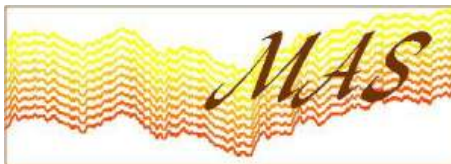


# Analyse Spatio-Temporelle de la mousson en Afrique de l'Ouest

Laurence **Viry**

en collaboration avec

Clémentine **Prieur**, Anestis **ANTONIADIS** et Hubert **Gallée**



# Projet Multidisciplinaire

Ce travail implique :

- **Physiciens :**

- ★ Hubert **GALLÉE** (LGGE, Grenoble),
- ★ Christophe **MESSAGER** (IFREMER, Brest).

- **Statisticiens :**

- ★ Anestis **ANTONIADIS** (SMS,LJK Grenoble),
- ★ Céline **HELBERT** (MOISE, LJK/INRIA Grenoble),
- ★ Clémentine **PRIEUR** (MOISE, LJK/INRIA Grenoble),
- ★ Laurence **VIRY** (CIMENT/MOISE, UJF/INRIA Grenoble).

- **Experts en Calcul Intensif :**

- ★ Laurence **VIRY** (CIMENT/MOISE, UJF/INRIA Grenoble).
- ★ Eddy **CARON** (GRAAL, ENS/INRIA Lyon).

# Contents

- 1 Modèle Physique
- 2 Modèle atmosphérique régional (code de simulation)
- 3 Modélisation stochastique
  - Motivations
  - Modélisation stochastique des Entrées/Sorties

# Sommaire

- 1 **Modèle Physique**
- 2 **Modèle atmosphérique régional (code de simulation)**
- 3 **Modélisation stochastique**
  - Motivations
  - Modélisation stochastique des Entrées/Sorties

## Contexte Applicatif (projet AMMA)

Le climat de l'Afrique de l'ouest est piloté par un système de mousson avec une activité dynamique de Mai à Septembre.

Le cumul des précipitations en zone Sahélienne est très faible (500-600 mm)

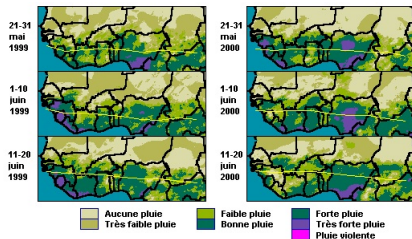
Une faible variabilité à la baisse de ce cumul peut entraîner un stress hydrique irréversible pour les cultures et la végétation.

Si ce cumul perdure d'une année sur l'autre, les conséquences deviennent durables sur les écosystèmes qui tendent à passer à des écosystèmes de type Saharien

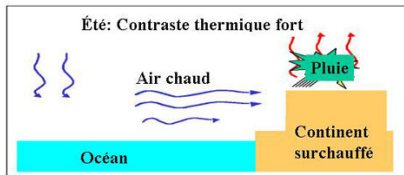
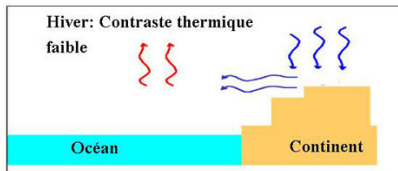
# Mousson en Afrique de l'Ouest

La mousson en Afrique de l'Ouest est liée à

- au déplacement semi-annuel de la Zone de convergence intertropicale (ZCIT)
- au gradient de température entre la zone (Sub)-Saharienne et la côte atlantique équatorial dans le golfe de Guinée.
- les alizés secs du nord-est (particulièrement leur forme plus intense l'harmattan) sont remplacés par les vents de la mousson du Sud-Ouest durant l'été.



# Contrastes thermiques



# Objectifs du projet

- Quels sont les principaux paramètres **responsables de la variabilité de la mousson** (température de surface de l'océan, albedo, végétation, ...)?
- Quelle est la réponse du climat Africain à la température de l'eau en surface du Golfe de Guinée?
- Quelle est la **dynamique spatio-temporelle** de la mousson en Afrique de l'Ouest?
- Comment effectuer une analyse de sensibilité?



# Sommaire

- 1 Modèle Physique
- 2 **Modèle atmosphérique régional (code de simulation)**
- 3 Modélisation stochastique
  - Motivations
  - Modélisation stochastique des Entrées/Sorties

# Contexte de l'analyse de sensibilité

Phénomènes physiques  $\Rightarrow$  Modèles Mathématiques  $\Rightarrow$  Codes de Simulation

**Modèle:** Le modèle mathématique étudié est décrit par une fonction déterministe  $\eta$  qui fait correspondre à chaque entrée  $(x_1, x_2, \dots, x_n)$ , une sortie scalaire ou vectorielle  $y = (y_1, y_2, \dots, y_p)$

**Enjeux:** Les codes sont de plus en plus complexes, ils prennent en compte de nombreuses variables en entrées et peuvent délivrer un grand nombre de résultats

**Objectifs:** Développer des méthodes performantes, même en grande dimension, permettant de mesurer la variabilité des sorties et de déterminer les facteurs en entrées les plus importants

# MAR - Modèle atmosphérique régional

- **Modèle mathématique** (modèle climatique régional)
  - ★ **Atmosphère**
    - Équations primitives hydrostatiques (Navier-Stokes)
    - Paramétrisation des processus dynamiques sous-maille (turbulence, diffusion horizontale, filtre numérique)
    - Cycle hydrologique: equations de conservation
      - vapeur d'eau, gouttelettes et cristaux nuageux,
      - gouttes de pluie et flocons de neige
    - Paramétrisation des nuages sous-maillles et convection atmosphérique
  - ★ **Surface**
    - conservation de la chaleur et de l'eau des sols

# Le code de simulation MAR

Exige **des ressources de calcul importantes** , avec un **grand nombre de runs**

- un run : 15 jours, ~15H CPU,
- pour une année simulée utilisée, on simule 2 années  $\Rightarrow$  **48 (2x24) runs** pour une année, on simule 17 années,
- 3 - 4 Go pour les Entrées/Sorties,
- **Interdépendance entre les Entrées/Sorties des runs**: les sorties de la simulation d'un mois sont utilisées en entrées du mois suivant.

$\Rightarrow$  **Déploiement sur grille** : méso-centre Grenoblois (CIMENT)

$\Rightarrow$  **On soumet en parallèle** sur une dizaine de noeuds **suivant la disponibilité des plates-formes**.

# DIET

**Approche** : déploiement sur grille avec le middleware **DIET**  
(Distributed Interactive Engineering Toolbox)  
<http://graal.ens-lyon.fr/DIET/>

**DIET** est développé par l'équipe **GRAAL** (Algorithms and Scheduling for Distributed Heterogeneous Platforms) du **LIP** (Laboratoire de l'Informatique du Parallélisme) de l' ENS Lyon.

C'est :

- Une **boîte à outils** dont le but est de rendre l'utilisation d'une grille informatique aussi transparente que possible.
- Un **'intergiciel'** qui se situe entre la partie applicative et les ressources informatiques disponibles sur la grille.

# DIET (Distributed Interactive Engineering Toolbox)

Permet de déployer de **grosses applications** sur des **grilles de calculateurs**

DIET utilise le **modèle client/serveur** :

- détermine les ressources les plus adaptées aux informations fournies par la requête du client,
- il se charge de :

**scheduling**: quelle machine est la mieux à même d'effectuer une tâche à un moment donné.

**Data managment**: persistance des données, (re)distribution des données, workflow...

**interopérabilité**: une grille pouvant être très hétérogène

# Sommaire

- 1 Modèle Physique
- 2 Modèle atmosphérique régional (code de simulation)
- 3 **Modélisation stochastique**
  - Motivations
  - Modélisation stochastique des Entrées/Sorties

# Méta-modèle

- Pour contourner les problèmes dus à **la complexité du modèle**,
- Pour obtenir **des informations globales** sur tout le domaine de variabilité des facteurs.

Nous construisons un **méta-modèle** par une **approche stochastique**,

Les principales propriétés requises pour le méta-modèle :

- tenir compte de **la dynamique spatio-temporelle**,
- faire un calcul rapide des estimations des sorties rendant possible le calcul des indices de sensibilité (indices de Sobol).

⇒ La première étape concerne **la modélisation fonctionnelle** des entrées/sorties

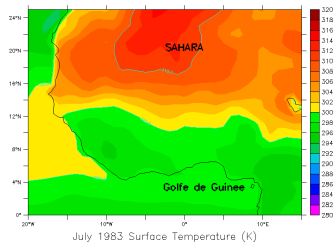
Température de surface de l'Océan  $\xrightarrow{\text{méta-modèle}}$  Précipitation



# Modélisation des entrées/sorties

## Entrées : température de surface de l'océan

- données climatologiques de Reynolds (satellites et in situ),
- données hebdomadaires sur 17 ans,
- observations sur un maillage  $\mathcal{G}$  couvrant la zone  $\mathcal{R}$ ,  $[5S : 5N] \times [30W : 10E]$ , avec un pas d'espace de  $1^\circ$  latitude et longitude (516 points).



## Sorties : précipitations

- sorties hebdomadaires de MAR sur 17 ans,
- obtenues sur un maillage  $\mathcal{G}'$  couvrant la zone  $\mathcal{R}'$ ,  $[10S : 30N] \times [20W : 20E]$ .

# Cadre Statistique

## Données :

La mousson en Afrique de l'Ouest est un phénomène périodique observé sur une période active de Mars à Octobre sur 17 années.

$\mathbf{X}^i := (X_i(x, t))_{x \in \mathcal{R}, t \in \mathcal{T}}$  TSO année  $i$

$\mathbf{Y}^i := (Y_i(x', t))_{x' \in \mathcal{R}', t \in \mathcal{T}}$  précipitations année  $i$

## Méthode :

Soit  $x_0 \in \mathcal{G} \subset \mathcal{R}$ .

Soit  $y_0 \in \mathcal{G}' \subset \mathcal{R}'$ .

L'entrée au point  $x_0$  est une trajectoire aléatoire  $(X_i^{x_0}), i = 1, \dots, 17$ .

La sortie au point  $y_0$  est une trajectoire aléatoire  $(Y_i^{y_0}), i = 1, \dots, 17$ .

# Modélisation statistique de la TSO

On suppose que  $X_i^{x_0}$  appartient à un espace de Hilbert fonctionnel  $\mathbb{H} \subset \mathbb{L}^2(\mathcal{T})$ .

$X^{x_0}$  est régulière, avec

- une fonction moyenne inconnue régulière  $\mathbb{E}X^{x_0}(t) = \mu_{X^{x_0}}(t)$ ,
- une fonction de covariance inconnue régulière  $\text{Cov}(X^{x_0}(s), X^{x_0}(t)) = G_{X^{x_0}}(s, t)$ .

On suppose qu'il existe une décomposition orthogonale de  $G_{X^{x_0}}$  :

$$G_{X^{x_0}}(s, t) = \sum_{m \geq 1} \rho_m(x_0) e_m(x_0, s) e_m(x_0, t), \quad s, t \in \mathcal{T},$$

avec des valeurs propres  $\rho_1(x_0) \geq \rho_2(x_0) \geq \dots$

# Décomposition de Karuhnen-Loève

On effectue une décomposition de Karuhnen-Loève :

$$X^{x_0}(t) = \mu_{X^{x_0}}(t) + \sum_{m=1}^{\infty} \alpha_m(x_0) e_m(x_0, t), \quad t \in \mathcal{T}.$$

Pour un point fixé de la grille, nous utilisons une troncature appropriée

$$\tilde{X}^{x_0}(t) = \mu_{X^{x_0}}(t) + \sum_{m=1}^N \alpha_m(x_0) e_m(x_0, t), \quad t \in \mathcal{T}.$$

# Hypothèse principale

La base tronquée, construite à partir du point fixé  $x_0$ , reste appropriée pour la décomposition de  $X^x(t)$ ,  $\forall x \in \mathcal{G}$ .

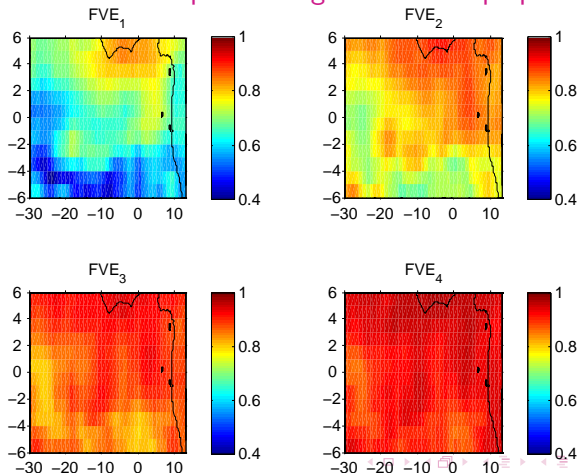
$$\tilde{X}^x(t) = \mu_{X^x}(t) + \sum_{m \geq 1}^N \alpha_m(x) e_m(x_0, t)$$

fournit une bonne approximation pour  $\{X^x(t), t \in \mathcal{T}, x \in \mathcal{G}\}$ .

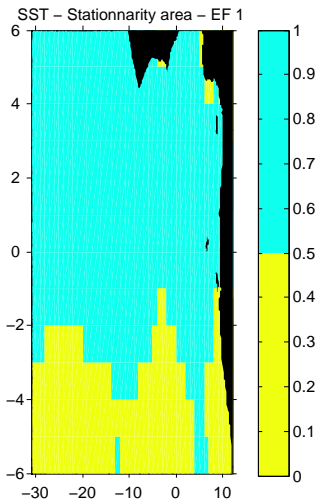
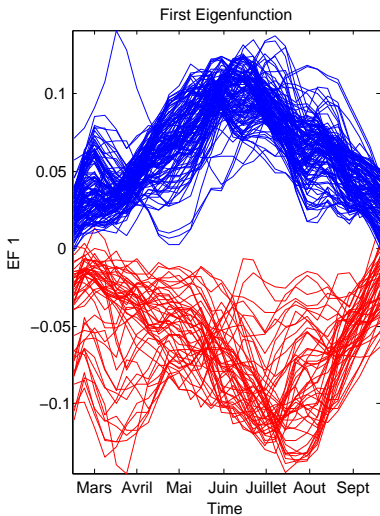
# Application

On effectue une décomposition de KL en chaque point  $x$  du maillage  $\mathcal{G}$ .

Le critère de troncature est le **pourcentage d'inertie expliquée**.

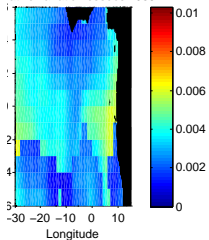


## Hypothèse de stationnarité de la base fonctionnelle.

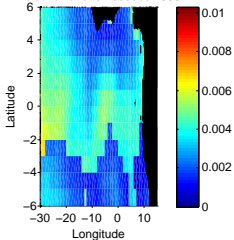


# Qualité de l'estimation de la SST sur $\mathcal{G}$

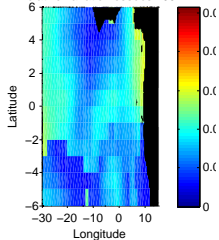
Relative variation est/obs 1985



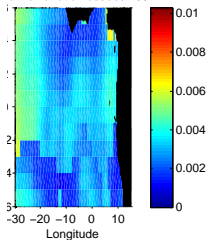
Relative variation est/obs 1988



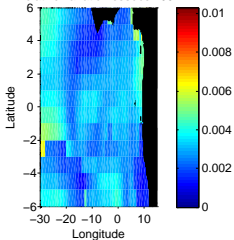
Relative variation est/obs 1991



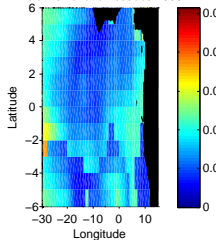
Relative variation est/obs 1994



Relative variation est/obs 1997



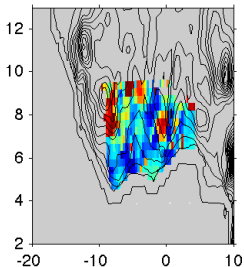
Relative variation est/obs 2000



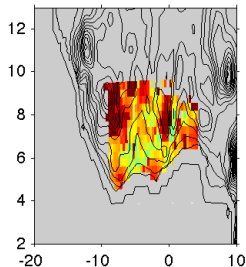


# Traitement identique sur les précipitations

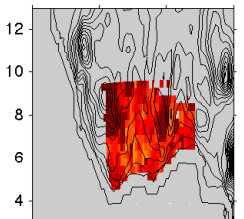
Pourcentage d'inertie par 1 fp



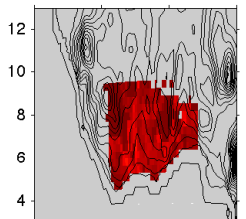
Pourcentage d'inertie par 2 fp



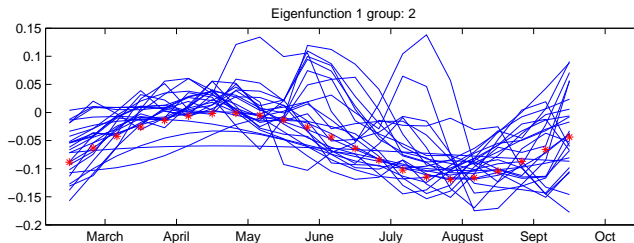
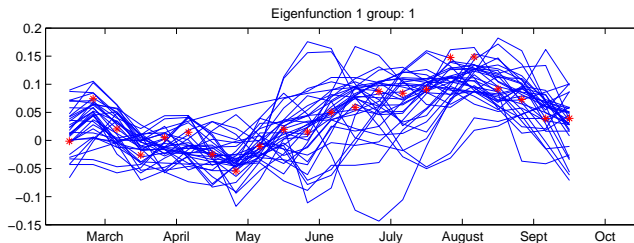
Pourcentage d'inertie par 3 fp



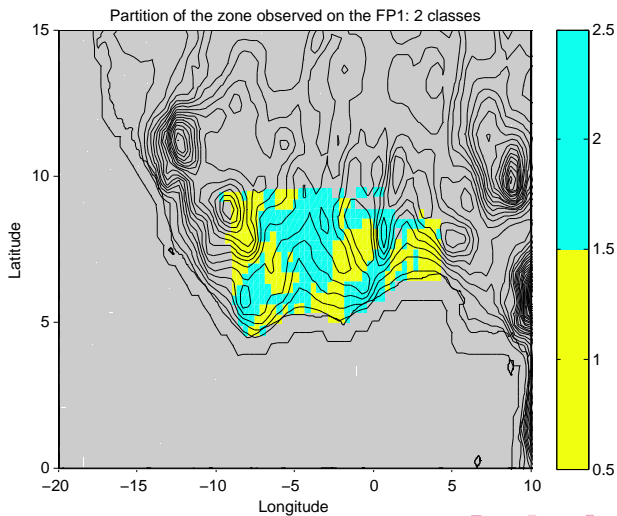
Décomposition de Karhunen-Loève - Précipitations



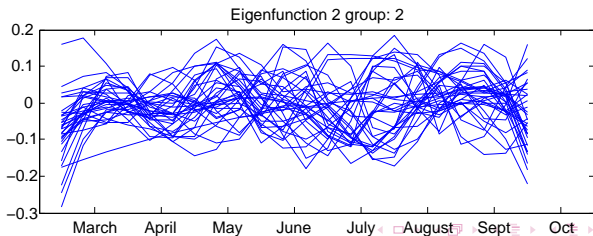
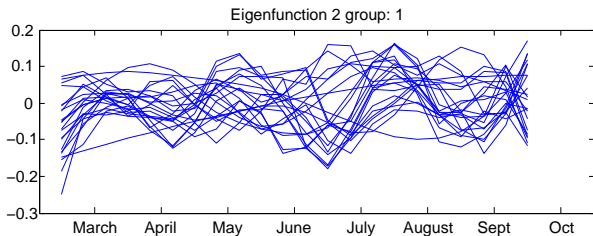
## Hypothèse de stationnarité de la base fonctionnelle : première fonction propre



## Visualisation des deux groupes sur la carte :



Hypothèse de stationnarité de la base fonctionnelle : deuxième fonction propre



# Modélisation des sorties du modèle

Modèle :

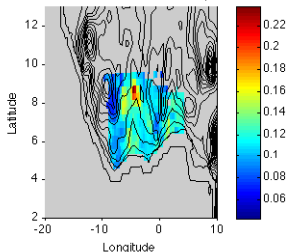
On divise  $\mathcal{R}'$  en deux régions,  $\mathcal{R}'_1$  et  $\mathcal{R}'_2$ . Soit  $y_0 \in \mathcal{R}'_1$ ,  $z_0 \in \mathcal{R}'_2$ .

$$\text{Si } y \in \mathcal{R}'_1, \tilde{Y}^y(t) = \mu_{Y^y}(t) + \sum_{k \geq 1}^K \beta_k(y) f_k(y_0, t).$$

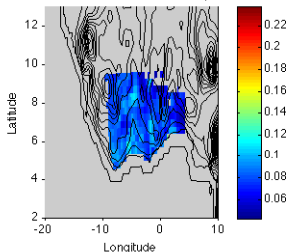
$$\text{Si } z \in \mathcal{R}'_2, \tilde{Y}^z(t) = \mu_{Y^z}(t) + \sum_{l \geq 1}^L \beta_l(y) g_l(z_0, t).$$

# Qualité de l'estimation des précipitations sur $\mathcal{G}'$

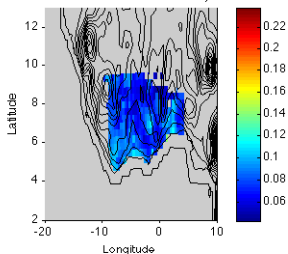
Relative variation estimated/observed - year 1983



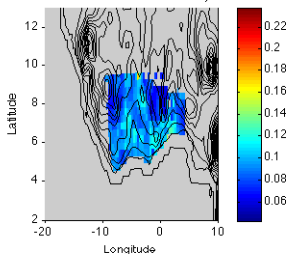
Relative variation estimated/observed - year 1984



Relative variation estimated/observed - year 1985



Relative variation estimated/observed - year 1986



# Conclusions et Perspectives

- ★ Notre **modélisation spatio-temporelle** pour les entrées (resp. les sorties) semble pertinente :
  - de faibles erreurs sur la grille,
  - validation physique des résultats, . . .
  - des interprétations physiques approfondies semblent opportunes.
  
- ★ Exécuter le code MAR à l'aide du middleware DIET:
  - a permis **la gestion d'un grand nombre de "runs"** sur un réseau local de machines de façon "scalable" et fiable,
  - permet d'envisager **la prise en compte de paramètres supplémentaires** du modèle (albédo, . . . )

## Conclusions et Perspectives (suite)

**Perspectives** : Nous voulons modéliser la régression de la pluviométrie sur la température de surface de l'océan.

**Méthodologie** : on régresse  $(\beta_1(y), \beta_2(y))$ ,  $y \in \mathcal{R}'$  sur  $(\alpha_1(x), \alpha_2(x))$ ,  $x \in \mathcal{R}$ .

Problème de grande dimension: approche fonctionnelle ou réduction de dimension.



Merci pour votre attention!