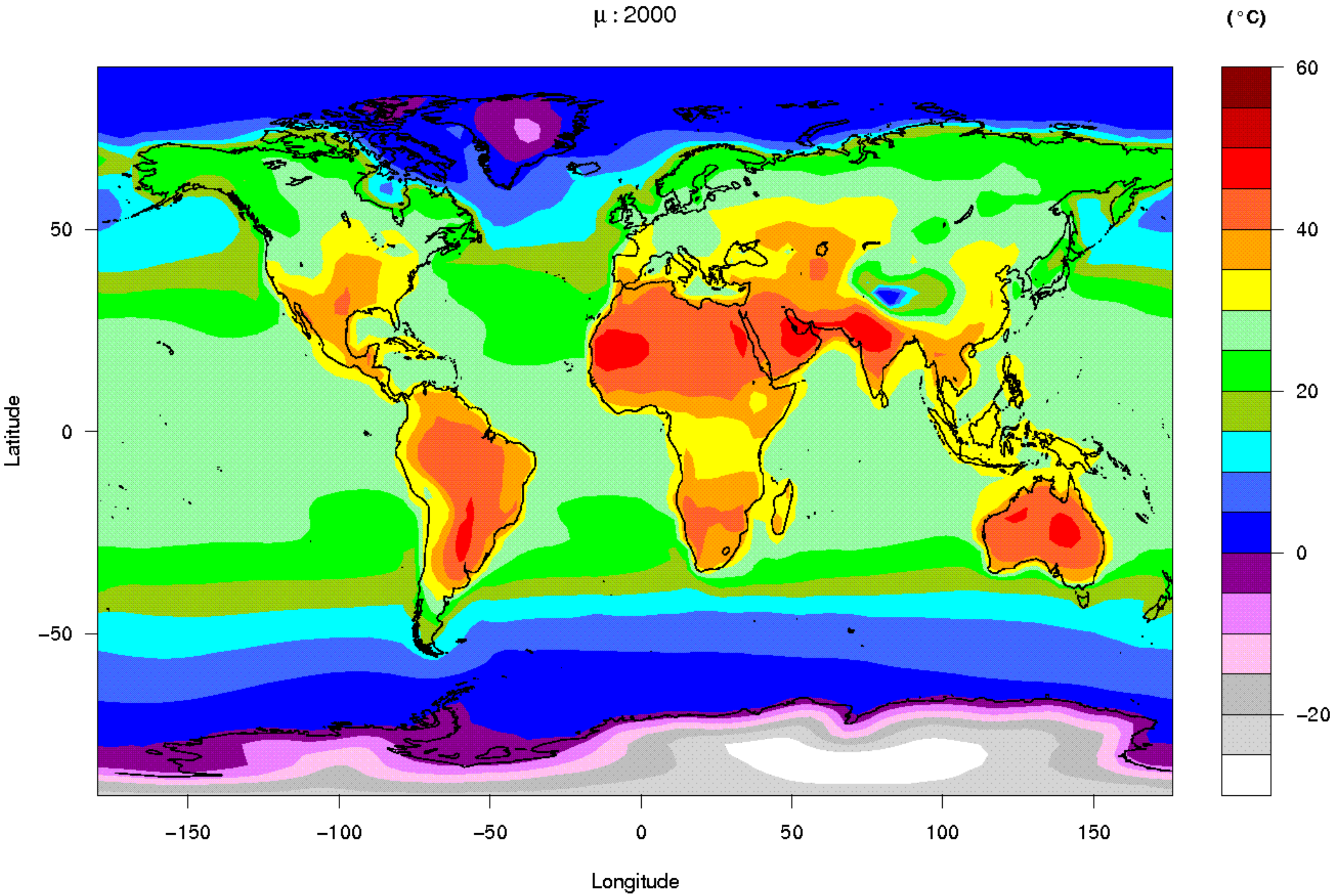
- **Evolution du quantile 0,90 sur la France**



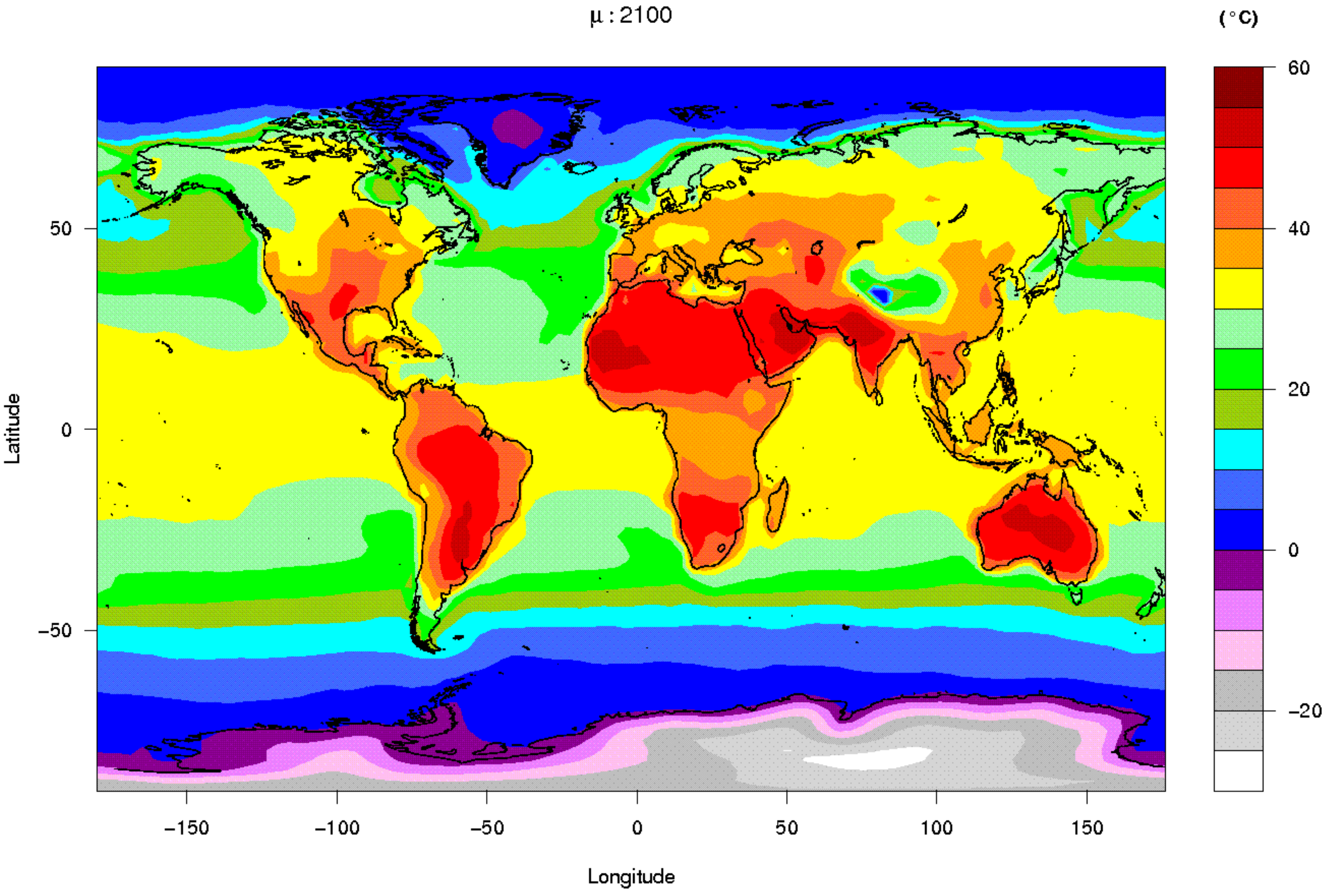
Paramètres de la GEV



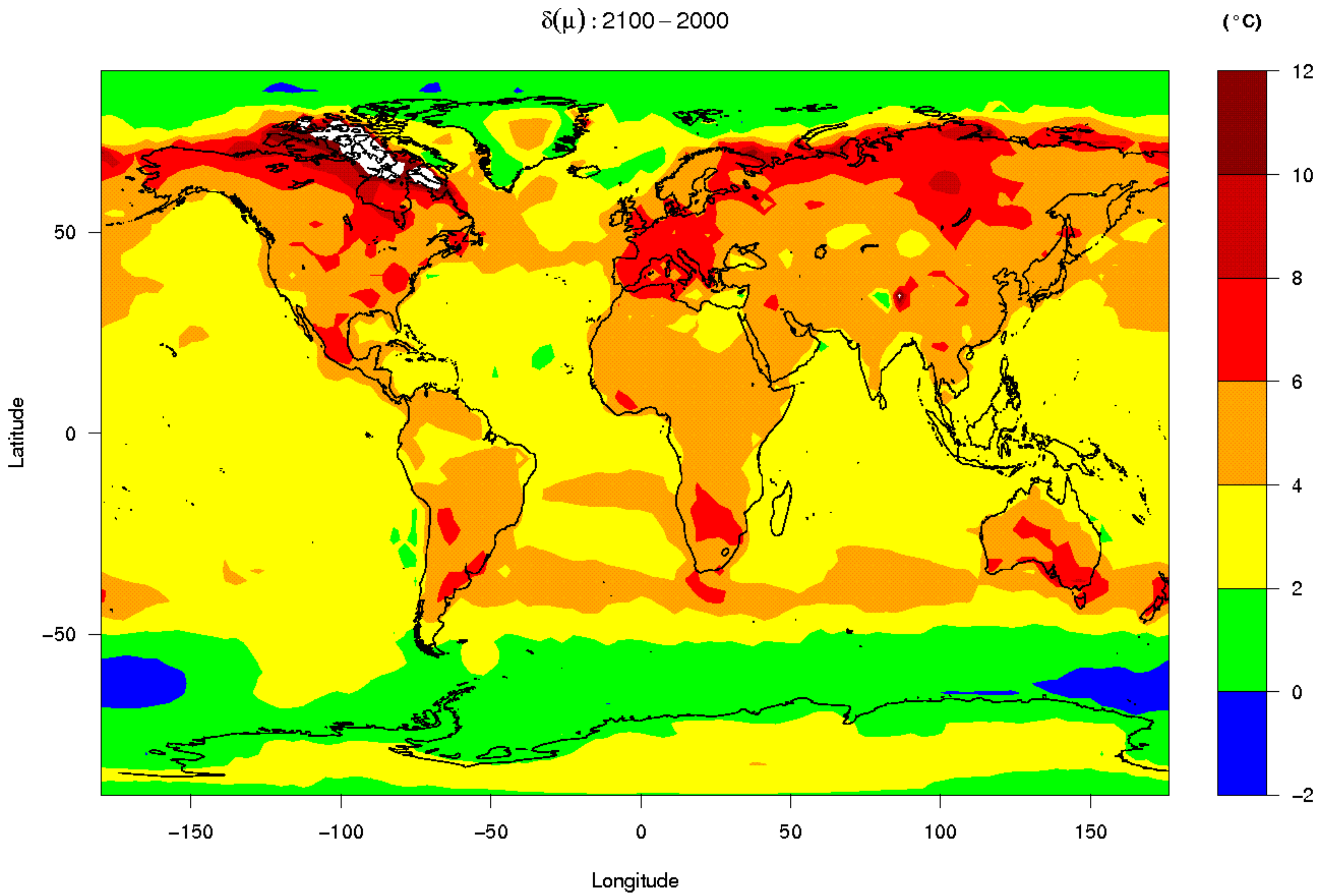
Paramètres de la GEV



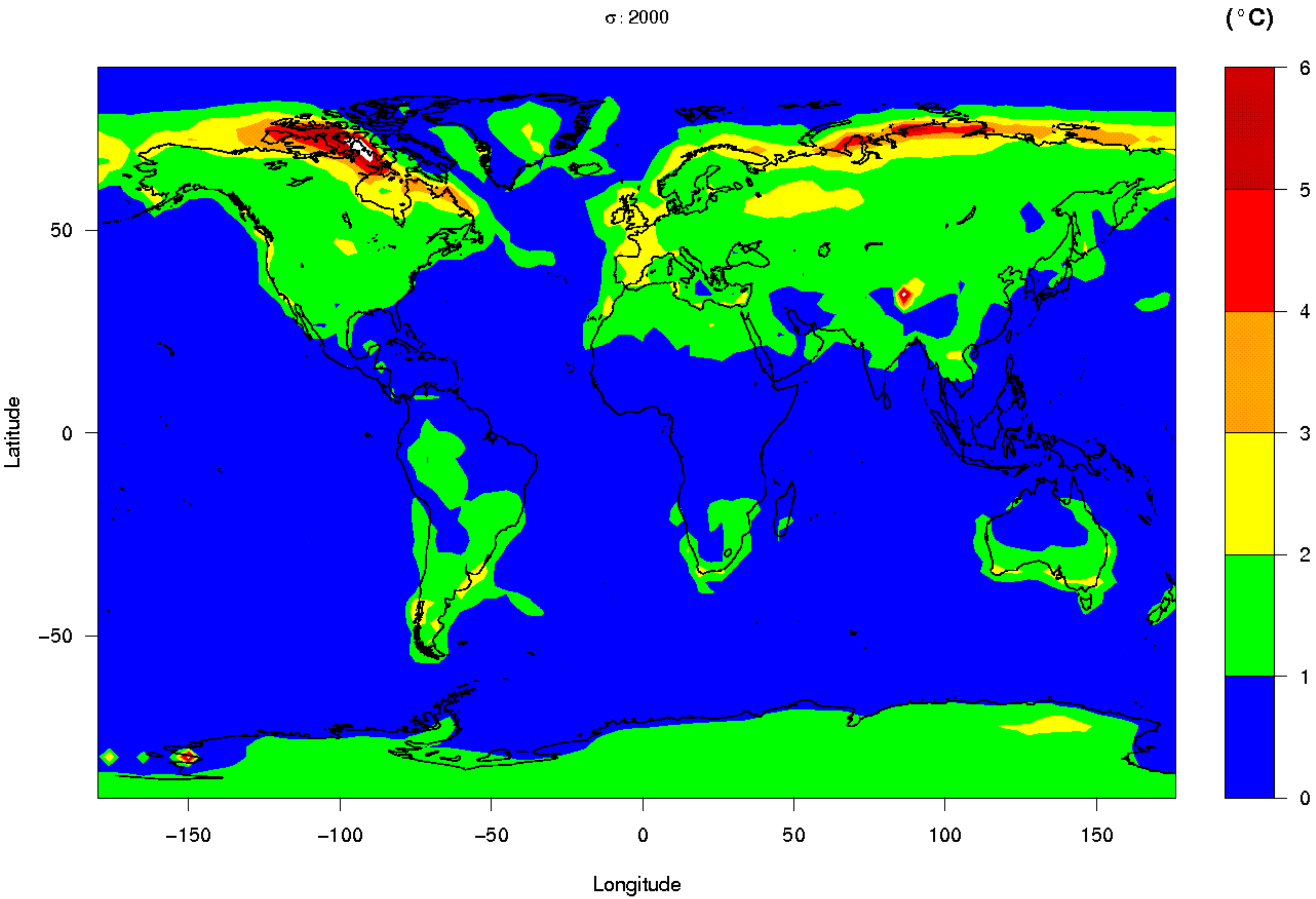
Paramètres de la GEV



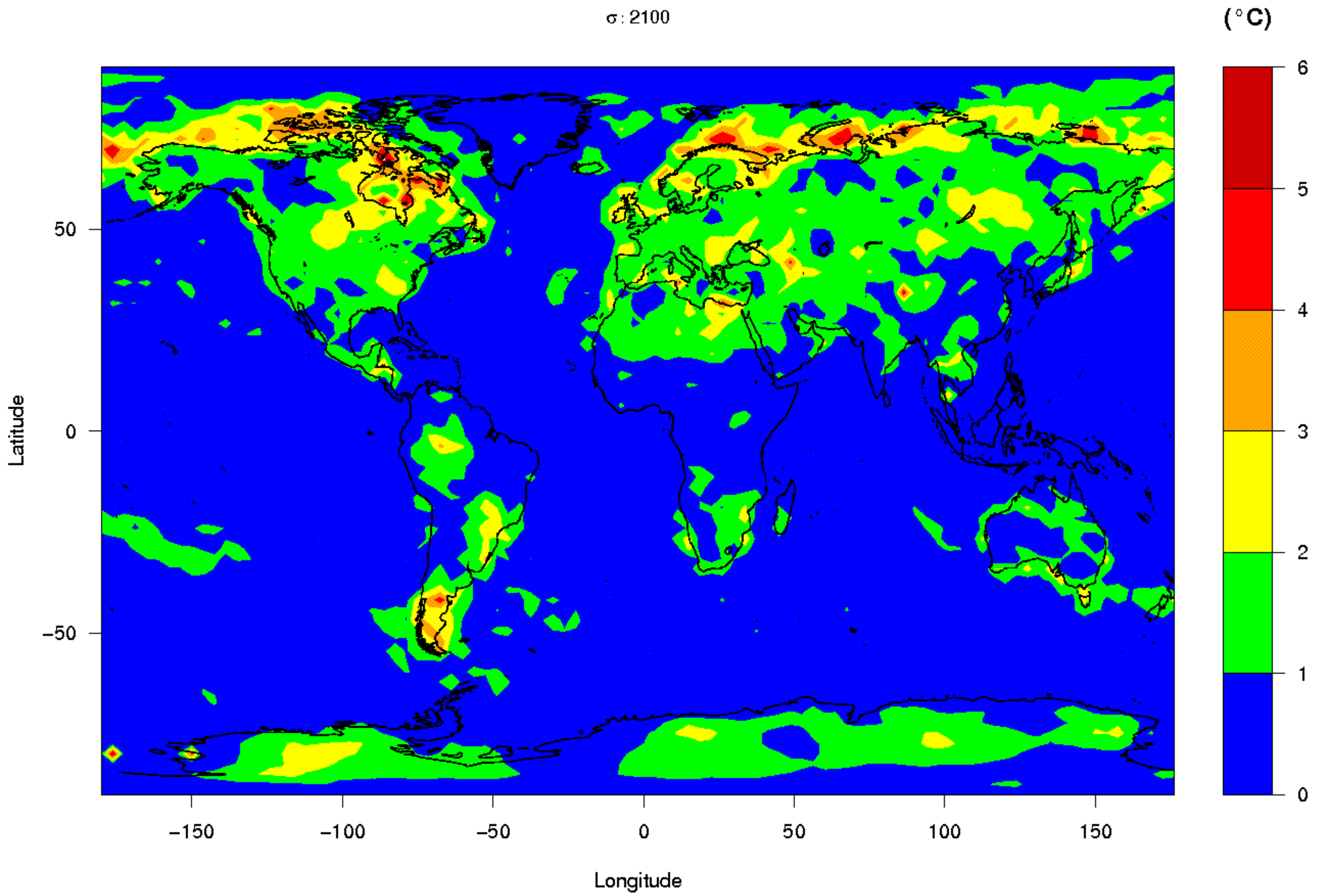
µ:2100-2000 différence



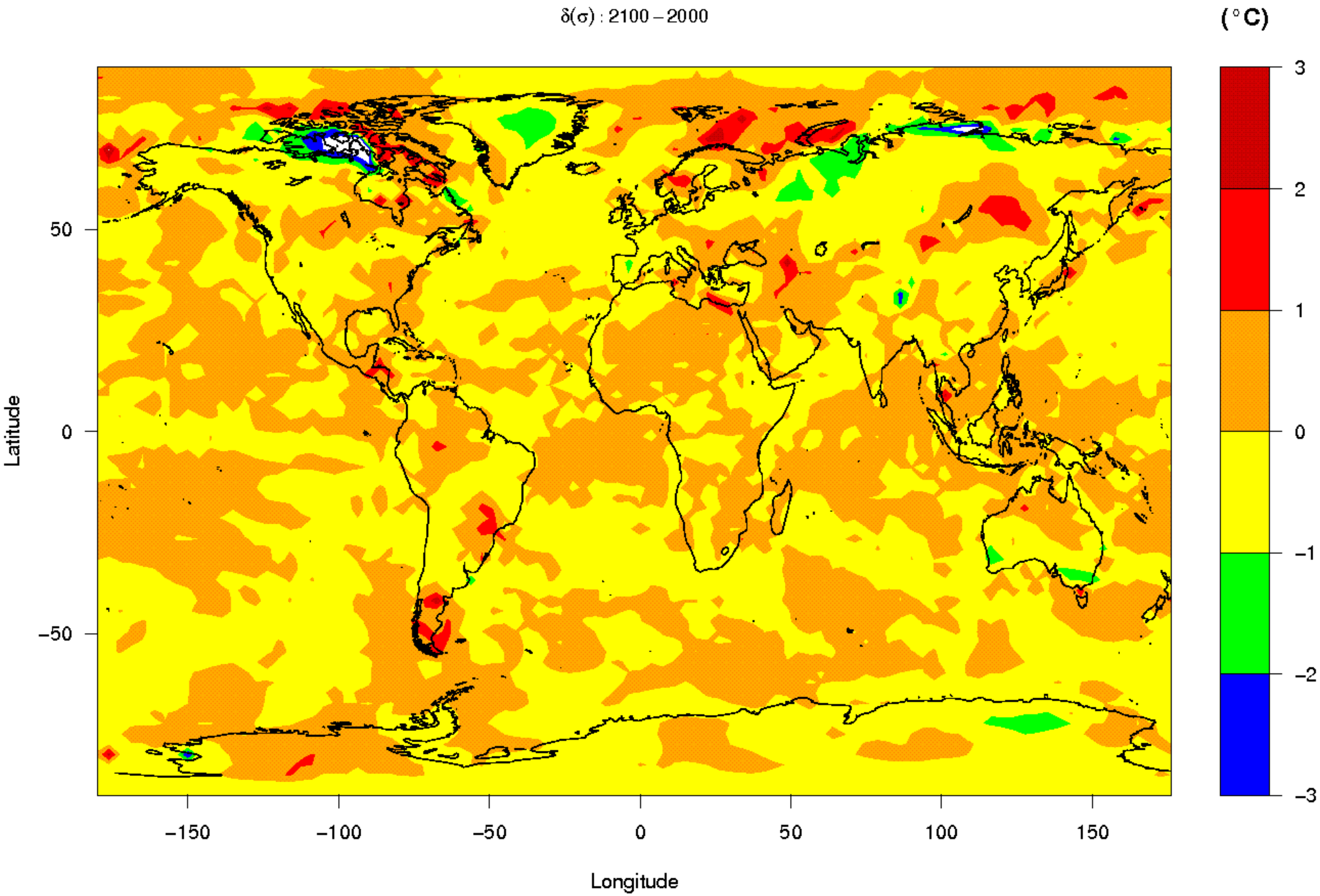
Paramètres de la GEV



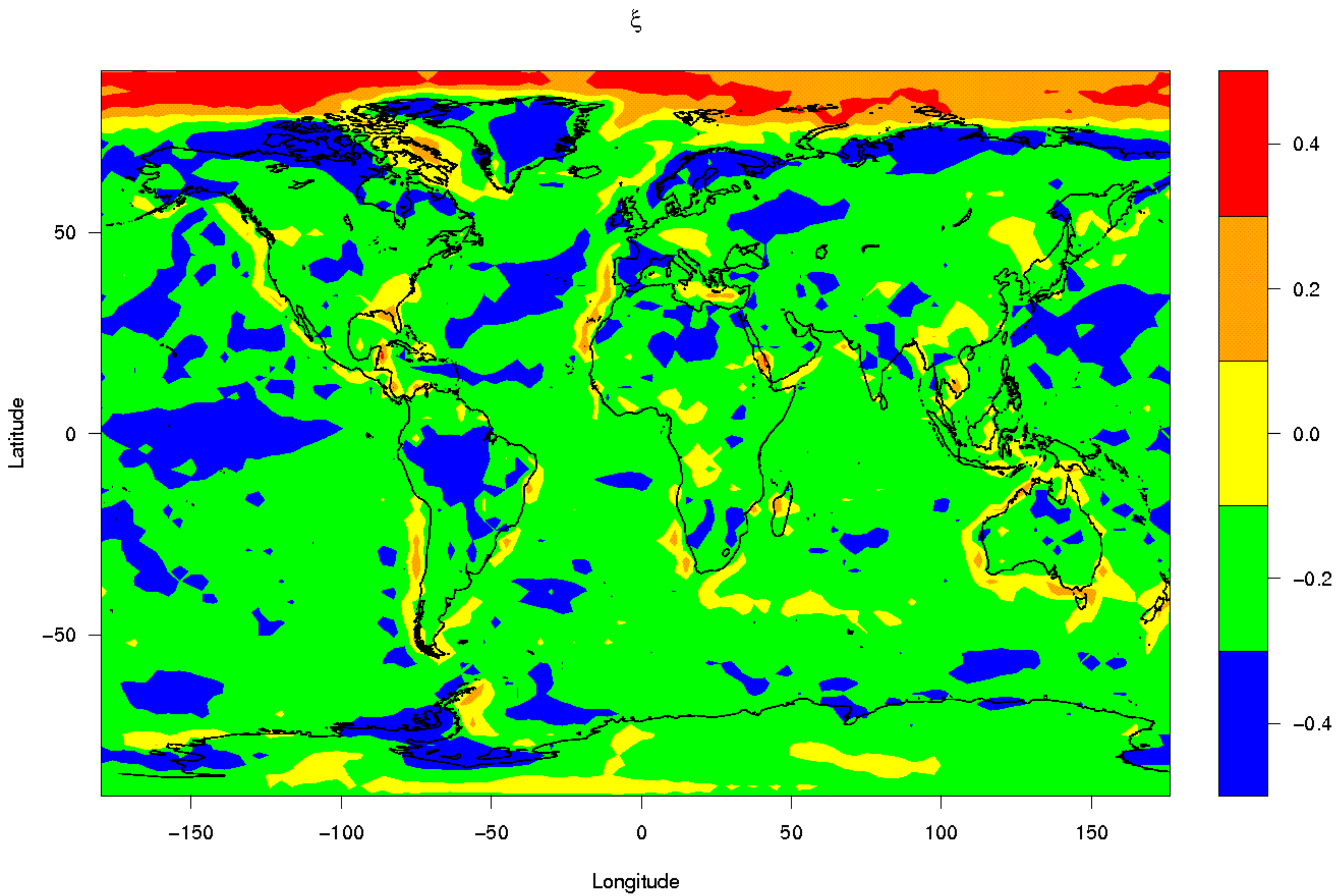
Paramètres de la GEV



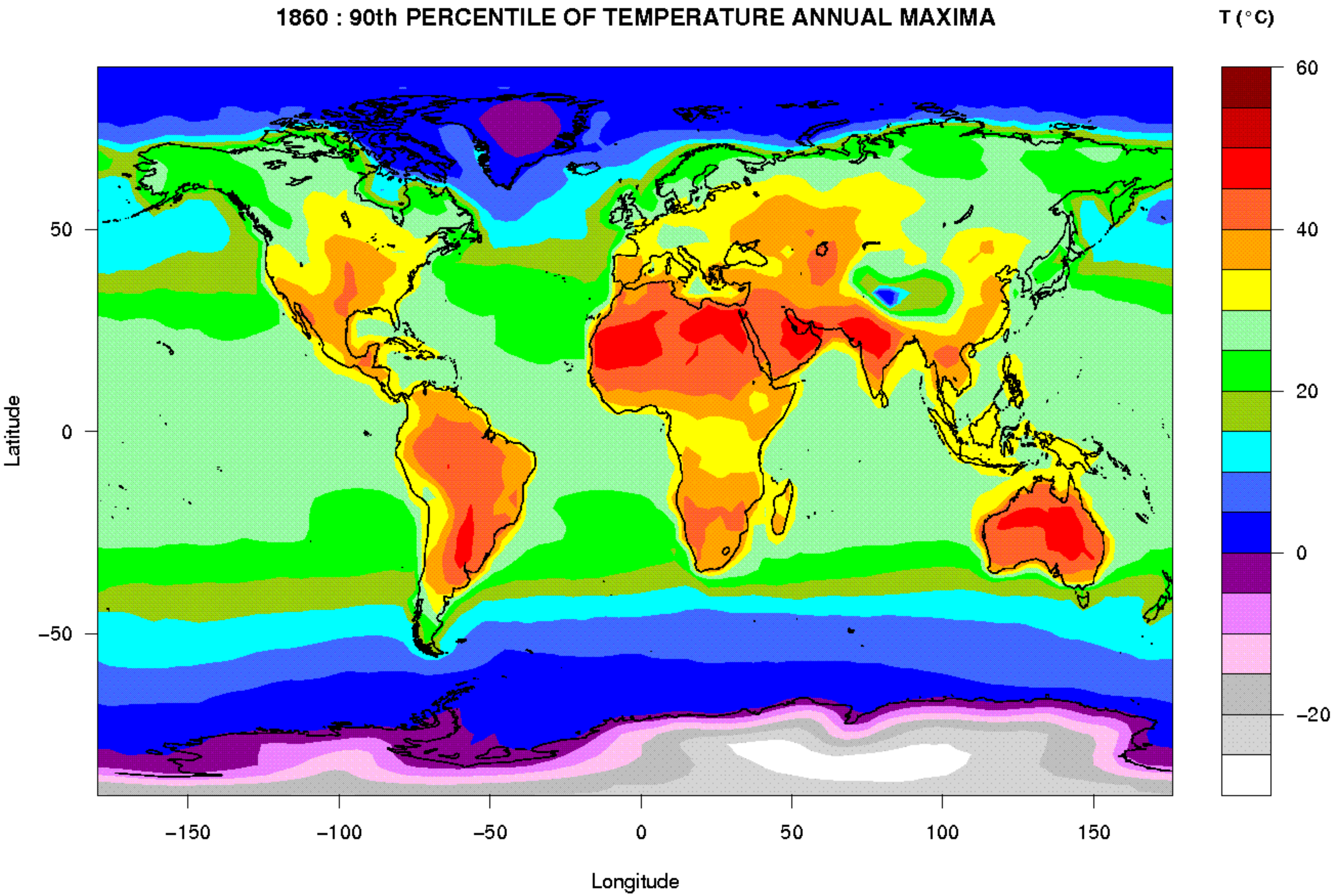
σ:2100-2000 différence



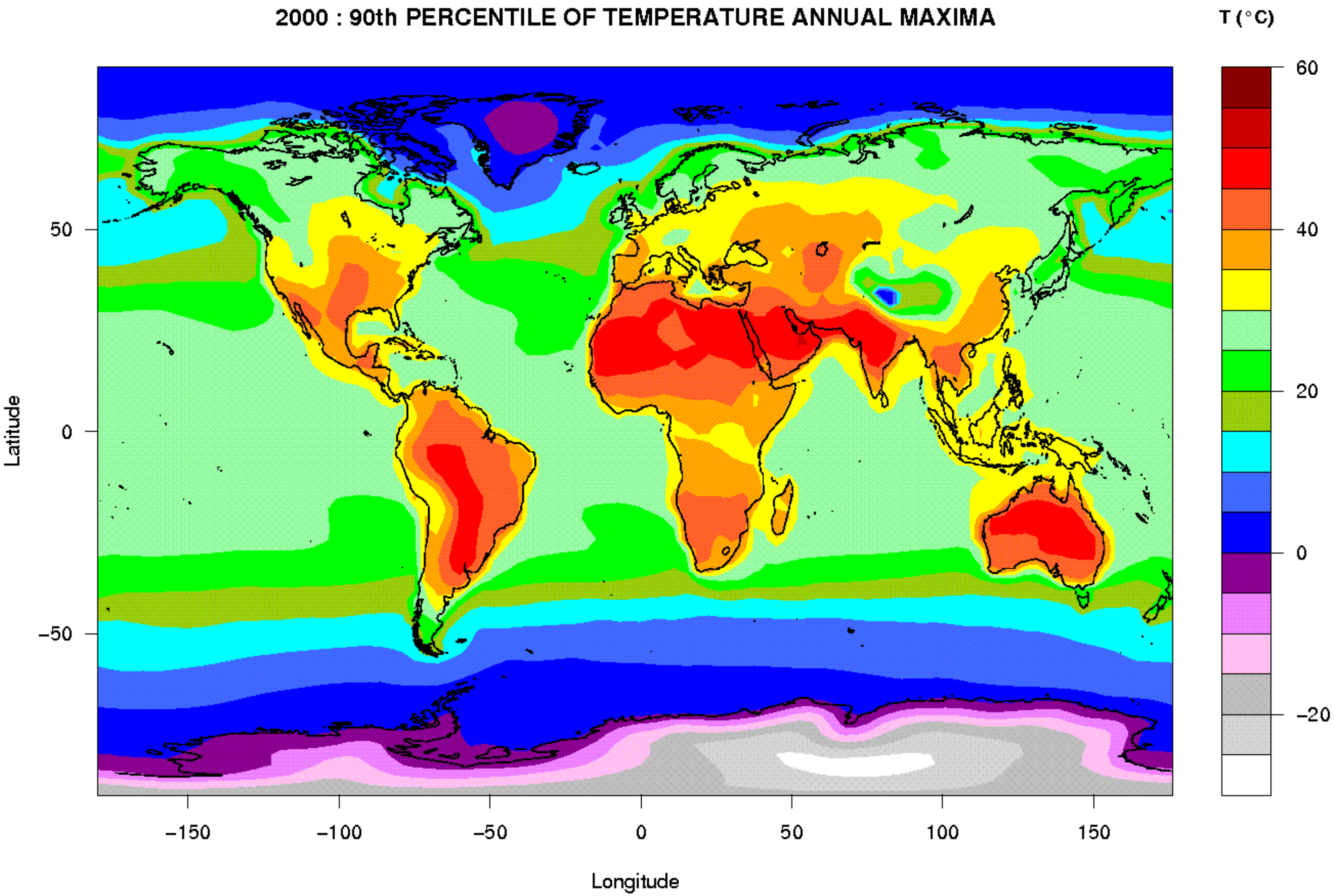
Paramètres de la GEV



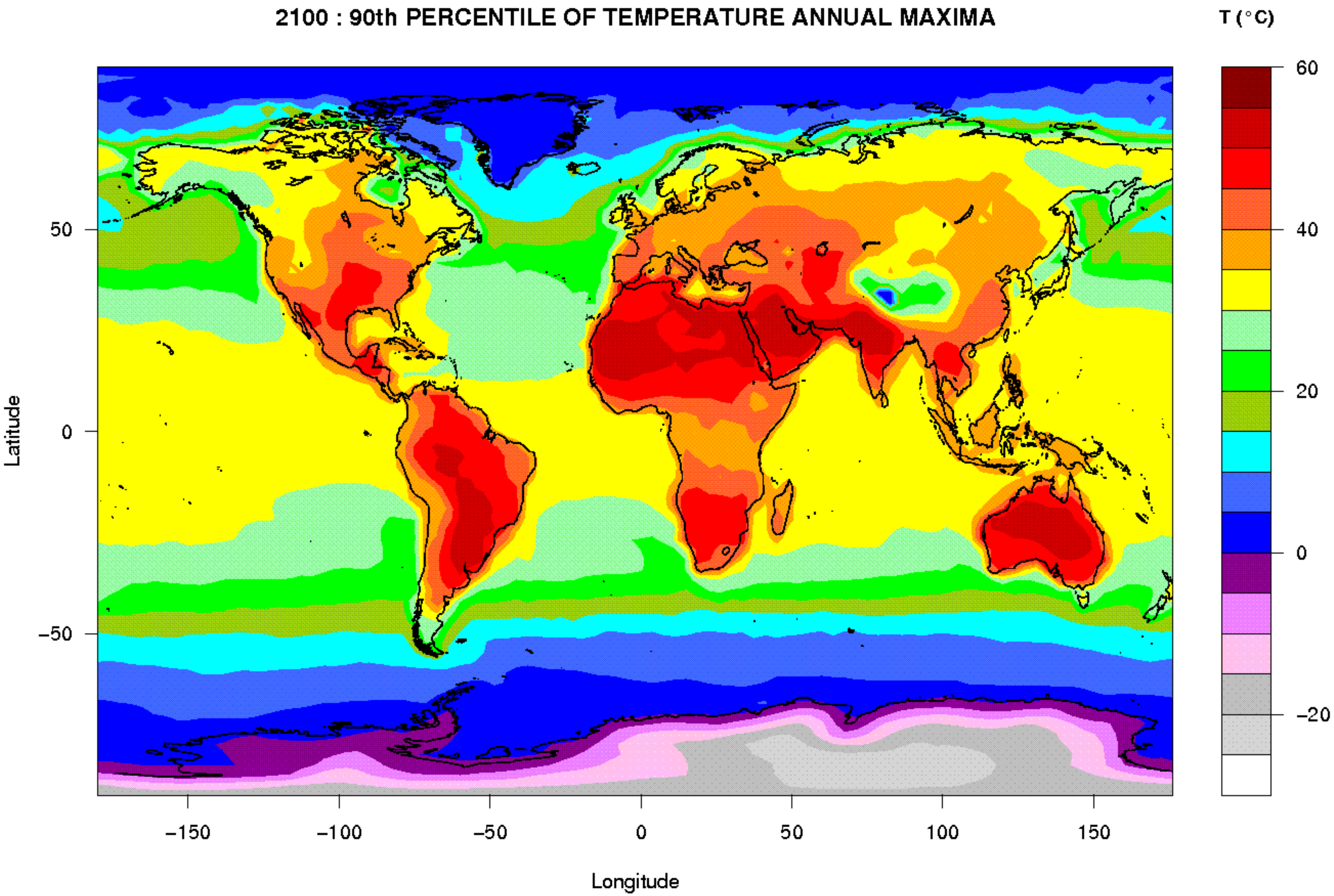
Quantile 0,90 du TX annuel



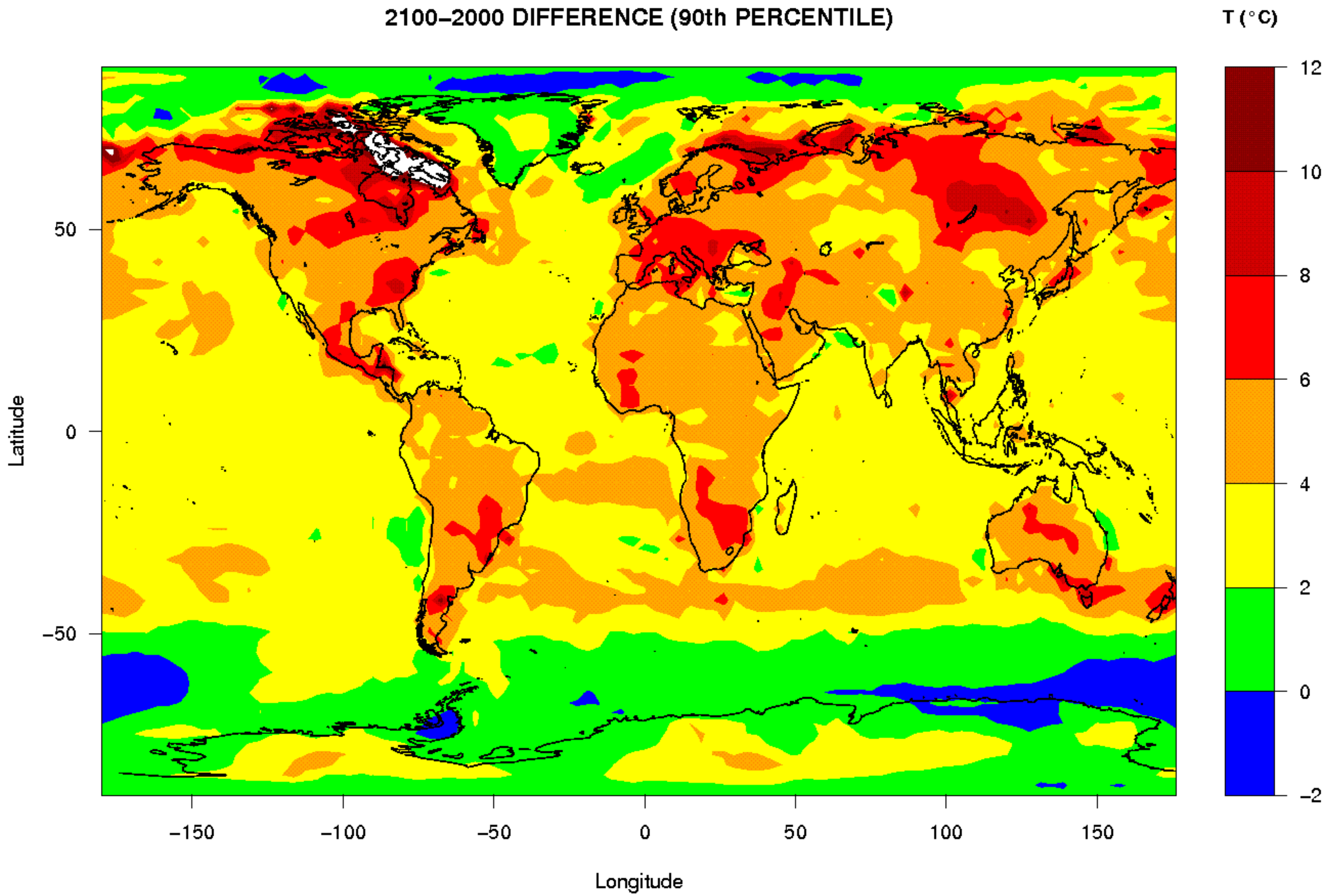
Quantile 0,90 du TX annuel



Quantile 0,90 du TX annuel



Quantile 0,90 du TX annuel : Différence 2100-2000



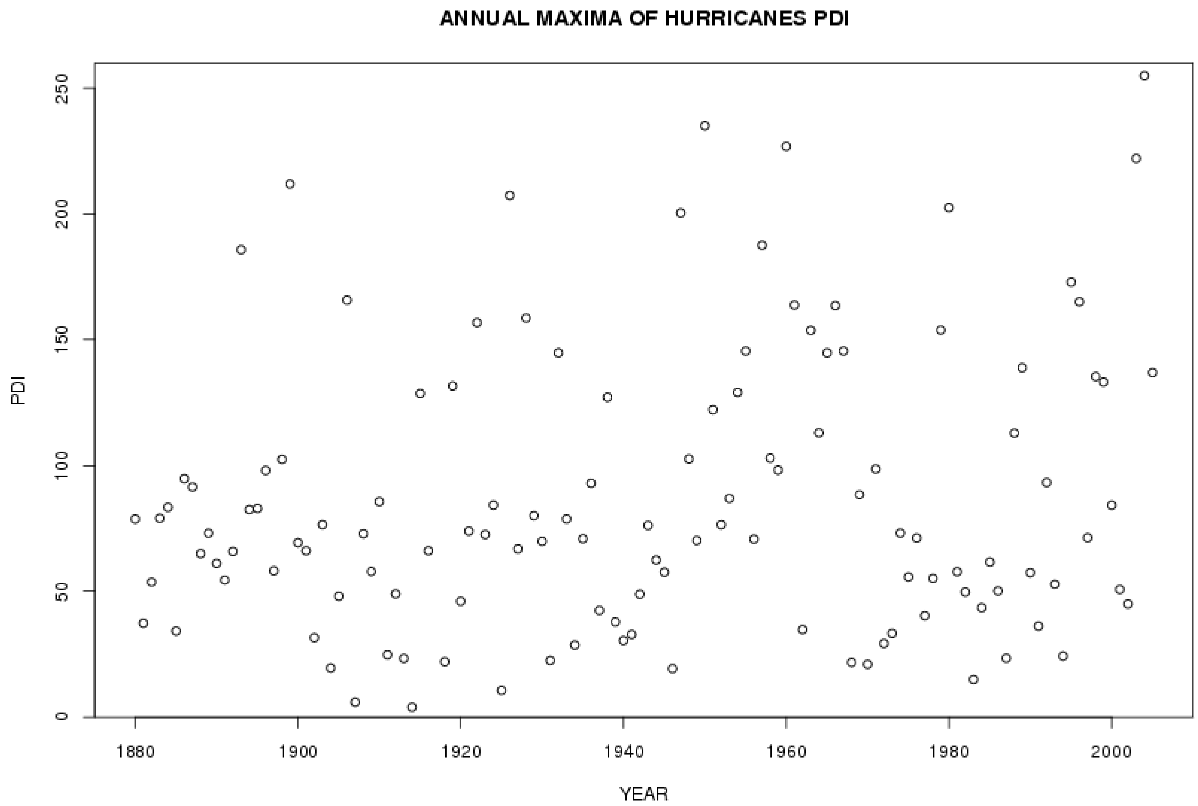
Exemple

Energie intégrée maximale (PDI) des cyclones sur l'Atlantique Nord

Avec Stéphane Hallegatte, CIRED, ENM

PDI maximale annuelle

- **Période : 1880-2005**



- **Prédicteurs potentiels : NAO, SOI, AMO, SST, Température Globale**

Modélisation VGAM de la GEV

- **Pourquoi?**

Variations conjointes de paramètres non indépendants: les modèles paramétriques sont difficiles à formuler

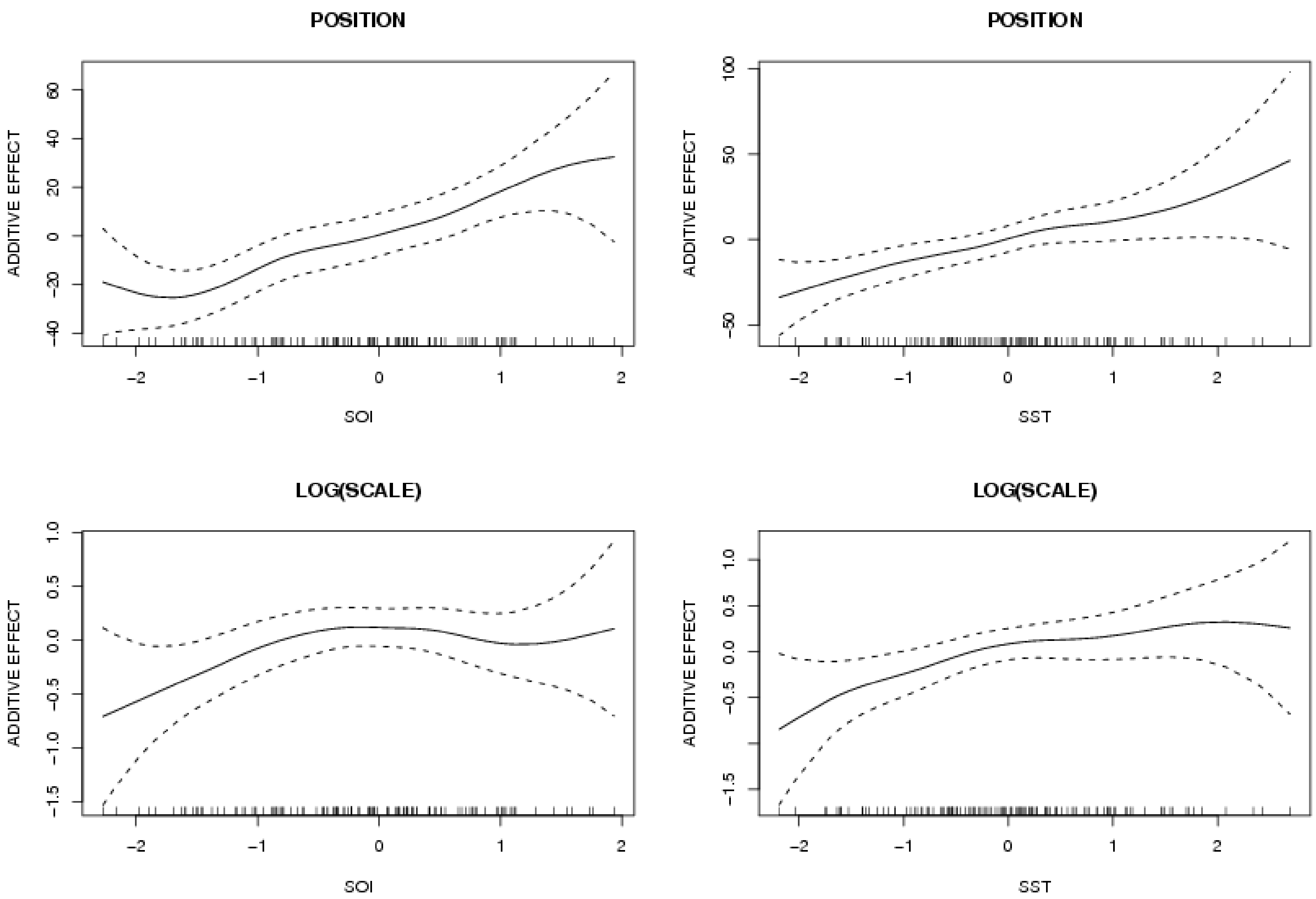
μ, σ modélisés comme des fonctions souples de covariables
ξ reste constant

$$\mu=\mu_o+f_1(X_1)+f_2(X_2)$$

$$\sigma=\sigma_o+g_3(X_3)+g_4(X_4)$$

$$\xi=\xi_o$$

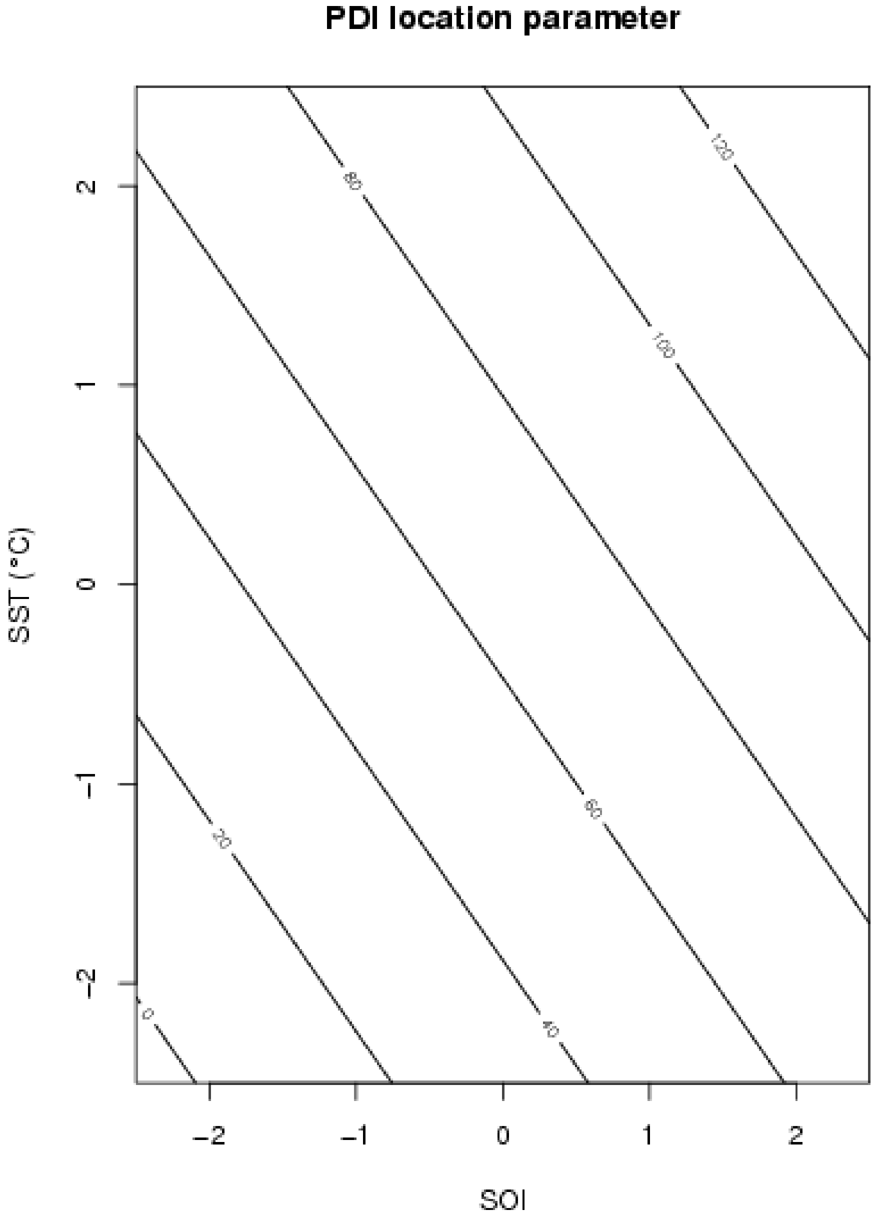
Estimation des effets additifs

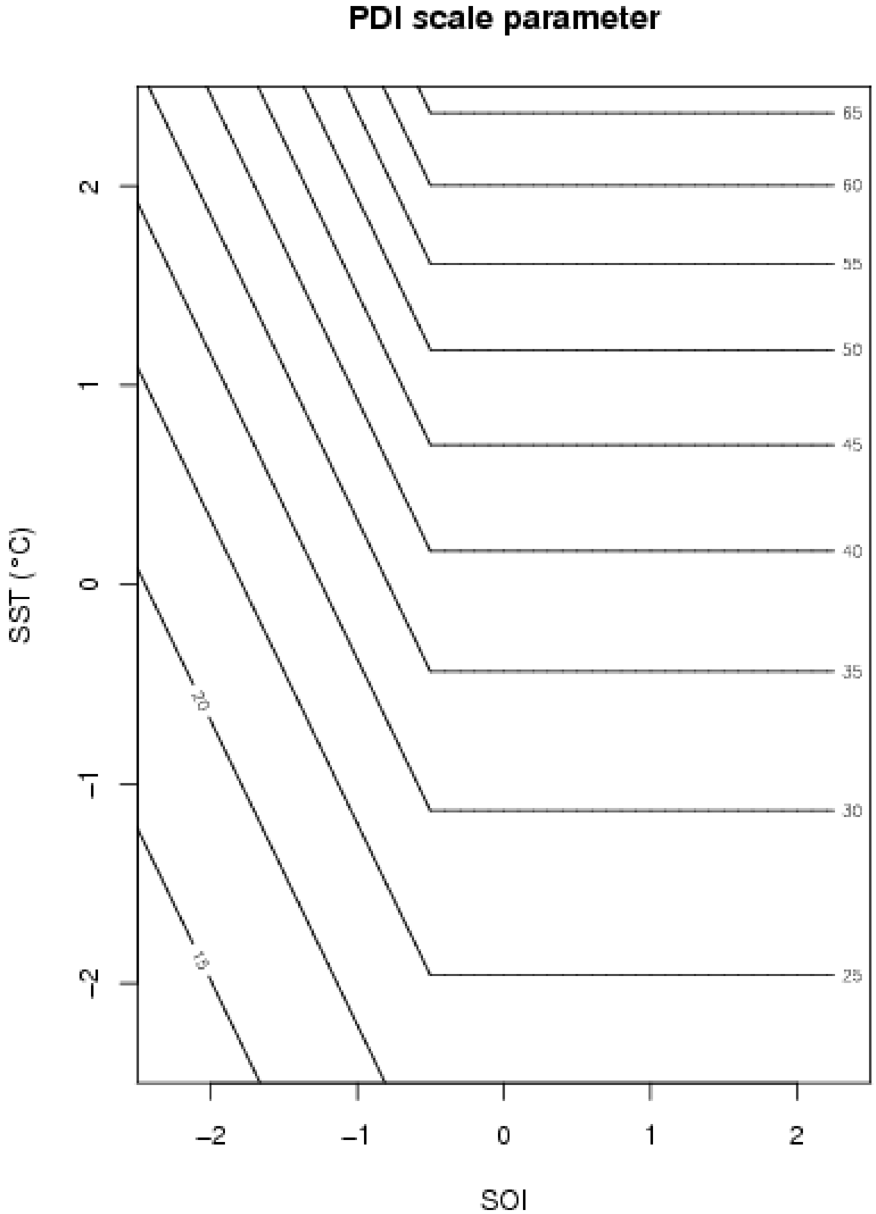


Modèle VGLM de la PDI

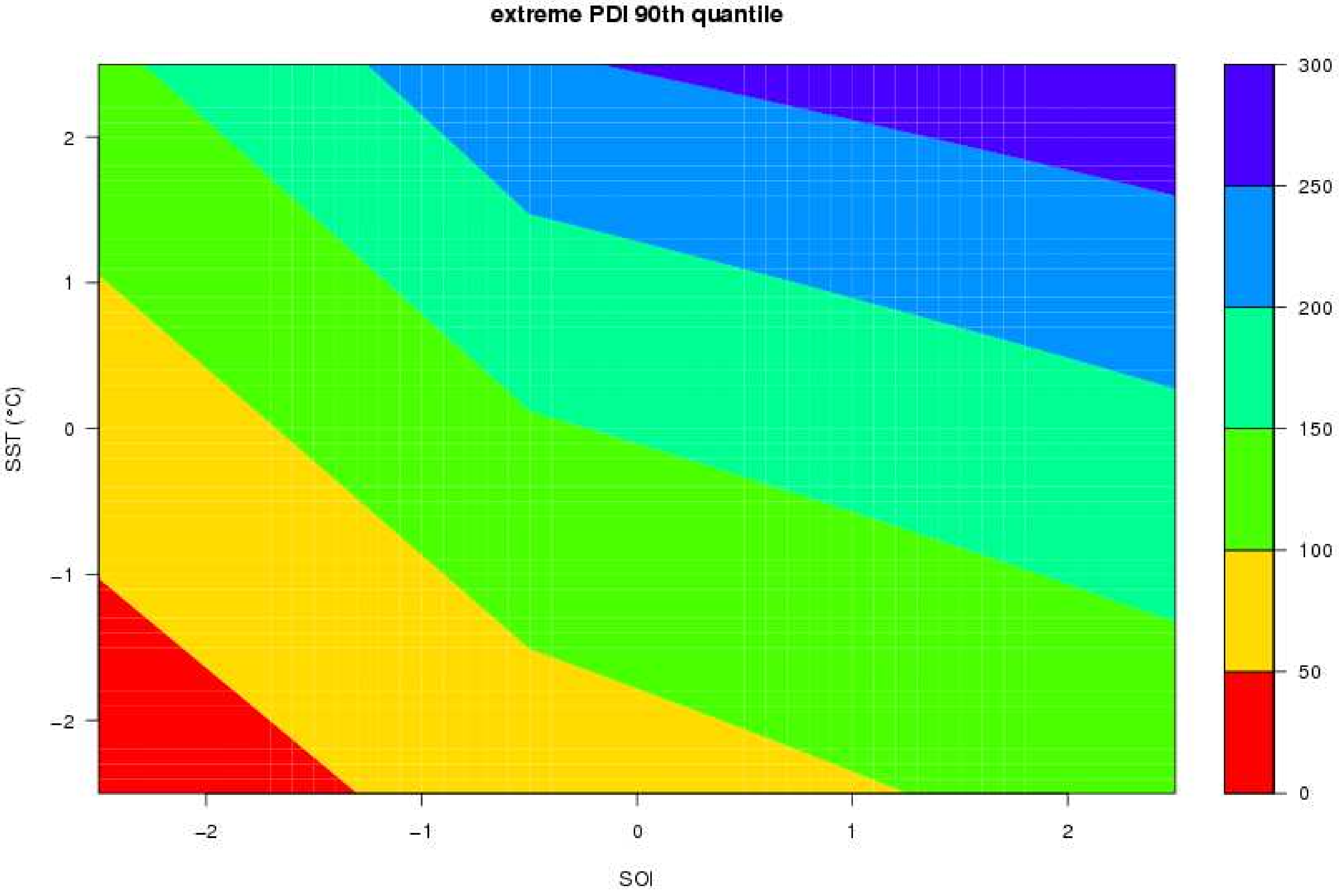
- Modèle paramétrique suggéré par les résultats du VGAM
- μ modélisé comme une fonction linéaire de SOI et SST
- σ modélisé comme une fonction linéaire de SST+modèle à rupture en SOI
- Inférence (tests de déviance)
  - Approximation Gumbel valide (p value : 0.36)
  - Rupture autour de -0.55hPa

Paramètres de la distribution Gumbel

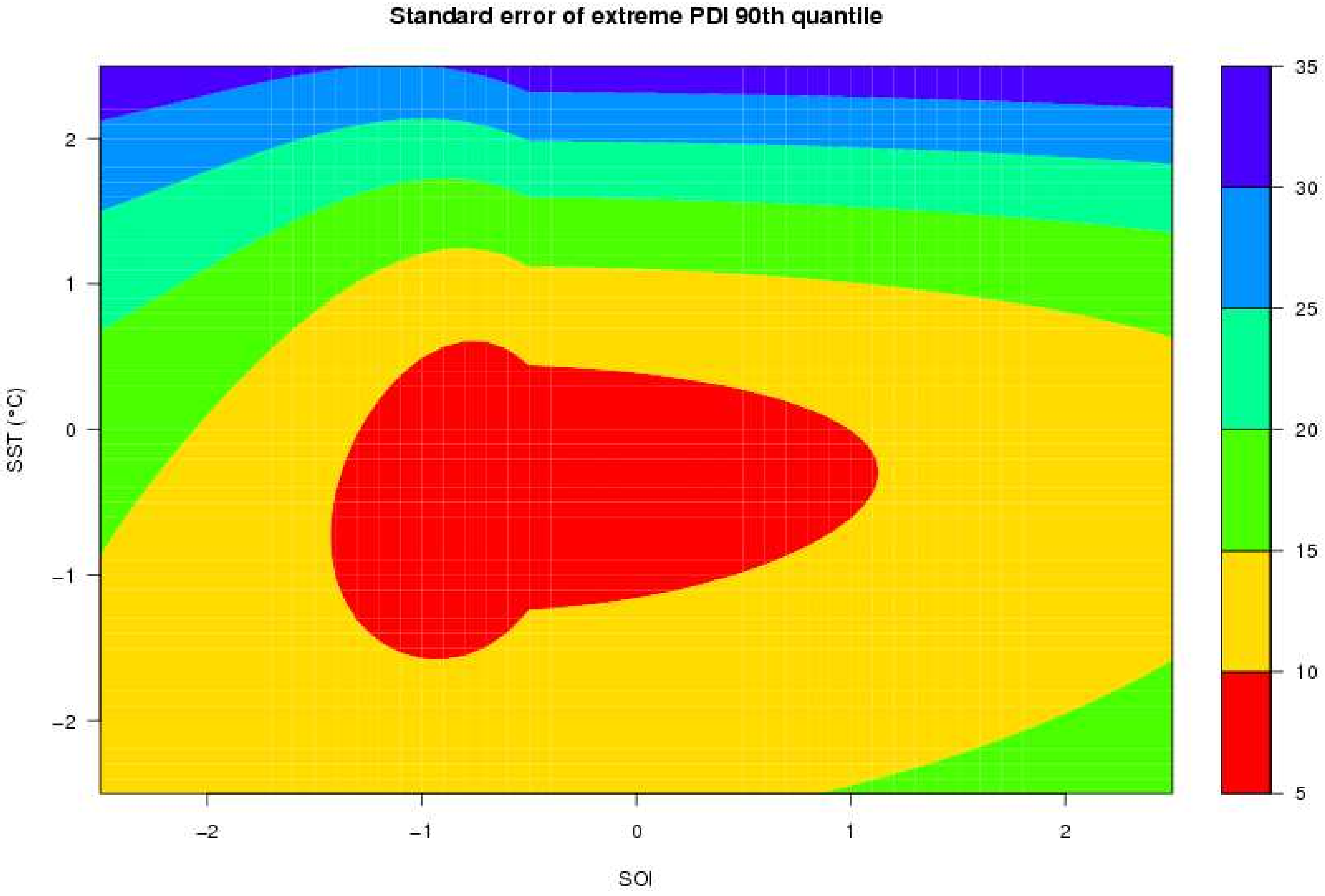




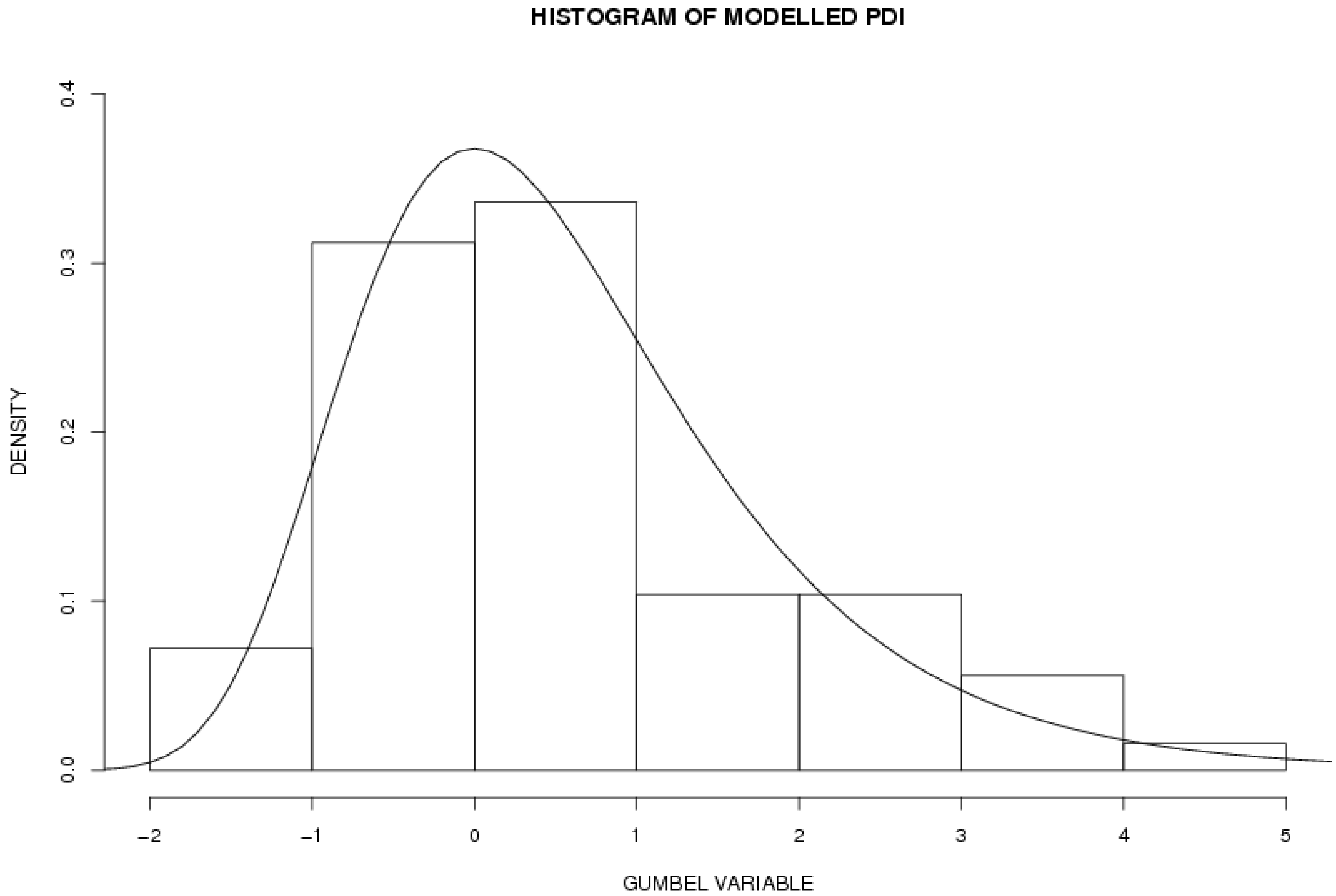
Quantile 0.90 de la PDI max annuelle



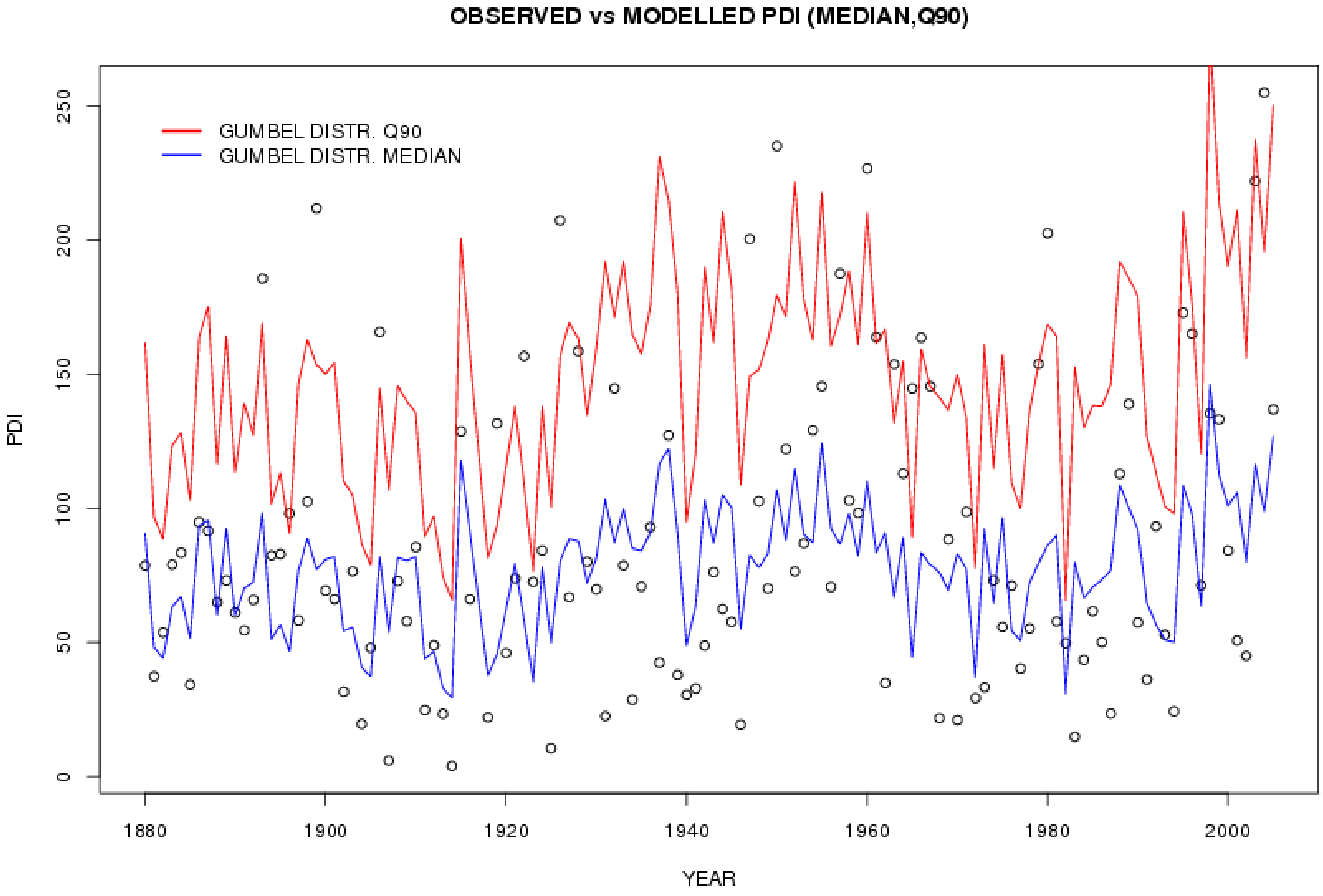
Erreur-type (méthode Delta)



Model fit


HISTOGRAM OF MODELLED PDI
DENSITY
0.0
0.1
0.2
0.3
0.4
−2
−1
0
1
2
3
4
5
GUMBEL VARIABLE

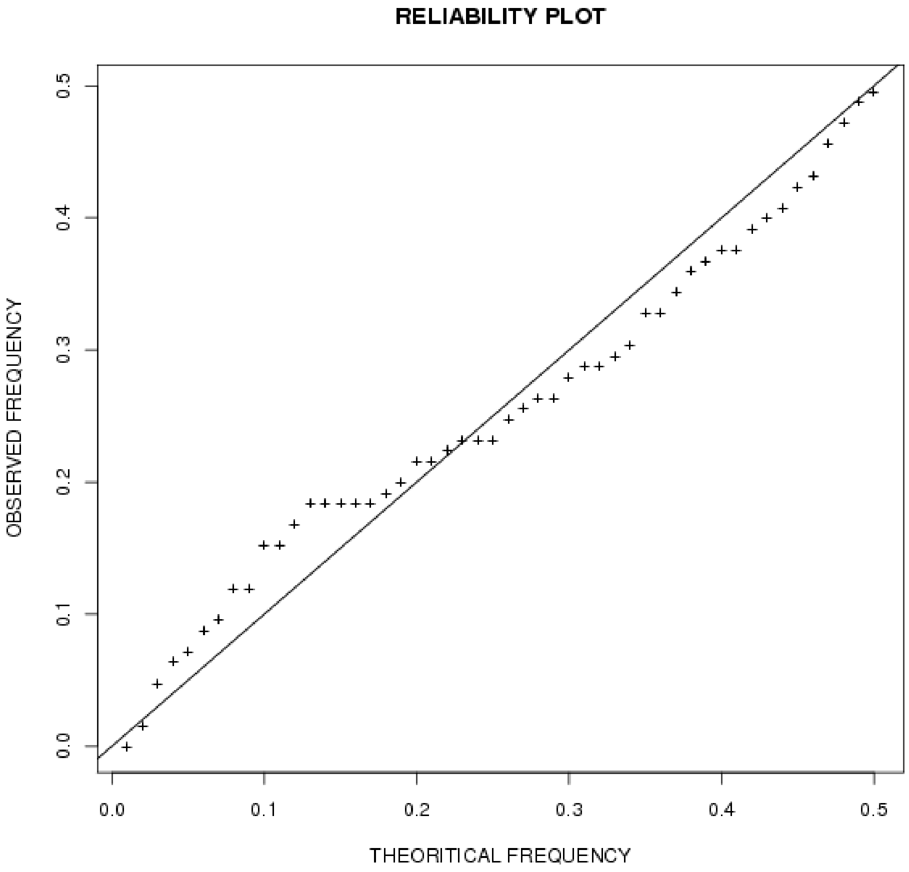
Prediction



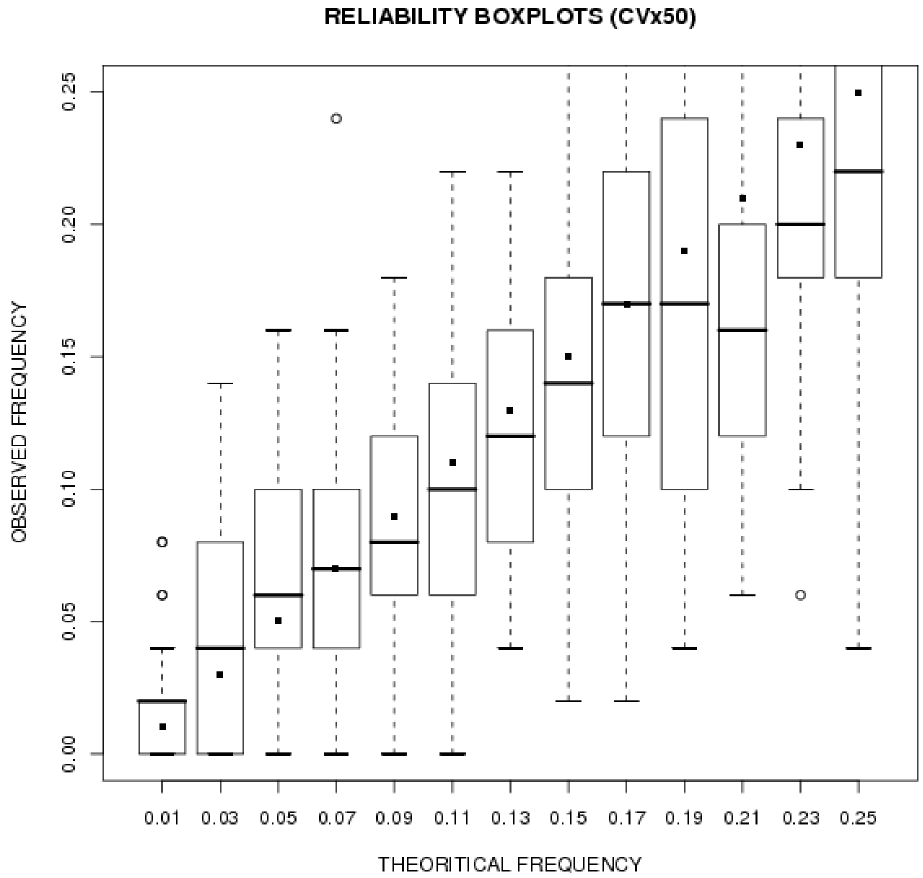
OBSERVED vs MODELLED PDI (MEDIAN,Q90)
GUMBEL DISTR. Q90
GUMBEL DISTR. MEDIAN
PDI
0
50
100
150
200
250
YEAR
1880
1900
1920
1940
1960
1980
2000

Reliability plots

- **Learning**



- **Cross validation**



CONCLUSION

- GAM & VGAM
- Basé sur les données ☺
- Technique exploratoire flexible ☺
- Inférences moins précises 😐
- Pas beaucoup de prédicteurs 😐
- Problèmes de convergence possibles 😐

Bibliographie

- GAM

  Hastie & Tibshirani, 1990

  Generalized Additive Models, Monographs on statistics and applied probability 43, Chapman & Hall/CRC, 335 p.

- VGAM

  Yee & Wild, 1996

  Vector Generalized Additive Models

  JRSS series B, Vol. 58, n°3, pp. 481-493

Bibliographie

- VGAM et extremes

Yee & Stephenson, 2007

Vector Generalized and Additive Extreme Value Models.

Chavez-Demoulin & Davison, 2005

Generalized Additive Modelling of sample extremes

Applied Statistics 54, 207-222.

Calcul

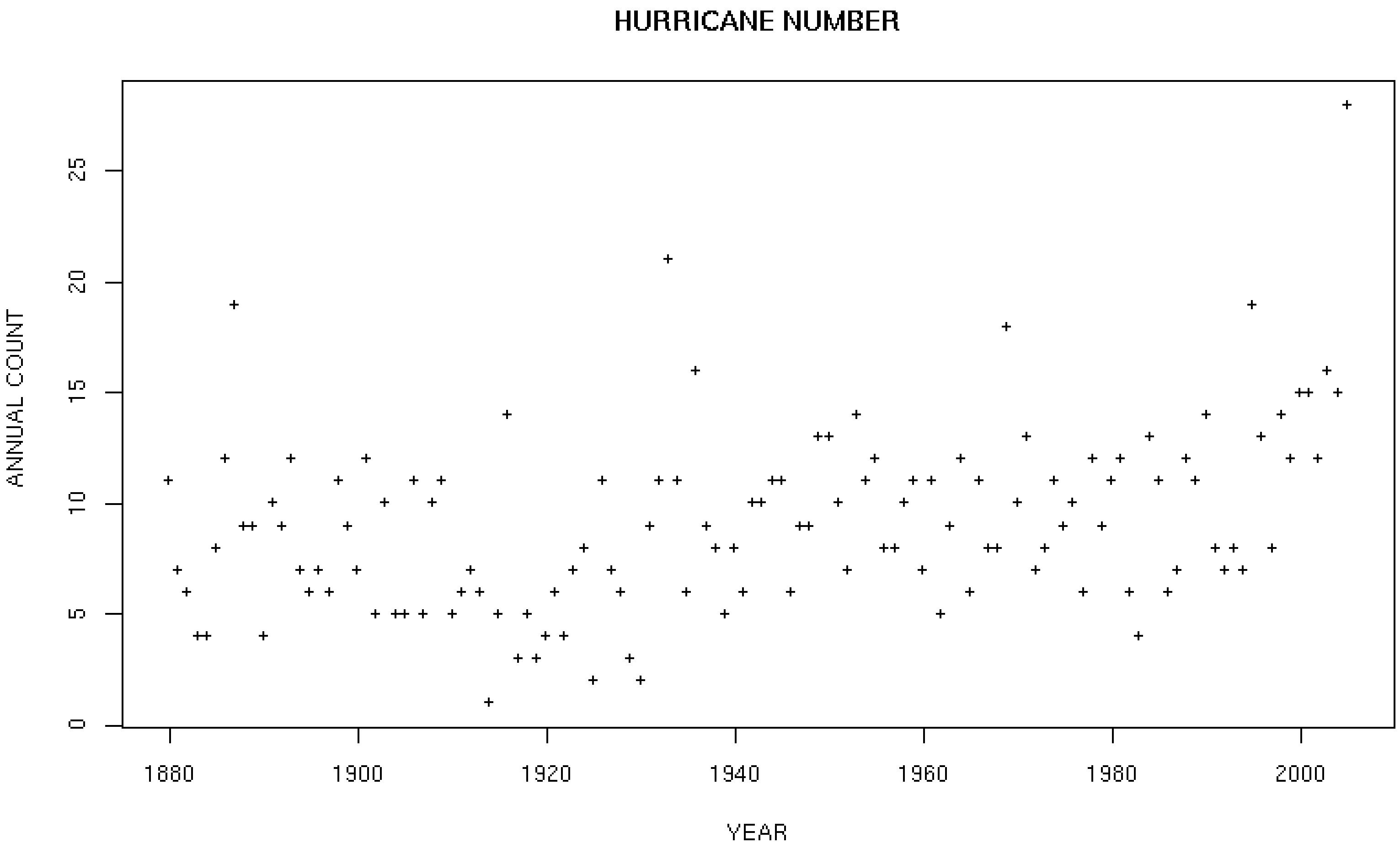
- Librairies R

  « gam » Hastie

  « VGAM » Yee, 2006

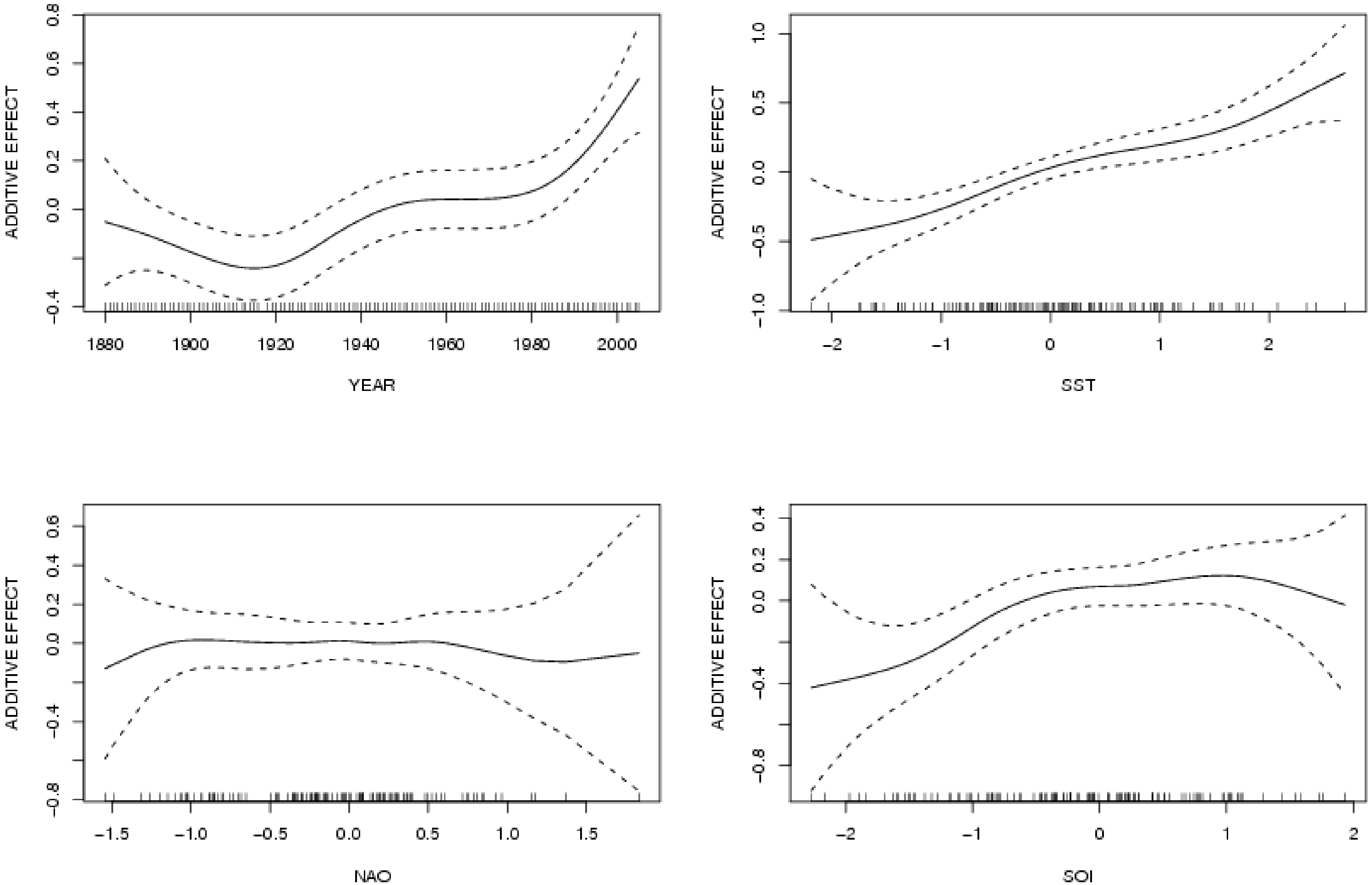
Nombre annuel de Cyclones

- Nombre de cyclones ~ distribution Poisson



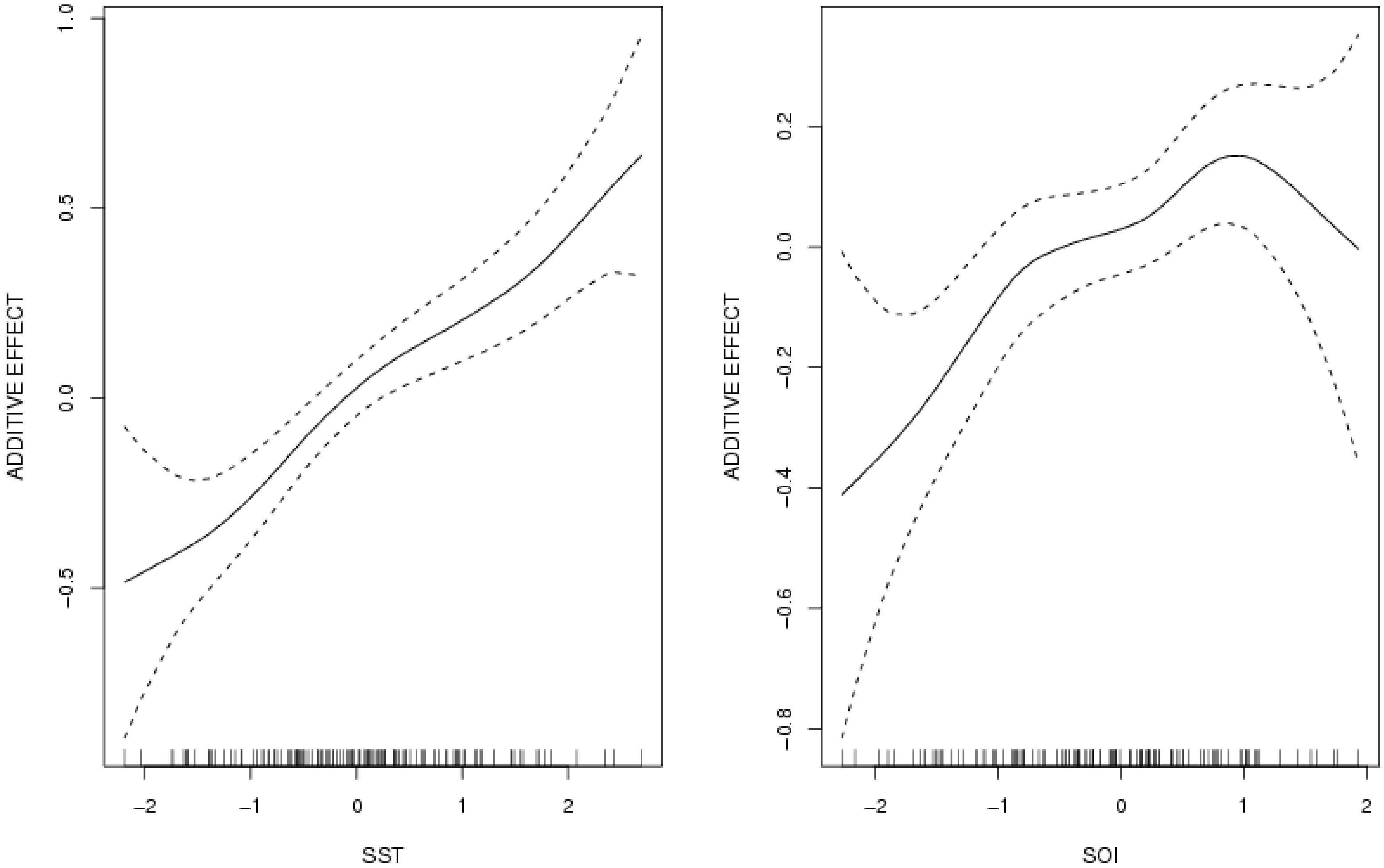
Facteurs influençant le nombre de cyclones

- Modélisation GAM, prédicteurs (YEAR,SST,SOI,NAO)



Modèle GAM

- $\text{Log}(\lambda) = \beta_o + S_1(\text{SST}) + S_2(\text{SOI})$



Modèle paramétrique

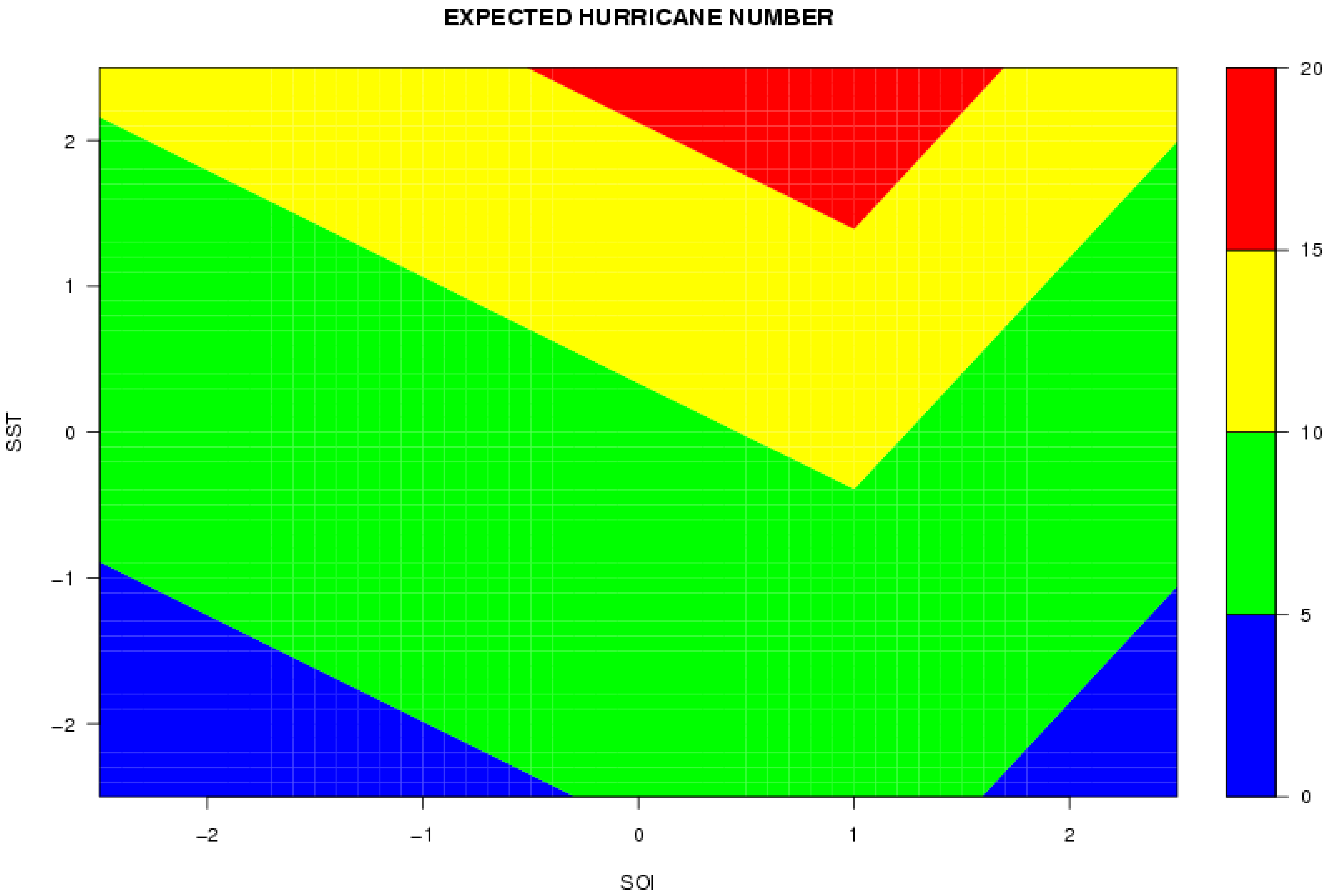
- "Modèle à ruptures" (avec contrainte de continuité) en SOI
  mis en évidence par le GAM

- $\log(\lambda) \quad = \beta o+\beta_{SOI}^{(1)}SOI+\beta_{SST}SST \qquad SOI<K$
  $\qquad\qquad = \beta o+\beta_{SOI}^{(1)}SOI+\beta_{SOI}^{(2)}(SOI-K)+\beta_{SST}SST \qquad SOI\geq K$

- Meilleur ajustement obtenu pour K=1
  log-likelihood=-316.16, à comparer avec -318.71 (linéarité)
  p value=0.02

- L'estimation du nombre de cyclones s'en déduit directement en fonction de SOI et SST

Nombre de cyclones prévu



Observé *vs* prévu: r=0.6