### UNIVERSITE CHEIKH ANTA DIOP DE DAKAR



### FACULTE : SCIENCES ET TECHNIQUES

### ECOLE DOCTORALE : MATHEMATIQUES ET INFORMATIQUE



UNIVERSITE PIERRE ET MARIE CURIE, PARIS, FRANCE

Année : 2016

### THESE DE DOCTORAT

<u>Spécialité :</u> INFORMATIQUE Option : Systèmes Complexes

> Présentée par : François KALY

# Titre: Etude statistique de la variabilité des teneurs atmosphériques en aérosols désertiques en Afrique de l'Ouest.

Soutenue le 13/04/2016 devant le jury composé de :

Président	MAMPASSI Benjamin	Professeur	UCAD, Dakar, Sénégal
Rapporteurs	DIAKHABY Aboubakary	Professeur	UGB, St Louis, Sénégal
	CHIAPELLO Isabelle	Chargé de Recherche avec HDR	LOA, Lille 1, France
	CAMARA Moctar	Maitre de Conférences	UASZ, Ziguinchor, Sénégal
Examinateurs	PICON Laurence	Professeur	UPMC, Paris, France
	NIANG Awa	Professeur	UCAD, Dakar, Sénégal
Directeur(s) de thèse	JANICOT Serge	Directeur de Recherche	UPMC, Paris, France
	MARTICORENA Béatrice	Chargé de Recherche avec HDR	LISA, Paris, France

Etude statistique de la variabilité des teneurs atmosphériques en aérosols désertiques en Afrique de l'Ouest.

François KALY

9 mai 2016

## Remerciements

Je souhaite tout d'abord remercier Serge JANICOT, Béatrice MARTICORENA et Awa NIANG qui m'ont permis d'aller au bout de ce travail. Durant cette période de thèse, ils ont toujours été présent et ont su me guider dans mon travail. Ils ont toujours su apporter le regard critique nécessaire sur mes travaux et me proposer des voies de recherche pertinentes. Ils ont su me donner l'envie et la motivation de mener ces travaux à bien, et pour cela je leurs dois beaucoup.

Merci à Madame Sylvie THIRIA et toute l'équipe MMSA (Modélisation et Méthodes Statistiques Avancées) du Laboratoire d'Océanographie et du Climat : Expérimentation et Approches Numériques (LOCEAN) de l'Université Pierre et Marie Curie pour m'avoir accueilli au sein de l'équipe et d'avoir faciliter mon intégration au sein de ce laboratoire. Ils m'ont fourni un cadre de travail idéal pendant ma thèse et m'ont permis de développer mes idées. Ils ont toujours su apporter leurs expertises et un regard critique nécessaire sur les aspects techniques et applicatifs de mon travail.

Je tiens à remercier également Madame Ndèye NIANG pour m'avoir accueilli à plusieurs reprises au CNAM pendant mes séjours à Paris. Elle a toujours su apporter son expertise et un regard critique nécessaire sur les aspects techniques et applicatifs de mon travail. Je souhaite aussi remercier Mory OUATTARA pour sa disponibilité, son aide et ses idées lors de nos discussions. Nos échanges m'ont toujours permis de prendre du recul pour mieux cerner les difficultés rencontrées.

Je tiens aussi à remercier tous les membres de mon jury d'avoir pris le temps de s'intéresser à mes recherches.

Je remercie également toute la famille pour le soutien pendant tout mon cursus. Je remercie infiniment mes parents pour tout ce qu'ils m'ont donné.

Enfin, je remercie le Département Soutien et Formation (DSF) de l'Institut de Recherche pour le Développement (IRD) pour m'avoir octroyé une allocation de recherche dans le cadre de cette thèse, me permettant ainsi de la préparer dans de très bonnes conditions.

## Résumé

Le but de cette thèse est de documenter la variabilité des teneurs en aérosols minéraux en Afrique de l'Ouest et de comprendre les mécanismes qui la contrôlent.

L'analyse des concentrations mesurées au sol au Sahel montre que ce sont les transports d'aérosols sahariens qui sont responsables des maxima de concentration observés en saison sèche Marticorena et al. [2010]. Ces évènements présentent une très forte variabilité à l'échelle intra-saisonnière et interannuelle mais une certaine persistance à l'échelle régionale. Quelles sont les conditions météorologiques qui expliquent cette variabilité? La réponse à cette question ne peut être obtenue qu'au travers d'une analyse systématique et conjointe des conditions météorologiques régionales et des mesures d'aérosol désertiques sur le continent (concentration de surface, épaisseurs optiques en aérosols) et au large de l'Afrique de l'Ouest. Les mesures d'épaisseur optique en aérosols (AOT) permettent la surveillance globale du contenu atmosphérique en particules. Cependant la relation permettant de retrouver les concentrations massiques en  $(PM_{10})$  à partir des mesures d'épaisseur optique n'est pas toujours simple. Ainsi dans un premier temps l'analyse de la variabilité de la concentration en  $PM_{10}$  des conditions météorologiques locales mesurées au sol, a permis de confirmer un cycle saisonnier régional présenté par Marticorena et al. [2010], caractérisé par un maximum de concentration en saison sèche et un minimum pendant la saison des pluies sur l'ensemble de trois stations. L'analyse du cycle diurne des concentrations a permis de voir qu'il est diffèrent selon les saisons. En saison sèche, le jet de basses couches (NLLJ en anglais) saharien semble moduler le cycle diurne des concentrations en particules mesurées au Sahel, tandis que l'évolution diurne des concentrations de poussière en saison des pluies apparaît être modulé par l'évolution des systèmes convectifs a l'échelle régionale. Dans un deuxième temps, une méthodologie basée sur une classification en types de temps a été proposée. Elle permet de mettre en évidence des situations météorologiques typiques pour lesquelles la relation AOT-  $PM_{10}$  est simplifiée, puis une modélisation de  $PM_{10}$  à partir de l'épaisseur optique. Pour cela deux approches ont été adoptées. D'abord une famille composée de six régimes de temps bruts a été définie décrivant la saisonnalité climatique. La mise en relation AOT- $PM_{10}$  permet de voir que les régimes de temps caractéristiques du flux d'harmattan présentent une relation meilleure en terme de corrélation par rapport à la relation avant la séparation en type de temps. Ensuite six régimes de temps ayant un impact sur la variabilité interannuelle des aérosols ont été définis. Les régimes de temps obtenus sont caractérisés par une forte variabilité interannuelle et moins persistante que pour les régimes de temps bruts. Les régimes de temps sont caractérisés par des anomalies de circulation atmosphérique, soit cyclonique, soit anticyclonique, d'échelle synoptique et centrée sur le nord du domaine. L'intégration de l'ensemble de cette approche en régimes de temps permet d'obtenir des résultats satisfaisants dans la prévision des valeurs de  $PM_{10}$ , ouvrant des perspectives en terme de système d'alerte précoce.

#### Mots-clefs

régimes de temps, 2S - SOM, carte topologique auto organisatrice, classification ascendante hiérarchique,  $PM_{10}$ , AOT, Afrique de l'Ouest.

### Statistical study of the variability of atmospheric dust content over West Africa

#### Abstract

The aim of this thesis is to document the variability of the atmospheric dust content in West Africa and to understand the mechanisms that control it. The analysis of the concentrations measured ground in the Sahel shows that it is the transport of Saharan aerosols that are responsible for the maximum concentration observed in the dry season [Marticorena et al., 2010]. These events have a very high variability at intra-seasonal and interannual scale with some persistence on a regional scale. What are the weather conditions that explain this variability? The answer to this question can be obtained only through a systematic analysis of regional weather and desert aerosol measurements on the continent (surface concentration, aerosol optical thickness) and of the West Africa. The aerosol optical thickness measurements (AOT) allow global monitoring of atmospheric particulate content. However the relationship allowing to find the concentrations (  $PM_{10}$ ) from the optical thickness measurement is not always simple. So at first the analysis of the variability of the  $PM_{10}$  concentration and local meteorological conditions measured on the ground allowed to confirmed a regional seasonal cycle presented by Marticorena et al. [2010], characterized by a maximum concentration during the dry season and a minimum during the rainy season on all three stations. The analysis of the diurnal cycle concentrations allowed seeing that it differ depending on the season. In the dry season, the Saharian low level jet (NLLJ) appears to modulate the diurnal cycle of particulate concentrations in the Sahel, while the diurnal evolution of concentrations of the rainy season appears to be modulated by the evolution of convective systems at regional scale. Secondly, a methodology based on a classification into weather types was proposed. It allows to highlight typical meteorological situations in which the AOT-  $PM_{10}$  relationship is simplified and modeling  $PM_{10}$  from the Aerosol Optical Thickness. For this, two approaches have been adopted. At first a family of six weather types has been defined describing the climate seasonality. The link between AOT and  $PM_{10}$  can allowed to see that the weather types characterise to harmattan flow have a better relationship in terms of correlation to the relationship before separating in weather type. Then six weather types affecting the interannual variability of aerosols were defined. The resulting weather patterns are characterized by high interannual variability and less persistent than for rough weather regimes. The weather patterns are characterized by atmospheric circulation anomalies, either cyclonic or anticyclonic at synoptic scale and centered on the northern area. The integration of all of this approach in weather types provides satisfactory results in forecasting  $PM_{10}$  values, opening perspectives in terms of early warning system.

#### Keywords

weather regims,  $2S_SOM$ , self-organizing map, hierarchical clustering,  $PM_{10}$ , AOT, West Africa

## Table des matières

Ta	able o	des figu	ures	8
$\mathbf{Li}$	ste d	les tab	leaux	12
In	trod	uction		<b>14</b>
1	Gér	néralité	s sur la circulation atmosphérique et les aérosols désertiques	
	$\mathbf{en}$	Afrique	e de l'ouest.	17
	1.1	Introd	uction	17
	1.2	Généra	alités sur la circulation atmosphérique en Afrique de l'Ouest	17
		1.2.1	La circulation en saison sèche	17
		1.2.2	La circulation en saison des pluies (saison de Mousson)	19
			1.2.2.1 Circulation dans les basses couches	20
			1.2.2.2 Circulations dans les couches moyennes et dans la haute	
			$troposphère \ldots \ldots$	21
			1.2.2.2.a Jet d'Est Africain $(JEA)$ et ondes d'Est	21
			1.2.2.2.b Jet d'Est Tropical $(JET)$	21
			1.2.2.2.c Jet d'Ouest Subtropical (JOST)	21
		1.2.3	Zone de Convergence Inter Tropicale $(ZCIT)$	22
	1.3	Généra	alités sur les aérosols désertiques	22
		1.3.1	Sources de Production et cycle de vie des aérosols désertiques	23
			1.3.1.1 Sources	23
			1.3.1.2 Cycle de vie des aérosols	24
			1.3.1.2.a Les émissions en zones source	25
			1.3.1.2.b Le transport des particules	27
		1 9 0	1.3.1.2.c Le depot des particules	28
		1.3.2	Variabilité des contenus atmospheriques en aerosols desertiques	28
			1.3.2.1 Variabilité journaliere	28
			1.3.2.2 Variabilité interappuelle	29 20
	1 /	Impag	ts des pérosols désortiques	30 31
	1.4	1/1	Impacts sur le climat	31
		1.4.1 1 4 2	La visibilité	32
		1.1.2 1 4 3	Impacts sur la santé la biogéochimie et la biodiversité	33
	1.5	Conclu	ision et hypothèses méthodologiques	34
<b>2</b>	Pré	sentati	on des données d'étude	35
	2.1	Introd	uction	35
	2.2	Présen	tation des stations de mesures et des données in-situ	35
		2.2.1	Présentation des stations de mesures	35
		2.2.2	Mesures in situ	36
			2.2.2.1 Mesures de concentrations en $PM_{10}$	36
			2.2.2.2 Mesures de l'Epaisseur Optique en Aérosol	37

		2.2.2.3 Mesures météorologiques in situ	37
2	2.3	Données météorologiques à grande échelle	37
2	2.4	Aerosol Index (AI)	38
3 I	Mét	hodes et approche statistique utilisées 4	41
3	3.1	Introduction	41
3	3.2	Méthode de filtrage de données	41
3	3.3	Analyse Factorielle des Correspondances	42
		3.3.1 Les données	42
		3.3.2 Le problème et la méthode	43
ŝ	3.4	Méthodes de Classification	43
		3.4.1 Généralités sur les méthodes de classification	43
		3.4.1.1 Les méthodes de hard subspace clustering	44
		3 4 1 2 Méthode de soft subspace clustering	45
		3.4.2 Présentation de l'Algorithme (2S-SOM)	45
		$3.4.2$ Treschauton de l'Higoronnie (25.50 M) $\cdots \cdots \cdots$	10 //6
		3 4 2 2 2 2 SOM	40 /0
ę	25	Evaluation des variables portinentes pour chaque classe de la CAH	40 51
د د	3.6	Los critòros d'évaluation d'una classification	51 54
U.	5.0	2.6.1 Critères d'évaluation interne	54 54
		2.6.2 Criticizes d'évaluation interne	56
		5.0.2 Criteres d evaluation externe	90
4	Vari	abilité temporelle des aérosols mesurés au sol au Sabel	59
-	4 1	Introduction	50
	1.1	Variabilitá des concentrations en poussiàres minérales	50
ר /	1.2 1.2	Analyse Comparative des cycles saisonniers des concentrations en $PM_{10}$ et	00
-	1.0	$d_{\rm e}$ l'AOT	77
5 ]	Rela	ations AOT- $PM_{10}$ en fonction des régimes de temps d'Afrique de	
1	L'O	uest	81
F	5.1	Présentation de la méthodologie	81
0		511 Objectifs et stratégie	81
		512 Présentation des données d'études	82
		51.3 Présentation de la méthodologie de filtrage	82
		514 Présentation de la méthodologie de classification par types de temps	83
		5.1.5 Sélection de Variables	81
F	ະາ	Págimag de temps heuts	80 80
ر	9.2	5.2.1 Détermination du nombre de régimes de temps bruts	09 00
		5.2.1 Determination du nombre de regimes de temps bruts	09 01
		5.2.2 Caracterisation des regimes de temps	00 91
		$5.2.2.1$ Matrices de confusion $\ldots \ldots \ldots$	92
		5.2.2.2 Distribution diurne à inter-annuelle	92 06
		5.2.2.3 Persistances et transitions inter-classes	
			90
		5.2.2.4 Structures spatiales et interprétations météorologiques et	90
		5.2.2.4 Structures spatiales et interprétations météorologiques et climatiques	90 99
		5.2.2.4Structures spatiales et interprétations météorologiques et climatiques $\dots \dots \dots$	90 99 25
		5.2.4Structures spatiales et interprétations météorologiques et climatiques95.2.3Relation AOT- $PM_{10}$ avec et sans régimes de temps125.2.4Conclusion12	90 99 25 38
E.	5.3	$5.2.2.4$ Structures spatiales et interprétations météorologiques et climatiques $5.2.3$ Relation AOT- $PM_{10}$ avec et sans régimes de temps $5.2.4$ Conclusion $5.2.$	90 99 25 38
Ę	5.3	5.2.2.4 Structures spatiales et interprétations météorologiques et climatiques   5.2.3 Relation AOT- $PM_{10}$ avec et sans régimes de temps   5.2.4 Conclusion   12   5.2.4 Conclusion   13   Régimes de temps obtenus sur les données en enlevant le cycle saisonnier   moyen 13	90 99 25 38 39
Ę	5.3	5.2.2.4 Structures spatiales et interprétations météorologiques et climatiques   5.2.3 Relation AOT- $PM_{10}$ avec et sans régimes de temps   5.2.4 Conclusion   5.2.4 Conclusion   12   5.2.3 Relation AOT- $PM_{10}$ avec et sans régimes de temps   5.2.4 Conclusion   13   Régimes de temps obtenus sur les données en enlevant le cycle saisonnier   moyen 15   5.3.1 Caractérisation des régimes de temps 15	90 99 25 38 39 39
Ę	5.3	5.2.2.4 Structures spatiales et interprétations météorologiques et climatiques   5.2.3 Relation AOT- $PM_{10}$ avec et sans régimes de temps   5.2.4 Conclusion   5.2.4 Conclusion   12   5.2.4 Conclusion   13   Régimes de temps obtenus sur les données en enlevant le cycle saisonnier   moyen 15   5.3.1 Caractérisation des régimes de temps 15   5.3.1.1 Matrices de confusion 15	99 99 25 38 39 39 39
Ē	5.3	5.2.2.4 Structures spatiales et interprétations météorologiques et climatiques   5.2.3 Relation AOT- $PM_{10}$ avec et sans régimes de temps   5.2.4 Conclusion   5.2.4 Conclusion   12   5.2.4 Conclusion   13   Régimes de temps obtenus sur les données en enlevant le cycle saisonnier   moyen 15   5.3.1 Caractérisation des régimes de temps 15   5.3.1.1 Matrices de confusion 15   5.3.1.2 Distributions diurnes à inter-annuelles 14	90 99 25 38 39 39 39 39 40

		5.3.1.4 Structures spatiales et interprétations météorologiques et	
		climatiques $\ldots \ldots 15$	2
		5.3.2 Relations AOT- $PM_{10}$ avec et sans régimes de temps	6
		5.3.3 Travail exploratoire sur un système d'alerte opérationnel 17	3
		5.3.4 Conclusion	7
	5.4	Conclusion	8
6	Cor	clusion générale et perspectives 18	0
Bi	ibliog	graphie 18	5
$\mathbf{A}$	Que	elques figures 19	4
	A.1	Figures régimes de temps bruts	4
		A.1.1 régimes de temps bruts	4
		A.1.2 régimes de temps desaisonalisés	6
		A.1.3 Système d'alerte	6
	A.2	Poster et Article	2

# Table des figures

1.1	Les Situations d'alizé maritime, d'alizé maritime continentalisé, d'alizé de	
	Nord-Est et d'harmattan en Anrique de l'Ouest. Le Dorgne (1988) (les	10
1 0	Evolution du jet postume durent une journée des mois de Luip et Ostobre	10
1.2	Evolution du jet nocturne durant une journee des mois de Juin et Octobre,	
	avec les données obtenues pendant la campagne de mésure intensive de 2006	10
1.0	du programme AMMA. (Source de la figure : Lothon et al. [2008])	19
1.3	La circulation atmospherique en saison des pluies (A) et en saison seche (B)	20
	en Afrique de l'Ouest.	20
1.4	Schéma représentant des éléments clés de la mousson africaine pendant l'été	
	boréal, la circulation méridienne-verticale et les vents zonaux dominants	
	(Peyrille, 2006)	22
1.5	Cartes de l'Afrique du nord récapitulant les différentes zones sources de	
	poussières désertiques identifiées par satellite (haut), en fonction de la com-	
	position des aérosols à distance (milieu) et principales zones sources identi-	
	fiées par l'auteur (bas) (source : <i>Scheuvens et al.</i> [2013])	24
1.6	Schéma du cycle des aérosols désertiques (Laurent, 2005)	25
1.7	Mouvement des particules au sol sous l'effet des vents en fonction de leur	
	diamètre. D'après Shao et Lu (2000)	26
1.8	Localisation des principaux processus météorologiques entraînant le soulè-	
	vement d'aérosols. Source : Kocha, 2011(Adapté de Bou Karam et al. [2008]).	27
1.9	Moyenne journalière des concentrations en aérosols désertiques (< $10 \mu m$ à	
	la station de Banizoumbou (Niger); [Marticorena et al., 2010].	30
1.10	Nombre de jours de brumes sèches et précipitations annuelles (mm) me-	
	surées à la station météorologique de Gao, Mali, de 1952 à 1984, $[N^{\prime}T\text{-}$	
	chayi Mbourou et al., 1994]	31
1.11	Schéma de l'impact d'une couche d'aérosols sur les rayonnements solaire et	
	terrestre	32
1.12	Moyennes journalières de $PM_{10}$ (lg / m3) par rapport à des moyens quo-	
	tidiens de visibilité (km) pour chaque station AMMA pendant les saisons	
	(a) humide et (b) sèche pour la période 2006-2008. Banizoumbou et M'Bour	
	sont représentés par des points gris et triangles bleus, respectivement	33
2.1	Localisation des trois stations de mesures installées en Afrique de l'Ouest	
	et vues des sites d'échantillonnage des aérosols à Banizoumbou (Niger),	
	Cinzana (Mali) et M'Bour (Sénégal). (adapté de Marticorena et al.(2010)) .	36
2.2	Représentation de la zone et des pixels (en rouge) pour lesquels on considère	
	les données de température, humidité spécifique et relative, géopotentiel,	
	vent zonal et méridien, hauteur de la couche limite produites par ECMWF.	
	Le point noir au centre représente la position de mesure de chaque site	38
2.3	Moyennes mensuelles des données de l' $AI$ obtenues entre 2006 et 2009	39

3.1 3.2	Carte topologique en $2 - D$ . Le réseau est constitué de deux couches : une couche d'entrée qui sert à la présentation des observations et une couche d'adaptation pour laquelle est défini un système de voisinage Recherche d'une partition adaptée aux classes recherchées. La méthode consiste à regrouper les neurones de la carte par classification ascendante hiérarchique ( $CAH$ ) et à tester les différentes partitions obtenues en fonc- tion des données expertisées	. 47
3.3	Regroupement des neurones de la carte par classification ascendante hiérar- chique : les feuilles de l'arborescence représentent les neurones (6 neurones pour cet exemple); l'axe des ordonnées donne, pour chaque regroupement ou palier agrégatif, l'indice d'agrégation pour la similarité choisie	. 52
4.1	Cycle moyen saisonnier de la concentration en $PM_{10}$ et de l'AOT de 2006	
4.2	à 2010	. 78
4.3	coefficient d'Angström	. 79
	zana); noir (Mbour))	. 79
4.4	Cycle saisonnier de l'AOT et la concentration en $PM_{10}$ pour Alpha $\leq 0.4$	80
$5.1 \\ 5.2$	Organigramme récapitulatif de l'approche développé dans ce travail Représentation des poids Alpha de chaque bloc sur chaque cellule de la	. 84
0.2	carte $2S_SOM$ 1pour les stations de Banizoumbou, Cinzana et Mbour	. 85
5.3	Histogramme des niveaux d'agrégation pour 30 classes (a) et Indices de Davies Bouldin (b) pour les stations de Banizoumbou. Cinzana et Mbour	86
5.4	Représentation des poids Alpha de la carte $2S - SOM2$ de chaque bloc sur chaque cellule de la carte pour les stations de (a : Banizoumbou , b :Cinzana	. 00
EE	et c : Mbour)	. 89
5.5	de Banizoumbou.	. 90
5.6	Cycle saisonnier d'occurrence des types de temps en saison sèche pour les	
5.7	trois stations (a : Banizoumbou; b : Cinzan; c : Mbour) Cycle interannuel d'occurrence des types de temps de chacune des trois	. 94
5.8	stations Persistances et transitions inter-classes	. 95
50	zana; c: Mbour)	. 97
5.9	Caracterisation annuelle des regimes de temps par les mois à travers un plan factoriel )	99
5.10	cartes moyennes des vents $(m/s)$ obtenus et de l'humidité $(g/kg)$ à 1000 hPa sur les stations de (a)Banizoumbou (b) Cinzana et (c) Mbour pour les	
۳ 1 1	six régimes de temps	. 103
5.11	cartes moyennes des vents $(m/s)$ obtenus et la temperature $(C)$ a 1000 $hPa$ sur les stations de (a)Banizoumbou (b) Cinzana et (c) Mbour pour les six	100
5.12	regimes de temps	. 106
<b>.</b>	chaque classe sur les stations de (a) Banizoumbou, (b) Cinzana et (c) Mbour pour les six classes	100
5.13	Cartes moyennes du cycle saisonnier des vents (m/s) et de la vitesse ver-	. 109
	ticale $(hPa/s)$ à 500 $hPa$ (valeurs négatives pour vitesses ascendantes et	
	inversement).	. 110

5.14	Cartes moyennes du cycle saisonnier des vents (vecteurs, m/s) de la vi- tesse verticale $(hPa/s)$ à 850 $hPa$ (couleur bleu pour valeurs négatives des
	vitesses verticales ascendantes et inversement)
5 15	Cartes movennes des vents (m/s vecteurs) et de la vitesse verticale ( $hPa/s$
0.10	couleur) à 500 $hPa$ dans chaque classe (couleur bleu : valeurs négatives
	vitesses ascendantes et coulour rouge : valour positives de vitesse verticale) 114
5 16	Vitesses ascendances et couleur rouge : valeur positives de vitesse verticale). 114
5.10	Cartes moyennes des vents (m/s, vecteurs) et de la vitesse verticale ( $nFa/s$ ,
	couleur) a 850 $nPa$ dans chaque classe (couleur bleu : valeurs negatives,
F 1 F	vitesses ascendantes et couleur rouge : valeur positives de vitesse verticale). 117
5.17	Rose des vents des mesures obtenues a la station d'étude (données in-situ)
	sur les stations de (a) Cinzana, (b) Banizoumbou et (c) Mbour pour les six
	classes
5.18	Profils de vent obtenu avec les classes de la base de données pour chacune
	des trois stations
5.19	Cartes moyennes des vents (m/s, vecteurs) à $925 hPa$ , de Géopotentiel
	(mgb, isolignes) à 925 $hPa$ et l'Aerosol Index (couleur) (A :Banizoumbou;
	$B:Cinzana; C:Mbour)  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  $
5.20	Nuage des points et droite de régression linéaire entre AOT et $PM_{10}$ avant
	la séparation en types de temps
5.21	Boxplot présentant la Médiane (lingne rouge), les quartiles $25\%$ et $75\%$ des
	concentrations en $PM_{10}$ et de l'AOT pour chaque régime de temps et pour
	chaque station
5.22	Nuage des points et droite de régression linéaire entre $PM_{10}$ et AOT pour
	chaque classe et pour les trois station
5.23	Droite de régression entre $PM_{10}$ et AOT pour chaque régime de temps 133
5.24	Boxplot présentant la Médiane (lingne rouge), les quartiles 25% et 75% des
	concentrations en $PM_{10}$ et de l'AOT pour chaque régime de temps et pour
	chaque station après le tri sur le coefficient d'Angstrom
5.25	Nuage des points entre $PM_{10}$ mesuré et $PM_{10}$ estimé pour chaque classe
	sur chaque station (Banizoumbou, Cinzana Mbour)
5.26	Cycle saisonnier d'occurrence des types de temps en saison sèche pour les
	trois stations (Banizoumbou; Cinzan; Mbour)
5.27	Evolution interannuelle d'occurrences des types de temps pour les trois
	stations (Banizoumbou; Cinzana; Mbour)
5.28	Persistance des régimes de temps sur 8 jours pour les trois stations (a :
	Banizoumbou; b : cinzana; c : Mbour)
5.29	Projection sur le premier plan factoriel année par année des régimes de
- 00	temps de chaque station
5.30	Projection de l'ensemble des types de temps des trois stations année par
	année
5.31	Cartes moyennes : (A : la Vitesse verticale $(hPa/s)$ à 500 $hPa$ (couleur),
	de Géopotentiel (mgb) à 500 $hPa$ (isolignes) et de vents (m/s) à 500 $hPa$ ;
	B : de Vents (m/s) et Humidite (g/kg) à 700 $hPa$ ; C : de Vents (m/s) et
F 00	Humidite $(g/kg)$ a 925 $hPa$ )
5.32	Anomalie de la vitesse verticale $(nPa/s)$ à 500 $nPa$ (couleur), de Geopo-
	tentiel (mgb) a 500 $hPa$ (isolignes) et de vents (m/s) a 500 $hPa$ (vecteurs)
	pour les trois stations. (Couleur marron : valeurs negatives de vitesse ver-
5 99	ticale, ascendance; Couleur verte : valeurs positive de vitesse verticale) 157 conte movenne des enemplies de Verte $(m/c)$ et Herri itté $(r/c) \ge 700 LD$
0.33	carte moyenne des anomanes de vents $(m/s)$ et flumidite $(g/g)$ à 700 nPa
594	pour les trois stations (A : Danizounipou; D : Onizana; O : Midour) 100 carte movenne des anomalies de Vents (m/s) et Humidité (m/s) è 025 $LD_{z}$
0.34	carte moyenne des anomanes de vents $(m/s)$ et numidite $(g/g)$ à 925 <i>nPa</i> pour les trois stations (A : Barizoumbou : D : Cincere : C : Macur) 169
	pour les trois stations (A : Danizounibou; D : Unizana; U : Mibour) 103

5.35	Anomalie de l'indice d'aérosol (couleur), de géopotentiel (mgb) à $925 \ hPa$	166
5.36	(isonglies) et de vents pour les trois stations Balizoumbou, Chizana et Moou. Nuages des points et corrélations entre anomalies AOT et $PM_{10}$ sur toute.	.100
0.00	la base avant séparation en régimes de temps	. 167
5.37	Nuages des points et les corrélations entre AOT et $PM_{10}$ dans chaque classe	
	sur les trois stations.	169
5.38	Droite de régression entre les anomalies de $PM_{10}$ et AOT pour chaque	
	régime de temps	170
5.39	Les Nuages de points entre PM estimée et PM mesurée dans chaque classe	
	sur chacune des trois stations (Banizoumbou; Cinzana; Mbour)	172
A.1	Cycle Diurne d'occurrence des types de temps bruts en saison sèche pour	
	les trois stations (A : Banizoumbou; B : Cinzan; C : Mbour)	. 194
A.2	: Diagramme diurne-mensuel des régimes de temps sur les trois stations	
	(Banizoumbou, Cinzana, Mbour)	195
A.3	Cycle Diurne d'occurrence des types de temps bruts en saison sèche pour	
	les trois stations (a : Banizoumbou; b : Cinzan; c : Mbour)	196
A.4	Carte (vent, geopotentiel (iso lignes)) à 1000 hPa et AI (Couleur) pour15%	
	des pm bruts les plus fort sur l'ensemble des trois stations	199
A.5	: Carte (vent, geopotentiel (iso lignes)) à 1000 hPa et AI (Couleur) pour 15%	
	des pm les plus fort sur la base desaisonnalisée pour l'ensemble des trois	<u> </u>
A 6	Weather types identification at M'Bour (Seneral) and implication on the	202
п.0	relationships between $PM_{10}$ concentrations and aerosols optical thickness	
	(AOT) Conférence Internationale AMMA Toulouse. France. du 02 Juillet	
	2012 au 06 Juillet 2012.	203
A.7	Multi-annual variability of mineral dust concentration over West Africa	
	monitored by the AMMA Sahelian Dust Transect Conférence : European	
	Geosciences Union avril 2013 à Vienne/Autriche (EGU 2013)	204

## Liste des tableaux

$3.1 \\ 3.2$	Le tableau de contingence croisant les caractères $X$ et $Y$
5.1	Nombre de poids important de chaque variable de chaque bloc sur chaque
	classe pour les stations de Banizoumbou, Cinzana et Mbour
$5.2 \\ 5.3$	Union des variables dominantes de chaque classe pour chacune des stations. 88 Nombre de variables significatives de chaque paramètre de bloc sur chaque
	classe pour les trois stations
$5.4 \\ 5.5$	pourcentage d'occurrence de chaque type de temps pour les trois stations 91 Pourcentage de confusion des classes de Banizoumbou et Cinzana (a) et des
	classes de Mbour et Cinzana (b) 92
5.6	Probabilité de passer d'une classe à une autre en quatre jours (les valeurs en gras montrent la probabilité qu'un type de temps reste sur lui en quatre
	jours). $\ldots$
5.7	Tableau des pourcentages de variance expliquée des deux premiers axes et
	leur somme
5.8	Significativité des variables in-situ dans chaque classe
5.9	Corrélations entre $PM_{10}$ et AOT par régime de temps et avant la séparation
	en régimes de temps pour les trois stations
5.10	Nombre d'observations avant et après le tri des cas de poussière
5.11	Corrélations entre $PM_{10}$ estimée et $PM_{10}$ mesurée de la base test pour
	chaque classe
5.12	Corrélations linéaire entre $PM_{10}$ estimé $PM_{10}$ mesuré du test et de l'ap-
	prentissage pour les trois stations
5.13	Pourcentage d'occurrence des données de chaque classe et de chaque station. 139
5.14	Pourcentage de confusion des régimes de temps entre la station de Cinzana et Banizoumbou d'une part et de Cinzana et Mbour d'autre part
5.15	Probabilité de passer d'une classe à une autre sur quatre jours pour les trois
	stations (a : Banizoumbou ; b : Cinzana ; c : Mbour)
5.16	Tableau des pourcentages de variance expliquée des deux premiers axes et
	leur somme pour chaque station
5.17	Relation entre les régimes de temps et la NAO dans chacune des trois sta-
	tion. (a : Banizoumbou; b : Cinzana; c : Mbour)
5.18	Corrélations linéaire entre $PM_{10}$ estimé $PM_{10}$ mesuré du test et de l'ap-
	prentissage pour les trois stations
5.19	Les seuils de $PM_{10}$ pour chaque filtrage et pour chacune des trois stations. 173
5.20	Tableau de confusion entre les classes de la base de données brutes corres-
	pondant aux maximums de $PM_{10}$ à 15% et celles des la bases de données
	désaisonnalisées correspondante par rapport au totale de confusion sur l'en-
	semble de la base de données pour les trois stations

5.21	Pourcentage de données des classes de la base de données brutes correspon-	
	dant aux maximums de $PM_{10}$ à 15% contenu dans les classes de la base de	
	données désaisonnalisées correspondantes	. 175
5.22	Tableau de confusion entre les classes de la base de données désaisonnali-	
	sées correspondant aux maximums de $PM_{10}$ à 15% et celles des la bases	
	de données brutes correspondante par rapport au totale de confusion sur	
	l'ensemble de la base de données pour les trois stations	176
5.23	Pourcentage de données des classes de la base de données désaisonnalisées	
	correspondant aux maximums de $PM_{10}$ à 15% contenu dans les classes de	
	la base de données brutes.	. 177

## Introduction

L'Afrique de l'Ouest est la source la plus importante d'aérosols désertiques. Elle contient la zone source la plus étendue, l'ensemble Sahara-Sahel, dont la production annuelle est estimée entre 400 et 1000Mt selon les sources et les méthodes d'estimations [Bach, 1976; Schütz et al., 1981; d'Almeida, 1987; Swap et al., 1992; Huneeus et al., 2011]. Les aérosols désertiques jouent un rôle très important sur l'atmosphère (réduction de visibilité, équilibre du climat), la santé (de nombreuses études ont été menées afin de mieux caractériser l'impact des aérosols désertiques sur l'homme et son environnement), la biodiversité (dégradation des récifs coralliens [Prospero et al., 2002], développement de champignons qui affectent les récoltes par exemple celle de bananes [Griffin et al., 2003]. Dans l'atmosphère, ces aérosols influencent le bilan radiatif terrestre en interagissant avec le rayonnement solaire et tellurique [Sokolik and Toon, 1999]. Les effets radiatifs des aérosols désertiques ont également des conséquences sur la dynamique atmosphérique à l'échelle locale et régionale. Cette interaction est telle qu'elle influence la qualité des prévisions en l'Afrique de l'Ouest [Tompkins et al., 2005]. Elle serait également à l'origine de diminution de l'activité cyclonique observée sur l'Atlantique nord-tropical [Evan et al., 2006 b]. Les aérosols désertiques constituent donc une incertitude importante dans la prévision de l'évolution du climat. Ils sont également soupçonnés de jouer un rôle dans l'occurrence des épidémies de méningites dans la « ceinture des méningites ». Les aérosols désertiques transportés à grande distance peuvent jouer un rôle déterminant sur la disponibilité de certains éléments nutritifs (comme le fer ou le phosphore) pour de larges régions océaniques et pour certains écosystèmes continentaux [Swap et al., 1992; Jickells et al., 2005].

L'évaluation des différents impacts des aérosols désertiques nécessite la quantification des teneurs en aérosols à proximité des zones sources. Une telle quantification impose de tenir compte de la très forte variabilité de ces teneurs atmosphériques induite par les conditions météorologiques qui conditionnent les émissions, le transport et le dépôt de ces aérosols.

La variabilité des contenus en aérosols désertiques au dessus de l'Atlantique Nord tropical a été largement étudiée ces dernières décennies en s'appuyant sur les données de la visibilité horizontale, mais aussi sur les épaisseurs optiques en aérosols dérivées d'observations au sol ou satellite. N'Tchayi Mbourou et al. [1994, 1997] par exemple, ont utilisé des données météorologiques de visibilité horizontale pour évaluer la variabilité spatiale et temporelle des aérosols désertiques en Afrique au nord de l'équateur. Moulin et al. [1997] ont mis en évidence par une analyse sur 11 ans (1982-1997) des épaisseurs optiques en aérosols dérivées de Météosat, une relation entre la variabilité interannuelle du transport d'aérosols désertiques depuis le Sahara et le Sahel vers la Méditerranée et l'Océan Atlantique, et l'indice d'Oscillation Nord-Atlantique.

Compte tenu des différents niveaux de transport de poussière, la visibilité horizontale ou l'épaisseur optique des aérosols ne peuvent pas toujours être de bons indicateurs de mesures de la variabilité des aérosols désertiques. Ainsi dans le cadre du programme AMMA, trois stations de mesures, le « Sahelian Dust Transect », ont été déployées en zone Sahélienne afin de documenter plus finement la variabilité des teneurs atmosphériques en

aérosols désertiques et de comprendre les mécanismes qui la contrôlent.

Ce travail s'inscrit dans le cadre du programme AMMA (Analyse Multiple-disciplinaire de la Mousson Africaine) et a pour objectifs de compléter nos connaissances sur l'étude de la variabilité des aérosols désertiques en Afrique de l'Ouest et de comprendre les conditions météorologiques qui expliquent cette variabilité. Cependant bien que cette thèse a un profil « Géophysique », elle a aussi un profil « Développement de méthodologies statistiques » qui occupe une place importante. Les avancées technologiques en terme de capacité de stockage d'informations, et la multiplication des sources d'informations, contribuent à la mise en place de bases de données complexes et de grande dimension. Dans les domaines tels que la génétique, la finance et les études métrologiques par exemple, on rencontre des données de grande dimension. De plus ces données peuvent présenter une structure en plusieurs blocs de variables caractérisant chacune une vue particulière sur les données recueillies selon une thématique spécifique. Cependant les progrès des outils informatiques ont permis de traiter d'immenses bases de données et de réaliser des représentations graphiques à travers des techniques et méthodes statistiques.

Dans cette thèse, on utilisera plusieurs techniques de la statistique. On procédera à des études exploratoires de données de nature climatologique et à des modélisations statistiques de processus physiques liés aux aérosols désertiques en mettant en place une méthodologie utilisant principalement les outils de la statistique non linéaire basés sur les cartes topologiques auto-organisatrices de Kohonen.

Plusieurs étapes ont été retenues afin d'atteindre nos objectifs :

**Etape 1** : La première étape de ce travail est largement basée sur l'analyse statistique des observations disponibles sur une période de 5 ans pour les trois stations sahéliennes du « Sahelian Dust Transect » (*SDT*). Ce transect, déployé par le *LISA* en 2006 dans le cadre du programme *AMMA*, est constitué des 3 stations, Banizoumbou (Niger), Cinzana (Mali) et M'Bour (Sénégal), fournissant des mesures dédiées au suivi des aérosols désertiques en Afrique de l'Ouest.

**Etape 2** : La deuxième étape du travail a consisté à réaliser une classification des données afin d'obtenir des types de temps en saison sèche sur les trois stations du SDT sur cette période de 5 ans. Cette analyse est menée en utilisant les données météorologiques ERA-interim d'ECMWF. Cette classification permet d'identifier les conditions météorologiques qui prévalent en saison sèche pour ces trois stations et d'identifier les mécanismes météorologiques et climatiques qui les gouvernent.

**Etape 3** : La troisième étape a consisté à analyser la variabilité des concentrations de surface  $(PM_{10})$  et d'épaisseurs optiques en aérosol selon ces types de temps. Il s'agit d'identifier les conditions météorologiques qui conditionnent et modulent le transport d'aérosols désertiques vers l'Atlantique Nord-Tropical et de quantifier leur occurrence. L'analyse statistique des relations entre concentrations de surface, contenus intégrés, et conditions météorologiques doit permettre de proposer une modélisation statistique des teneurs en aérosols mesurées sur les trois stations du SDT.

Après l'introduction, nous décrirons dans le **chapitre 1**, les principales caractéristiques du climat et des aérosols désertiques en Afrique de l'Ouest. On présentera d'abord les éléments du climat nécessaire à la compréhension de notre étude, puis le cycle des aérosols désertiques, les principales zones sources d'émissions et la variabilité temporelle des émissions et du transport des aérosols depuis l'Afrique de l'Ouest. Le chapitre 2 présentera les stations de mesure et les données utilisées dans le cadre de cette étude, et les résultats déjà obtenus avec les données in-situ.

Le chapitre 3 présentera un état de l'art des différents développements statistiques permettant de traiter notre problématique. Nous détaillerons la méthode, non linéaire de Subspace clustering utilisée avec des tests statistiques pour développer une approche de classification avec sélection de variables utilisée dans le chapitre 5.

Le chapitre 4 présentera l'étude de la variabilité des aérosols mesurés au sol au Sahel. Ainsi, nous présenterons la variabilité de la concentration en  $PM_{10}$  à différentes échelles de temps allant du jour à quelques années en fonction des conditions météorologiques locales, puis une analyse comparée des cycles saisonniers de la concentration en  $PM_{10}$  et de l'épaisseur optique des aérosols.

Le chapitre 5 quand a lui, présentera l'analyse climatologique en types de temps en saison sèche sur les trois stations du SDT sur une période de 5 ans. Cette analyse sera menée en utilisant les données météorologiques ERA-interim d'ECMWF. Puis une analyse de la variabilité des concentrations de surface et d'épaisseurs optiques en aérosol selon les types de temps sera développée. Il s'agit d'identifier les conditions météorologiques qui conditionnent et modulent le transport d'aérosols désertiques vers l'Atlantique Nord-Tropical et de quantifier leur occurrence. Dans ce chapitre deux approches de détermination des régimes de temps ont été adoptées. On commencera par montrer les résultats sur les données brutes c'est à dire des données sans aucun filtrage. Puis on montrera les résultats sur les données filtrées par enlèvement du cycle saisonnier moyen.

Enfin, nous conclurons en rappelant les principaux résultats obtenus sur ce travail et nous aborderons les perspectives inspirées par cette étude.

### Chapitre 1

## Généralités sur la circulation atmosphérique et les aérosols désertiques en Afrique de l'ouest.

#### 1.1 Introduction

Le climat de l'Afrique de l'ouest s'inscrit dans celui du climat tropical caractérisé par l'alternance d'une saison sèche qui dure environ huit mois et d'une saison pluvieuse comprise entre juin et septembre, régulée par le déplacement en latitude de la zone de convergence intertropicale. Ainsi la circulation atmosphérique dans les basses couches est contrôlée par le flux de mousson et le flux d'harmattan qui peuvent être intenses d'une saison à l'autre et jouent un rôle très important sur la présence d'aérosols dans l'atmosphère. Dans ce chapitre nous décrivons succinctement le climat de l'Afrique de l'ouest. Nous présentons, notamment, la circulation générale dominante en Afrique au nord de l'équateur en saison sèche comme en saison des pluies et les éléments clés influençant le cycle atmosphérique des poussières désertiques. Puis, nous abordons les aérosols désertiques d'une manière générale, leur cycle de vie, leurs variabilités et leurs impacts.

# 1.2 Généralités sur la circulation atmosphérique en Afrique de l'Ouest

#### 1.2.1 La circulation en saison sèche

En saison sèche, la région Ouest Africaine est sous l'influence des hautes pressions de l'hémisphère Nord, constituées de cellules semi-permanentes, rattachées aux anticyclones des Açores et de Libye [Le Borque, 1988]. L'extension, la forme, l'intensité des cellules qui peuvent être séparées ou soudées, donnent aux alizés qu'elles engendrent une direction, une vitesse et des caractères différents. On distingue ainsi l'alizé maritime, l'alizé maritime continentalisé, l'alizé continental de nord-est et l'alizé d'est ou harmattan (Figure 1.1, Le Borgne [1988] ), d'où une grande variété de temps. Par ailleurs, en Afrique de l'Ouest en saison sèche, la circulation atmosphérique est plus dominée par l'alizé continental de nordest et l'harmattan [Hastenrath, 1988]. L'harmattan règne lorsque les cellules des Açores et de Libye sont soudées. L'alizé est alors de direction zonale, il est très chaud et très sec, à cause de son long trajet désertique continental. L'humidité relative peut tomber à moins de 10% [Le Borqne, 1988], ce qui empêche le développement de formations nuageuses. Le ciel est clair et l'amplitude diurne très forte, avec des températures très élevées le jour, et basses la nuit, dues à un important refroidissement radiatif par absence de vapeur d'eau et de couverture nuageuse. Ainsi, l'harmattan est un vent extrêmement sec, chaud la journée, froid durant la nuit et souvent chargé de poussières [Hamilton et al., 1945].

Chapitre 1. Généralités sur la circulation atmosphérique et les aérosols 18 désertiques en Afrique de l'ouest.



Figure 1.1 – Les Situations d'alizé maritime, d'alizé maritime continentalisé, d'alizé de Nord-Est et d'Harmattan en Afrique de l'Ouest. Le Borgne (1988) (les flèches noires désignent la direction du vent)

#### Description du jet de basses couches

La saison hivernale, en Afrique au nord de l'Equateur, est également marquée par la formation, durant la nuit, des jets de basses couches (Low Level Jets, LLJs) dans la circulation d'harmattan couramment appelé jet nocturne ou Nocturnal Low-Level Jet (NLLJ). Ces jets correspondent à un maximum de vent, qui se forme dans les basses couches, au niveau de la couche d'inversion nocturne [Blackadar, 1957; Holton, 1967; Banta et al., 2006; Lothon et al., 2008] et disparaît pendant la journée. Les LLJs se forment au-dessus du continent Africain tout au long de l'année (surtout sur le Bodelé [Bou Karam et al., 2008) mais leur fréquence d'occurrence est plus importante en hiver qu'en été. Ainsi, 50%des nuits hivernales connaissent la formation de LLJs contre 30% en saison des pluies [Bou Karam et al., 2008]. De plus, la fréquence d'occurrence et l'intensité des LLJs sont accentuées par la présence d'orographie qui influence le flux général d'Harmattan. La Figure 1.2 montre un exemple des variations du profil vertical du jet de basses couches observé au cours d'une journée des mois de Juin et Octobre pendant la période de mesure intensive du programme AMMA en 2006[Lothon et al., 2008]. Au-dessus du Sahara, beaucoup d'auteurs ont montré que les heures d'occurrence du LLJ évoluent tôt le matin (entre 3h et 6h) et son développement s'explique par le mécanisme d'oscillation d'inertie [Blackadar, 1957]. En effet après le coucher du soleil, le développement de la couche d'inversion nocturne par le refroidissement de la surface continentale, conduit à un découplage entre cette surface et les masses d'air qui se trouvent au-dessus de la couche d'inversion et qui ne subissent plus la force de friction exercée par la surface mais seulement le gradient de pression et la force de Coriolis [Hoxit, 1975; Garratt et al., 1992; Mahrt, 1999; Mauritsen and Svensson, 2007]. Ce changement entraine une instabilité qui provoque des oscillations autour du vent géostrophique, formant les jets nocturnes [Blackadar, 1957; Hoxit, 1975; Banta et al., 2003] au sommet de l'inversion. L'intensité des LLJs est d'autant plus élevée que l'inversion de température est importante, ce qui a pour effet de réduire d'avantage les effets de friction. Le cycle diurne des LLJs est en décalage de phase avec celui des vents de surface; les maximums de vents associés aux LLJs sont présents la nuit tandis que les maximums de vents de surface se produisent à la mi-matinée Washington et al. [2006]; Todd et al. [2008]. En Afrique de l'Ouest ces conditions sont favorisées par les reliefs [Todd et al., 2007]. Ainsi, les jets nocturnes ainsi que les forts vents de surface qui y sont associés, pilotent environ 60% du soulèvement d'aérosols désertiques dans le Sahara [Todd et al., 2007; Knippertz, 2008; Bou Karam et al., 2008; Schepanski et al., 2009].



Figure 1.2 – Evolution du jet nocturne durant une journée des mois de Juin et Octobre, avec les données obtenues pendant la campagne de mesure intensive de 2006 du programme AMMA. (Source de la figure : *Lothon et al.* [2008]).

#### 1.2.2 La circulation en saison des pluies (saison de Mousson)

A partir du mois d'Avril jusqu'au mois d'octobre, la région passe progressivement du régime d'alizé continental au régime de mousson (Figure 1.3). Avant l'installation de la mousson, on assiste à une période qui s'étend sur à peu près trois mois (Avril-Mai-Juin) où les pluies sont centrées sur la côte guinéenne. Cette période est marquée vers sa fin par le « saut de mousson », représentant le déplacement, relativement rapide, de la zone de pluie de  $5^{\circ}N$  à  $10^{\circ}N$  en 20 jours environ, et caractérisant ainsi le réel démarrage des pluies au-dessus du Sahel (Onset en anglais). La date climatologique de l'Onset (1968-1990) est le 24 Juin avec 8 jours d'écart type [*Sultan and Janicot*, 2003 b]. L'Onset est alors défini comme étant la transition entre deux régimes quasi permanents : le pré-Onset, où le flux de mousson est présent au-dessus du Sahel sans pour autant que les pluies s'installent durablement, et le régime de mousson, qui lui désigne le début réel de la saison

#### Chapitre 1. Généralités sur la circulation atmosphérique et les aérosols 20 désertiques en Afrique de l'ouest.

des pluies au Sahel [Sultan and Janicot, 2003 b]. Issu de l'anticyclone de Sainte-Hélène, le vent de mousson est frais et humide, convergent, instable et initiateur de précipitations. De direction sud-ouest, il est surmonté, principalement en été, de deux noyaux de vents d'est : le Jet d'Est Africain (JEA) dans les couches moyennes, et dans la haute troposphère le Jet d'Est Tropical (JET) qui a son origine au-dessus du sud-est de l'Asie.

#### 1.2.2.1 Circulation dans les basses couches

- Flux d'Harmattan : L'harmattan est un vent chaud, sec et le plus souvent poussiéreux (alizé continental) d'Afrique de l'Ouest qui souffle vers le sud en provenance du Sahara et du Sahel et affecte le golfe de Guinée en hiver. Entre la fin Novembre et le milieu du mois de Mars, l'harmattan est associé au régime anticyclonique (hautes pressions) habituel de la saison sèche des climats tropicaux. Il en résulte ainsi un ciel absolument clair et une très forte luminosité bien que parfois réduite par les poussières atmosphériques. Lorsque l'harmattan souffle, l'inhibition pluviométrique qui l'accompagne est absolue.

- Flux de mousson : En été, le gradient de température méridien provoque une circulation du sud au nord. Ainsi les alizés de l'hémisphère sud dépassent l'équateur et s'orientent vers le nord-est. Ce vent des basses couches est appelé flux de mousson. Il est plus froid que l'harmattan, plus chargé en humidité par son parcours océanique et son épaisseur dépasse rarement 2km.



Figure 1.3 – La circulation atmosphérique en saison des pluies (A) et en saison sèche (B) en Afrique de l'Ouest.

- Front Inter Tropical (FIT, ITD en anglais) : Le FIT est la ligne de convergence au sol entre l'Harmattan caractérisé par l'air sec du nord-est et la mousson chargée d'air humide du sud-ouest. Le FIT fait partie des paramètres les plus importants dans les prévisions pluviométriques au Sahel. Ainsi, lorsque le FIT passe au nord d'une région, l'épaisseur d'air humide dans cette zone devient plus grande et les pluies plus à très probables.

#### 1.2.2.2 Circulations dans les couches moyennes et dans la haute troposphère

#### 1.2.2.2.a Jet d'Est Africain (JEA) et ondes d'Est

C'est un vent thermique d'Est appelé Jet d'Est Africain (JEA), situé entre  $15^{\circ}W$ et  $15^{\circ}E$ . Il est lié à la différence de température entre l'air chaud du Sahara où siège une importante dépression thermique et l'air frais venant du golfe de Guinée. Ce vent d'Est migre vers le nord en période de mousson. Il souffle en moyenne troposphère entre 700 et 500 hPa (3-5 km) [Burpee, 1972, 1974; Charney and Stern, 1962]. Ce jet joue un rôle important dans le climat sahélien. En saison des pluies, sa structure zonale est perturbée par le développement d'ondes d'Est et de lignes de grains [Kiladis et al., 2006]. Des instabilités se formant au sein de l'AEJ donnent naissance aux ondes d'Est, ondes synoptiques de périodicité 3-5 jours et de longueur d'onde 3000 km Kiladis et al. [2006]; Leroux et al. [2010]. Elles se présentent sous forme d'ondulations méridiennes du champ de vent, se déplaçant d'est en ouest dans la moyenne troposphère entre juin et octobre. Il s'agit de phénomènes importants de la météorologie de l'Afrique de l'Ouest dans la mesure où ils modulent l'activité des systèmes précipitants par l'alternance de conditions favorables ou défavorables au déclenchement des mouvements convectifs qu'ils induisent. Des études réalisées par des auteurs comme [Lafore and Moncrieff, 1989] ou [Weisman and Rotunno, 2004] ont montré que le JEA participe aussi à l'organisation de la convection en orientant le système ou en augmentant le cisaillement dans les basses couches.

#### **1.2.2.2.b** Jet d'Est Tropical (JET).

C'est un vent thermique d'est qui s'étend de l'Asie à l'Atlantique. Il est lié au gradient thermique établi entre les hauts plateaux du Tibet (sud-est asiatique) et l'océan indien. En Afrique de l'Ouest, il est situé entre 5° N et 10°N dans la haute troposphère à un niveau compris entre 100 hPa et 200 hPa (11-15 km). Il s'étend vers l'ouest au-dessus de l'Afrique pendant la période de mousson Charney and Stern [1962]; Burpee [1972, 1974].

#### 1.2.2.2.c Jet d'Ouest Subtropical (JOST)

Les jets subtropicaux de l'hémisphère nord (Figure 1.4) prennent leur origine dans la zone de subsidence de la cellule de Hadley où les vents des hautes couches dévient vers l'Est sous l'influence de la force de Coriolis. Il s'agit des vents d'Ouest, situés autour des latitudes 30° nord. Ces courants de jets subtropicaux sont situés aux environs de 200 hPa (15 km) entre les circulations atmosphériques tropicales et celles de moyenne latitude *Charney and Stern* [1962]; *Burpee* [1972].



Figure 1.4 – Schéma représentant des éléments clés de la mousson africaine pendant l'été boréal, la circulation méridienne-verticale et les vents zonaux dominants (Peyrille, 2006).

#### **1.2.3** Zone de Convergence Inter Tropicale (ZCIT)

En Afrique de l'ouest, le climat est caractérisé par un déplacement de la Zone de Convergence Intertropicale entre 5° et 18° de latitude nord, siège des plus fortes précipitations (zone de forte ascendance sur la Figure 1.4). La position la plus au sud est atteinte en janvier-février et la plus septentrionale est atteinte au mois d'août [Sivakumar, 1988; Sultan and Janicot, 2003 b; Sultan et al., 2007]. Une fois que la position la plus au nord de son déplacement est atteinte, la ZCIT reprend sa descente vers le sud dès la fin du mois d'Août. Cette migration sud-nord de la ZCIT, qui suit le mouvement zénithal du soleil avec un retard d'environ six semaines [Sultan and Janicot, 2003 b], est favorisée par le réchauffement progressif de la zone continentale saharienne d'avril à juillet. La ZCITest alimentée par la convergence d'humidité de basses couches provenant de la convergence entre les alizés de l'Atlantique sud, chargés de vapeur d'eau et dirigés par les hautes pressions de l'anticyclone de Ste Hélène vers le nord du golfe de Guinée, et l'Harmattan au nord. Le FIT est la trace au sol de la discontinuité entre ces deux masses d'air, et se situe une quinzaine de degrés de latitude au nord de la ZCIT, car la convergence maximale d'humidité de la ZCIT ne peut se faire qu'avec une épaisseur suffisante de flux de mousson, donc assez loin au sud du FIT.

#### 1.3 Généralités sur les aérosols désertiques

La présence d'aérosols désertiques en suspension dans l'atmosphère est un fait très marqué en Afrique de l'Ouest. Ainsi au cours des dernières décennies, l'analyse des poussières minérales est devenue un des sujets majeurs traités dans les études environnementales. L'aérosol désertique, encore appelé poussière minérale, fait partie de la famille des espèces présentes dans l'atmosphère sous forme particulaire et appelées aérosols. Les aérosols désertiques sont produits par l'érosion éolienne des sols des régions arides et semi-arides de la planète, d'où l'appellation d'aérosol désertique. Sous l'action du vent, les particules désertiques émises forment des nuages de poussière qui peuvent couvrir des surfaces de plusieurs millions de kilomètres carrés [Schütz et al., 1981]. Du fait de leur durée de vie assez courte et de la localisation géographique des sources, les aérosols désertiques sont distribués de façon très variable selon les régions [Andreae, 1995]. En outre, il est maintenant évident que la présence de poussières minérales dans l'atmosphère a de nombreuses conséquences sur l'environnement et sur l'homme. Ainsi la source de production, le cycle de vie et l'impact des aérosols sur l'environnement seront l'objet de cette partie.

#### 1.3.1 Sources de Production et cycle de vie des aérosols désertiques

#### 1.3.1.1 Sources

Depuis un certain nombre d'années, l'analyse des poussières minérales s'est imposée comme l'un des sujets majeurs traités dans les études environnementales. Ainsi le Nord de l'Afrique, où est localisée la plus vaste zone aride et semi-aride de la planète, le Sahara et le Sahel, est reconnu comme étant la première région productrice d'aérosol désertique [Schütz et al., 1981; d'Almeida, 1987; Laurent et al., 2008; Huneeus et al., 2011]. Les émissions annuelle estimées à partir de mesures de concentration et de dépôt sont de l'ordre de 400 à 700 Mt selon les sources [Bach, 1976; Schütz et al., 1981; d'Almeida, 1987; Swap et al., 1992]. Différentes estimations des émissions d'aérosols désertiques ont également été établies à partir de modèles atmosphériques globaux décrivant le cycle de ces aérosols. Les émissions annuelles estimées sur le Nord de l'Afrique par les différents modèles ayant participé à l'exercice d'intercomparaison dédiés aux aérosols désertiques d'AEROCOM sont en moyenne de 982 Mt avec un écart-type du même ordre de grandeur (751 Mt) et des estimations qui varient de 204 à 2888 Mt [Huneeus et al., 2011]). Ces estimations sont donc aujourd'hui encore entachées d'une forte incertitude. Une étude réalisée par Prospero et al. [2002], à partir d'indicateurs semi quantitatifs de présence d'aérosols absorbants (Aerosol Index, AI) dérivés de l'instrument TOM (Total Ozone Mapping Spectrometer), identifie comme principales sources sur le Nord de l'Afrique le sud-ouest du Sahara, une zone le long de la frontière entre la Mauritanie et Mali, le sud de l'Algérie et une zone située au dessus du désert Libyen et du désert Nubien. Cette étude montre aussi que la dépression de Bodélé (situé au Tchad) est également une des sources les plus actives de poussières désertiques au-dessus du continent Africain. Brooks et Legrand et al. [2000] ont également mis en évidence quatre zones sources au nord de l'Afrique à partir des *IDDI* (Infrared Difference Dust Index), cohérentes avec celles identifiées par Prospero et al. [2002]. En utilisant aussi les AI - TOMS, Engelstaedter et al. [2006] retracent la carte des zones d'émissions de poussières en quatre zones qui est en accord avec celle donnée par les auteurs précédents (Figure 1.5). Plus récemment, Schepanski, Tegen, Laurent, Heinold, and Macke [2007], en utilisant des images infrarouge du capteur SEVIRI pour la période allant de Mars 2006 à février 2007 ont identifié la zone source couvrant l'ouest du Niger, le nord du Mali et le sud de l'Algérie, comme potentiellement la plus active en été. L'ensemble des zones sources identifiées a été récapitulé par Scheuvens et al. [2013] en fonction des méthodes d'identification employées (Figure 1.5).



Chapitre 1. Généralités sur la circulation atmosphérique et les aérosols 24 désertiques en Afrique de l'ouest.

Figure 1.5 – Cartes de l'Afrique du nord récapitulant les différentes zones sources de poussières désertiques identifiées par satellite (haut), en fonction de la composition des aérosols à distance (milieu) et principales zones sources identifiées par l'auteur (bas) (source : *Scheuvens et al.* [2013])

#### 1.3.1.2 Cycle de vie des aérosols

Le cycle de vie des aérosols désertiques se décompose en trois phases : les émissions depuis les zones source, le transport dans l'atmosphère et le dépôt par voie sèche ou humide (voir Figure 1.6). Ces aérosols se déposent en partie dans les zones sources, et peuvent être déposés en dehors, durant leur transport qui peut les emporter à plusieurs milliers de kilomètres des sources.



Figure 1.6 – Schéma du cycle des aérosols désertiques (Laurent, 2005).

#### 1.3.1.2.a Les émissions en zones source

Les aérosols désertiques sont produits par érosion éolienne dans les zones arides et semi arides. Les processus d'émissions sont essentiellement contrôlées par les paramètres météorologiques [Marticorena and Bergametti, 1995; Brooks and Legrand, 2000; Prospero, 1999] et par les caractéristiques de la surface [Marticorena and Bergametti, 1995]. En effet les émissions d'aérosols désertiques par érosion éolienne sont un phénomène à seuil qui ne se produit que lorsque la force de friction du vent exercée sur les grains du sol devient supérieure aux forces qui les maintiennent au sol, c'est à dire la gravité et des forces de cohésion interparticulaires. Quand ce seuil est dépassé, les grains du sol les plus facilement mobilisables (~  $100\mu m$ ) sont alors mis en mouvement horizontalement, suivant des trajectoires balistiques qualifié de mouvement de saltation (Figure 1.6). Au cours de leur chute, les particules impactent la surface du sol, s'entrechoquent et/ou se brisent, ce processus est appelé « sandblasting ». Il permet la libération de particules fines constituant l'aérosol désertique qui peut être mis en suspension dans l'atmosphère et transporté à moyenne et longue distance.

Les particules de taille comprise entre 20 et  $70\mu m$  se déposent à proximité des zones source de production alors que les particules de taille inférieure à  $20\mu m$  peuvent être transportées sur de longues distances (voir Figure 1.7).



Figure 1.7 – Mouvement des particules au sol sous l'effet des vents en fonction de leur diamètre. D'après Shao et Lu (2000).

Les fortes vitesses de vent à l'origine des émissions d'aérosols désertiques en Afrique de l'Ouest, sont causées par différents processus météorologiques et dynamiques, dont certains sont spécifiques au climat Africain [Knippertz and Todd, 2012; Knippertz and Stuut, 2014]. La Figure 1.8 ci-dessous résume les phénomènes météorologiques responsables des émissions de poussières désertiques en Afrique de l'Ouest. - Les jets de basses couches (voir section 1.2.1.1) sont présents tout au long de l'année au dessus du continent avec une fréquence atteignant 50 à 80% des jours en hiver et de 30 à 40% des jours en été [Bou Karam et al., 2008]. Fiedler et al. [2013] suggèrent que la contribution moyenne du jet de basses couches dans l'émission des poussières minérales pendant la saison des pluies est comprise entre 5 et 35% en Afrique de l'Ouest. Cependant en saison sèche le jet nocturne de basses couches se développe plus au sud du Sahara et au nord du Sahel et en saison des pluies, il se développe au sud du Sahel [Fiedler et al., 2013]. Globalement, selon [Schepanski et al., 2009] et [Fiedler et al., 2013], plus de 60% des soulèvements de poussières sont liés au LLJ selon les régions et la saison.



Figure 1.8 – Localisation des principaux processus météorologiques entraînant le soulèvement d'aérosols. Source : Kocha, 2011(Adapté de *Bou Karam et al.* [2008]).

- Courants de densité orographiques : L'orographie entraine l'évaporation de la vapeur d'eau soumise à un mouvement ascendant entrainant le refroidissement par évaporation de l'air ambiant pendant la journée. Cet air se refroidi et devient plus dense que l'air environnant [Droegemeier and Wilhelmson, 1987] et subit un mouvement descendant atteignant la surface sous forme de rafales de vent et se propageant parfois bien loin provoquant ainsi des soulèvements d'aérosols [Charba, 1974; Knippertz et al., 2007]. Les courants de densité issus du refroidissement par évaporation sont limités aux régions montagneuses et se produisent le plus souvent au cours de l'après-midi lorsque la convection est maximale [Bou Karam et al., 2008]. Sur le Sahara, ils peuvent être observés le plus fréquemment près des montagnes de l'Atlas au Maroc en raison du transport de l'air humide en provenance de l'Atlantique [Knippertz et al., 2007].

- Cyclones sahéliens secs : Le cisaillement de vent horizontal au niveau du front intertropical renforcé par la présence des jets nocturnes (induits par le blocage orographique des masses d'air froides provenant de la Méditerranée) entraîne la formation de cyclones au sud des massifs de l'Hoggar et de l'Aïr, au dessus du Sahel [*Bou Karam et al.*, 2009]. Ils se développent pendant la saison de mousson lorsque le *FIT* est au niveau du Sahel. Ces cyclones entrainent de forts soulèvements d'aérosols et favorisent leur mélange sur la verticale [*Bou Karam et al.*, 2008]. *Knippertz and Stuut* [2014] indique que les systèmes météorologiques d'échelle synoptique tropicale entrainant la production de poussière en été au Sahel et au sud du Sahara sont les tourbillons de surface cycloniques associés aux ondes d'Est africaines.

- Convection sèche : Dans la région où on a la dépression thermique, la convection sèche et la forte turbulence entraîne un flux important d'aérosols dans la couche limite Saharienne qui peut aller jusqu'à 6 km [Kocha, 2011; Balme and Greeley, 2006].

#### 1.3.1.2.b Le transport des particules

Une fois que les particules de poussière sont injectées dans l'atmosphère, elles peuvent être transportées à plusieurs milliers de kilomètres de leurs sources et atteindre d'autres continents. Selon d'Almeida [1987], 60% des poussières sahariennes sont transportées vers le golfe de Guinée, 28% vers l'Atlantique et 12% vers l'Europe. Plus récemment Yu et al. [2015] ont estimé le transport transatlantique de poussières sahariennes en utilisant 7 années de mesures du lidar embarqué CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization). Sur la bande latitudinale de  $10^{\circ}S$  à  $30^{\circ}N$ , ils ont estimé qu'annuellement 182 Tg de poussières sont transportées à  $15^{\circ}W$  près de la côte Nord Africaine dont 132 Tg et 43 Tg atteignent respectivement  $35^{\circ}$  W et  $75^{\circ}$  W. Cependant, il existe un cycle saisonnier et une distribution méridionale du transport, contrôlé par le déplacement de la zone de convergence inter tropicale. En hiver et au printemps, le transport de poussière est déplacé vers le Sud, entre  $0^{\circ}et10^{\circ}N$  et affecte significativement l'Amérique du Sud. En été, le transport de poussière se produit principalement sur la bande  $10^{\circ} - 20^{\circ}N$  et affecte la mer des Caraïbes.

#### 1.3.1.2.c Le dépôt des particules

Le cycle de l'aérosol désertique se termine par le dépôt des particules au niveau des continents ou des surfaces océaniques, en conditions d'atmosphère sèche ou humide. Les dépôts secs et humides sont les processus par lesquels les particules de poussière dans l'atmosphère sont éliminées [Duce and Tindale, 1991; Schulz and Lanzerotti, 2012]. La distance parcourue par une particule dépend essentiellement de sa taille, les particules les plus fines peuvent être transportées plus loin que les particules grossières. Le dépôt sec est généralement le processus dominant au voisinage des zones source grâce à la fois à la présence de grosses particules de poussière et le climat sec qui prévaut dans les déserts. Le dépôt humide est généralement le processus dominant pour les particules de poussière loin des régions de source [Bergametti et al., 1989].

Les tempêtes de sable et poussières et les évènement de transport de poussières désertiques peuvent avoir des conséquences sur l'homme et son environnement. Ainsi, il est important de quantifier la variabilité des contenus atmosphériques des aérosols désertiques et de comprendre les mécanismes qui la contrôlent.

#### 1.3.2 Variabilité des contenus atmosphériques en aérosols désertiques

La variabilité des aérosols désertiques en Afrique a fait l'objet d'une attention particulière ces dernières décennies. Ainsi des études ont mis en évidence une forte variabilité temporelle à des échelles allant du jour à l'année mais aussi une forte variabilité spatiale. Ces études s'appuient essentiellement sur différentes mesures indicatives de la présence et de la quantité de poussières présentes dans l'air, comme la visibilité horizontale reportées par les stations météorologiques, ou l'épaisseur optique en aérosol mesurée au sol ou dérivées de mesures satellitales.

#### 1.3.2.1 Variabilité journalière

La variabilité des contenus atmosphériques en aérosols à l'échelle de la journée est difficile à quantifier par manque d'observations à cette échelle de temps. Les séries de données permettant de renseigner les aérosols désertiques sur la période de temps la plus longue sont les observations effectuées dans les stations météorologiques. De telles données sont en effet disponibles depuis 1854 pour certaines stations sahéliennes. La "mesure" la plus courante et la plus facilement accessible renseignant sur la possible présence d'aérosols désertiques est la mesure de la visibilité horizontale dont la réduction en zone désertique est très majoritairement due à la présence de poussières dans l'atmosphère. Cependant, cette mesure, basée sur l'observation de repères visuels, n'est pas très précise et peut également dépendre fortement de l'observateur [*Middleton*, 1986]. Ces données sont donc généralement utilisées pour renseigner la fréquence des évènements de poussières en considérant qu'une visibilité horizontale inférieure à 1 km correspond à une tempête de poussières locale et qu'une visibilité réduite à 5 km est indicatrice d'une brume sèche (c'est-à-dire de la présence d'aérosols dans l'atmosphère mais non nécessairement soulevés localement). Les mesures de réduction de visibilité des stations météorologiques réalisée avec des pas de

29

temps de 3 ou 6h donnent des indications sur le cycle diurne des contenus atmosphériques en aérosols désertiques en terme d'occurrence. N'Tchayi Mbourou et al. [1997] ont effectué une étude statistique fréquences de réductions de visibilité au sol dans 45 stations météorologiques environ réparties en Afrique au nord de l'équateur. Cette étude montre que l'évolution diurne des aérosols désertiques est variable suivant les stations. Dans certaines stations, on note des écarts importants entre les valeurs minimales observées la nuit et les valeurs maximales enregistrées entre 9 heures et 15 heures. Ces résultats sont cohérents avec ceux d'Ozer [2001] qui montrent que la fréquence de réduction de la visibilité associées aux lithométéores, c'est à dire aux nuages de poussières, présente un maximum le jour entre 09 :00 et 15 :00 et un minimum la nuit entre 21 :00 et 03 :00. N'Tchayi Mbourou et al. [1997] estiment que le maximum d'occurrence à midi peut être expliqué comme suit : à la tombée de la nuit, la stabilité atmosphérique causée par le refroidissement de la surface, favorise le dépôt de poussières à la surface et augmente donc la visibilité la nuit ; le matin, l'inversion de température nocturne disparaît et les couches atmosphériques de surface deviennent instables, la turbulence et la vitesse des vents augmentent. Dans l'après midi, l'augmentation de la hauteur de la couche de mélange et la réduction des vents à la surface expliquent la diminution de l'occurrence des aérosols désertiques. Ce cycle diurne (avec un maximum d'émission aux environs de midi et un minimum d'émission la nuit) semble bien établi pendant la saison sèche, mais Bou Karam et al. [2008] suggèrent que pendant la saison humide, des soulèvements importants d'aérosols désertiques ont également lieu au-dessus du Sahel pendant la nuit. Flamant et al. [2007] et Flamant et al. [2009] ont également observé de forts épisodes de soulèvement de jour comme de nuit provoqués par la propagation pendant plusieurs jours de courants de densité issus de systèmes convectifs.

#### 1.3.2.2 Variabilité annuelle et intra saisonnière

La saisonnalité des contenus en aérosols désertiques a été largement étudiée ces dernières années en s'appuyant sur les épaisseurs optiques en aérosols dérivées d'observations satellitaire dans le domaine visible avec une très bonne fiabilité ou même des données de visibilité horizontale. Ce qui a permis de montrer que aérosols désertiques en Afrique de l'Ouest présentent un cycle saisonnier bien marqué. Ainsi N'Tchayi Mbourou et al. [1997] utilisant les données de visibilité horizontale, montrent que la répartition annuelle des aérosols désertiques varie avec la latitude du site étudié. Il suggère que les variations saisonnières et annuelles des aérosols désertiques sont liées à la position du Front Intertropical qui sépare les vents d'harmattan secs fréquemment chargés de poussière et le flux humide de mousson. En janvier-février, lorsque le *FIT* est très bas en latitude, les aérosols désertiques sont présents sur tout le continent au nord de l'équateur. Par contre, lorsque la FITatteint sa position la plus haute en latitude en juillet-août, les aérosols sont concentrés à la périphérie du Sahara. Des études réalisées à partir mesures d'épaisseur optique obtenues sur les stations localisées en zone sahélienne tels que Banizoumbou, Ouagadougou, Agoufou, Dakar ont également montré des cycles saisonniers marqués [Holben et al., 2001; Oquinjobi et al., 2008]. Plus récemment, les données de concentration en  $PM_{10}$  acquises de 2006 à 2008 sur un ensemble de trois stations de mesures, le « Sahelian Dust Transect » déployé en zone Sahélienne dans le cadre du projet AMMA (Analyse Multidisciplinaire de la Mousson Africaine) ont également permis de décrire leur cycle saisonnier à l'échelle régionale (Marticorena et al. [2010] Figure 1.9). Selon Marticorena et al. [2010], les concentrations en aérosols désertiques sont maximales en fin d'hiver et début de printemps, en raison du transport d'aérosols sahariens dans le flux d'Harmattan.





Figure 1.9 – Moyenne journalière des concentrations en aérosols désertiques (<  $10\mu m$  à la station de Banizoumbou (Niger); [Marticorena et al., 2010].

La variabilité des concentrations en saison sèche est principalement contrôlée par celle des émissions depuis les différentes zones sources du Sahara [*Marticorena et al.*, 2010]. En début de saison des pluies, des concentrations journalières extrêmement fortes sont également observées. L'analyse des conditions météorologiques locales montre que ces fortes concentrations sont liées à des émissions locales par les systèmes convectifs de méso-échelle. La diminution des concentrations en saison des pluies semble liée à l'effet conjugué de la diminution des vents de surface dans le front de rafale de ces systèmes, de l'augmentation du couvert végétal et donc des seuils d'érosion et du dépôt des aérosols émis par les précipitations.

#### 1.3.2.3 Variabilité interannuelle

La variabilité des contenus atmosphériques en aérosols désertiques d'une année à l'autre a fait l'objet de beaucoup d'études. Ainsi l'intensité des cycles saisonniers observés en Afrique au nord de l'équateur présente une forte variabilité interannuelle. Les maxima d'épaisseur optique en aérosols observés en 1998-99 et 1999-2000 à Illorin (Nigéria, 8°19'N) présentent une variabilité d'un facteur 2 [*Pandithurai et al.*, 2001]. A plus long terme, des augmentations significatives des contenus en poussière au large de l'Afrique ont été observées au cours des périodes de sécheresse au Sahel (de 1952 à 1984) (*N'Tchayi Mbourou et al.* [1994, 1997] Figure 1.10). De la même façon, les concentrations en aérosols désertiques transportés sur l'Océan Atlantique, observées à La Barbade entre les années 1960 et 1980 ont augmenté d'environ un facteur 4 [*Prospero and Nees*, 1986]. Ces augmentations ont été attribuées à l'émergence de nouvelles sources liées à la diminution observée du couvert végétal [*Tucker et al.*, 1991].



Figure 1.10 – Nombre de jours de brumes sèches et précipitations annuelles (mm) mesurées à la station météorologique de Gao, Mali, de 1952 à 1984, [N'Tchayi Mbourou et al., 1994]

Une forte variabilité d'une année à l'autre du transport de poussières depuis le nord de l'Afrique a également été mise en évidence par *Chiapello and Moulin* [2002]. Ainsi ils ont montré une forte variabilité interannuelle en hiver (entre décembre et mars), avec des années marquées par une faible présence des concentrations de poussières comme l'année 1986 et des années marquées par une forte présence en aérosols tel que l'été de l'année 1989. *Chiapello et al.* [2005] suggèrent que les conditions de sécheresse au Sahel affectent les émissions et le transport des aérosols désertiques en été comme en hiver et d'une année à l'autre car le contenu atmosphérique en aérosols minéraux est lié aux précipitations des années précédentes. Ce qui implique que l'émission des poussières est contrôlé par un changement de la couverture végétale au Sahel. Une analyse à long-terme d'observations satellites et de concentrations de surface mesurées à la Barbade a mis en évidence une augmentation régulière des émissions sahéliennes depuis 1965, attribuée à la dégradation des sols par l'activité anthropique au Sahel [*Moulin and Chiapello*, 2006].

#### 1.4 Impacts des aérosols désertiques

De nombreuses études ont été menées afin de mieux caractériser l'impact des aérosols désertiques sur l'homme et son environnement. Ainsi les aérosols désertiques jouent un rôle très important sur l'atmosphère (équilibre du climat, réduction de visibilité), la santé, la biodiversité.

#### 1.4.1 Impacts sur le climat

Les aérosols désertiques influencent le bilan radiatif terrestre dans le visible, principalement par rétrodiffusion, mais aussi par absorption du rayonnement tellurique (IR)[Sokolik and Toon, 1999]. Ils constituent donc une incertitude importante pour la prévision de l'évolution du climat. Ces effets radiatifs ont des conséquences sur la dynamique atmosphérique à l'échelle locale et régionale. Elle serait aussi à l'origine de l'inhibition de l'activité cyclonique sur l'Atlantique nord-tropical [Evan et al., 2006 b]. Par ailleurs les impacts des aérosols sur le bilan radiatif sont variables selon les régions, impliquant des effets direct et indirect [*Miller and Tegen*, 1998; *Levin et al.*, 1996].

- L'effet direct : les aérosols réfléchissent une partie du rayonnement solaire vers l'espace (« effet parasol »). Ils agissent sur le bilan radiatif par diffusion du rayonnement solaire, provoquant ainsi le refroidissement de la surface. Localement, le forçage radiatif direct peut représenter jusqu'à  $-60W/m^2$  à la surface et plusieurs  $W/m^2$  au sommet de l'atmosphère [Miller and Tegen, 1998].
- L'effet indirect des aérosols désertiques a également fait l'objet de recherches approfondies (voir par exemple Mahowald and Kiehl [2003], Chen et al. [1998], Pruppacher and Klett). Les aérosols peuvent influencer le bilan radiatif en agissant en tant que noyaux de condensation nuageuse et noyaux glaçogène [Hoose et al., 2008].



Figure 1.11 – Schéma de l'impact d'une couche d'aérosols sur les rayonnements solaire et terrestre

#### 1.4.2 La visibilité

Les aérosols désertiques peuvent constituer une nuisance à cause de la gêne occasionnée par les fortes réductions de visibilité horizontale liées aux tempêtes de poussières, notamment en termes de qualité de l'air ou de gestion du transport aérien. Sous l'action du vent, les particules désertiques peuvent être soulevées et forment des nuages de poussière qui couvrent des surfaces de plusieurs millions de kilomètres. Ainsi au cours de violentes tempêtes de sable la visibilité peut être réduite et descendre en dessous de 1000 mètres voir nulle si on est à proximité des zones sources (Figure 1.12). Elle s'améliore à mesure que l'on s'éloigne des zones sources. Le soulèvement des poussières désertiques peut bloquer la circulation routière et aérienne et même paralyser toutes les activités socio-économiques.



Figure 1.12 – Moyennes journalières de  $PM_{10}$  (lg / m3) par rapport à des moyens quotidiens de visibilité (km) pour chaque station AMMA pendant les saisons (a) humide et (b) sèche pour la période 2006-2008. Banizoumbou et M'Bour sont représentés par des points gris et triangles bleus, respectivement

#### 1.4.3 Impacts sur la santé, la biogéochimie et la biodiversité

Les aérosols désertiques ont un impact avéré sur la santé humaine. Leur toxicité peut affecter les voix respiratoires et engendrer de graves conséquences chez les enfants et les personnes âgées. Les aérosols désertiques sont également soupçonnés de jouer un rôle dans l'occurrence des épidémies de méningites au Sahel dans la « ceinture des méningites ». Sultan et al. [2005] ont établi un lien entre les poussières désertiques et les épidémies de méningite au Sahel. En effet, ces épidémies se déclarent en saison sèche quelques semaines après le maximum des teneurs atmosphériques en aérosols désertiques [Marting and Chiapello, 2013; Deroubaix et al., 2013]. Deroubaix et al. [2013] montrent que les plus fortes incidences se produisent toujours à la fin de la saison sèche lorsque la température et les concentrations de poussières sont maximales. Thomson et al. [2006] suggèrent également que des concentrations en poussières anormalement fortes en fin d'automne, pourraient augmenter l'incidence des épidémies de méningites. Les aérosols désertiques sont également impliqués dans les échanges biogéochimiques. A plus long terme, l'érosion puis l'exportation de la fraction argileuse, la plus fertile, des sols peut participer aux processus de désertification dans des zones déjà fragilisées par les conditions climatiques peu favorables aux activités agricoles [Rajot, 2001]. Les aérosols désertiques transportés à grande distance peuvent aussi jouer un rôle déterminant sur la disponibilité de certains éléments nutritifs (comme le fer ou le phosphore) pour de larges régions océaniques où les apports de fer par les aérosols désertiques permettent le développement du phytoplancton [*TALBOT et al.*, 1986; *Jickells et al.*, 2005] et pour certains écosystèmes continentaux, par exemple la forêt amazonienne [*Swap et al.*, 1992].

#### 1.5 Conclusion et hypothèses méthodologiques

Dans cette partie, nous avons fait une description du climat et des poussières en Afrique de l'Ouest. Ainsi nous avons pu voir que les conditions météorologiques à l'échelle régionale et la circulation générale de l'atmosphère participent à chaque étape du cycle des poussières. De ce fait, les émissions et le transport d'aérosols désertiques depuis l'Afrique au Nord de l'Equateur présentent une variabilité temporelle très marquée à l'échelle journalière, saisonnière et interannuelle, bien que le cycle diurne n'a pas été bien documenté à cause du manque de données à cette échelle de temps. Par ailleurs cette variabilité est étudiée en s'appuyant largement sur des données de visibilité horizontale, mais aussi sur les épaisseurs optiques en aérosols. Bien que la variabilité des aérosols et le lien avec la météorologie soit évident, le rôle des situations et conditions météorologiques qui entrainent cette variabilité en saison sèche au Sahel n'est pas clair. L'objectif de ce travail s'inscrit dans celui du programme AMMA qui a permis de déployer des stations de mesures en zone Sahélienne afin de documenter plus finement la variabilité des teneurs atmosphériques en aérosols désertiques. Nous allons d'abord étudier la variabilité des aérosols mesurés au sol à travers une analyse du cycle saisonnier et du cycle diurne. Pour intégrer l'effet des variation de l'altitude de transport de poussière, il est intéressant d'étudier la relation liant les concentrations de surface  $(PM_{10})$ , et les épaisseurs optiques en aérosols (AOT). Cependant cette relation est caractérisée par sa très forte dépendance à la situation géographique et aux conditions météorologiques [Yahi et al., 2013]. La relation ne sera pas la même en saison sèche ou saison des pluies, à Mbour ou à Banizoumbou. Par ailleurs, des difficultés peuvent être identifiées dans cette étude et proviennent à la fois :

- Des mesures qui n'observent pas le même phénomène : mesures au sol pour les  $PM_{10}$ (concentration massique des  $PM_{10}$  entre 2,5 m et 9m du sol), ou mesures d'une grandeur intégrée sur la colonne atmosphérique pour les AOT.
- D'un échantillonnage temporel différent pour chacune des grandeurs. La solution à ces difficultés peut être obtenue à travers une analyse des conditions météorologiques régionales et des mesures d'aérosols désertiques sur le continent ( $PM_{10}$ , AOT) et en utilisant une méthode statistique basée sur une classification des situations météorologiques en types de temps.

Comme toutes les variables du système atmosphérique sont liées, pour obtenir des types de temps il est nécessaire d'employer une méthode s'affranchissant des redondances et permettant de tenir compte de la spécificité d'un type de variable par rapport à l'autre. Le choix des variables atmosphériques utilisées à l'échelle régionale est déterminant pour l'analyse de la synchronisation du climat avec la relation entre AOT et  $PM_{10}$ . La méthode développée par Yahi et al. [2013] a permis d'étudier la relation entre ces deux paramètres sur deux stations. Mais elle ne comportait pas d'analyse préalable pour sélectionner les variables météorologiques les plus pertinentes. Le fait de confirmer cette relation avec une autre méthode et d'autres variables en plus renforcera notre confiance dans l'hypothèse d'un lien entre les concentrations de poussières et conditions météorologiques régionales en saison sèche. L'approche statistique développée pour ce travail et qui constitue le cœur de la thèse permet de mettre en évidence des situations météorologiques typiques pour lesquelles la relation  $PM_{10} - AOT$  pourra être simplifiée. L'objet du chapitre suivant est de présenter les stations de mesures, l'instrumentation et toutes les mesures utilisées dans cette thèse.
# Chapitre 2

# Présentation des données d'étude

# 2.1 Introduction

Afin de compléter les connaissances sur les aérosols désertiques en Afrique de l'Ouest, un ensemble de trois stations de mesures, a été déployé en 2006 en zone Sahélienne par le LISA dans le cadre du programme de recherche international AMMA. Il s'agissait de documenter le plus finement possible la variabilité des teneurs atmosphériques en aérosols désertiques et de comprendre les mécanismes qui la contrôlent. Dans ce chapitre, l'inventaire des différents jeux de données utilisés dans cette thèse sera fait.

# 2.2 Présentation des stations de mesures et des données insitu

# 2.2.1 Présentation des stations de mesures

Les trois stations du "Sahelian Dust Transect" : M'Bour (Sénégal; 14.39°N, 16.96°W), Cinzana (Mali, 13.28°N, 5.93°W) et Banizoumbou (Niger, 13.54°N, 2.66°E) sont localisées en zone sahélienne. La Figure 2.1 présente la localisation des stations et illustre le déploiement instrumental. Les stations sont opérationnelles depuis janvier 2006. Nous avons utilisé pour ces études les mesures acquises de 2006 à 2010. Les stations, les différents dispositifs instrumentaux et les données in situ sont décrits en détail par *Marticorena et al.* [2010] et rappelés ci-dessous.



Figure 2.1 – Localisation des trois stations de mesures installées en Afrique de l'Ouest et vues des sites d'échantillonnage des aérosols à Banizoumbou (Niger), Cinzana (Mali) et M'Bour (Sénégal). (adapté de Marticorena et al.(2010))

**Mbour** est une ville du Sénégal, située sur la Petite Côte, à environ 85 km de Dakar, la capitale. Les instruments de mesures des poussières et des paramètres météorologiques sont installés sur le toit de l'un des bâtiments de l'*IRD* (Institut de Recherche pour le Développement) situé en bordure de mer.

**Banizoumbou** est un village du Niger situé à environ 60 km de la capitale Niamey. A Banizoumbou, la pluie annuelle mesurée est 533 mm pour la période 1991-1998 (Leduc, 2003). Les instruments de mesure sont installés à 2,5 km du village.

**Cinzana**, village du Mali, est situé sur la latitude 13°5N. La moyenne annuelle des précipitations des dix dernières années est de 650 mm. Le site de mesure est installé à la station de Recherche Agronomique, de l'Institut d'Economie Rurale du Mali, situé à 35 km à l'est de la province de Ségou.

# 2.2.2 Mesures in situ

Pour cette étude, nous avons utilisé des données météorologiques à grande échelle et trois types de mesures in situ : les concentrations en particules de taille inférieure à  $10 \mu m$   $(PM_{10})$ , les épaisseurs optiques en aérosols (Aerosol Optical Thickness (AOT), en anglais ) et les données météorologiques locales.

## **2.2.2.1** Mesures de concentrations en $PM_{10}$

Les concentrations de surface en aérosols sont mesurées avec un pas d'échantillonnage de 5 min au moyen d'un système de microbalance (Tapered Element Oscillating Microbalance TEOM1400A, de Thermo Scientific) équipé d'une tête de prélèvement standard permettant une collecte des particules de diamètre  $< 10\mu m$  ( $PM_{10}$ ). Ce diamètre de coupure ne permet pas de couvrir la totalité du spectre granulométrique des aérosols désertiques, mais il permet une mesure homogène sur l'ensemble des sites, comparable aux données de la littérature et aux mesures des réseaux de surveillance de la qualité de l'air. En moyenne journalière, de 2006 à 2008, les concentrations mesurées varient de 10 à 4000  $\mu g.m^{-3}$  [Marticorena et al., 2010]. Elles sont donc parmi les plus fortes que l'on puisse mesurer dans le monde et excèdent très largement les standards européens ou américains de qualité de l'air [Marticorena et al., 2010]. Le pas d'acquisition des mesures (5 min) permet de documenter, en plus des échelles hebdomadaires, saisonnières et pluriannuelles, des échelles de temps fines, notamment le cycle diurne et l'effet du passage des systèmes convectifs de méso-échelles. Cependant on peut rencontrer des données manquantes sur certaines heures de certaines années. La plupart du temps, les données manquantes correspondent à des événements tels que des pannes de matériel, d'électricité ou des pannes de connexions informatiques, les opérations d'entretien épisodique sur les instruments (le remplacement des filtres, le nettoyage, le contrôle, etc.).

# 2.2.2.2 Mesures de l'Epaisseur Optique en Aérosol

Les mesures d'épaisseur optique permettent de déterminer le contenu en aérosols intégré sur l'ensemble de la colonne atmosphérique. Elles sont réalisées par des photomètres solaires du réseau AERONET/PHOTONS, un réseau de photomètres solaires de marque CIMEL, opérationnel depuis les années 90 (http://aeronet.gsfc.nasa.gov) [Holben et al., 1998] et dont la composante française (PHOTONS) est gérée par le LOA (laboratoire d'Optique Atmosphérique, Lille). Les mesures d'éclairements et de luminances réalisées par ces radiomètres passifs permettent d'accéder à l'épaisseur optique en aérosols, c'està-dire au contenu en aérosol intégré sur la colonne atmosphérique. Les mesures étant réalisées du visible au proche infrarouge, la dépendance spectrale de l'épaisseur optique, le coefficient d'Angtröm, donne une indication sur la taille de l'aérosol. Ce paramètre permet de discriminer des particules de tailles contrastées que l'on peut associer à différents types d'aérosols. Notamment, on peut distinguer les aérosols désertiques de taille supermicronique dont le coefficient d'Angström est proche de 0 des aérosols submicroniques produits par la pollution ou les feux de savane dont le coefficient d'Angström est le plus souvent supérieur à 1. Pour cette étude, nous avons utilisé l'épaisseur optique des aérosols à 440 nm et le coefficient d'angström entre 440 et 870nm, c'est à dire la gamme spectrale la plus étendue possible.

# 2.2.2.3 Mesures météorologiques in situ

Les paramètres météorologiques de base (vitesse et direction de vent, température de l'air, humidité relative) sont mesurés et enregistrés avec un pas de temps de 5 minutes. La totalité de l'instrumentation est de marque Campbell Scientific Instruments. La vitesse et la direction de vent sont mesurées par des Windsonic2D, la température et l'humidité relative par des capteurs 50Y ou HMP50 et les précipitations avec des gauges à augets basculants ARG100. L'acquisition des données est faite à l'aide de centrales d'acquisitions CR200.

# 2.3 Données météorologiques à grande échelle

Les données météorologiques de grande échelle utilisées pour ce travail, sont les réanalyses météorologiques ERA - Interim du Centre Météorologique Européen (European Center for Meteorological Weather Forecast). Ces données ré-analysées sont le résultat d'une assimilation entre les observations multi-sources (satellites, sondes, stations en surface et en altitude, avions, bateaux etc.) inégalement reparties dans l'espace et des prévisions de modèles à courte échéance (3 heures) par le modèle météorologique global. Les produits sont fournis selon des grilles superposées en altitude correspondant aux 9 niveaux de pressions (1000 hPa, 925 hPa, 850 hPa, 700 hPa, 500 hPa, 400 hPa, 300 hPa, 250 hPa, 200hPa) et dont chaque maille carrée de 0.5° de coté décrit l'état de l'atmosphère à l'altitude considérée. Pour notre étude, nous avons retenu les variables de température, humidité spécifique et relative, géopotentiel, vent zonal et méridien, hauteur de la couche limite (Boundary Layer Height; *BLH*) qui vont nous servir à la classification en type de temps. Les données ERA-Interim à notre disposition sont produites toutes les 3 heures (0h, 3h, 6h, 9h, 12h, 15, 18h et 21h) pour la période 2006 à 2010. En nous inspirant des travaux de [Yahi et al., 2013], pour chacune des variables, on a extrait 17 points de grille (Figure 2.2) au tour de la station de mesure in-situ sur chacun des 9 niveaux d'altitude qui vont constituer la base de données de chaque station. L'étude porte sur les données de la saison sèche c'est-à-dire d'octobre à mai de chaque année de 2006 à 2010. Ainsi pour chaque station, on a formé une base de donnée de dimension (9727x935). Chaque observation de la base est un vecteur de dimension (935=17x9x6 + 17) soit 17 pixels sur la zone considérée sur les 9 niveaux et pour chaque pixel les six variables météorologiques, plus les 17 pixels de la *BLH* donnée uniquement en surface (1000 hPa). Les paramètres sont renseignés de 0h à 23h par pas de 3h. La Figure 2.2 ci dessous présente la zone d'étude, les stations de mesures sont représentées par des points noirs et les points rouges autour de chaque station représentent les 17 points de grilles utilisés pour chaque station.



Figure 2.2 – Représentation de la zone et des pixels (en rouge) pour lesquels on considère les données de température, humidité spécifique et relative, géopotentiel, vent zonal et méridien, hauteur de la couche limite produites par ECMWF. Le point noir au centre représente la position de mesure de chaque site.

# 2.4 Aerosol Index (AI)

Plusieurs données géophysiques spatiales d'aérosols sont disponibles dans la littérature grâce à la télédétection spatiale. Ainsi on peut distinguer quatre produits qui sont : l'AI (Aerosol Index) issu des mesures de l'instrument OMI (Ozone Monitoring Instrument), l'IDDI (Infrared Difference Dust Index) produits à partir des mesures Météosat dans l'IR thermique, les AOT de la sonde MISR (Multi-angle Imaging Spectro-radiometer) et les AOT "deep-blue" obtenues par l'instrument MODIS (Moderate Imaging Spectro-radiometer). L'IDDI a été récemment upgradé pour produire une série continue de 1981 à 2006. Ce produit ne permet donc pas de couvrir notre période d'étude. Les AOT - MISR sont reconnues comme étant l'estimation la plus fidèle des AOT mesurées au sol, cependant, l'échantillonnage temporel est trop faible pour une utilisation "climatologique" (Banks et al., 2013). Dans ce travail, nous avons utilisé les AI (Aerosol Index) issus de l'instrument OMI qui couvraient notre période d'étude.

L'Aérosol Index (AI) est obtenu à partir des mesures de l'intensité du rayonnement solaire à deux longueurs d'onde dans l'ultra-violet des sondes TOMS (Total Ozone Monitoring System; 1997-2004)) puis OMI (depuis 2005). La mesure de l'AI est basée sur le contraste spectral dans l'UV entre l'intensité du rayonnement solaire effectivement mesurée avec la présence d'aérosols et l'intensité calculée en considérant une atmosphère purement moléculaire (i.e. sans les aérosols) (Deroubaix 2013). L'AI - TOMS a été défini la première fois à partir des longueurs d'onde 340 et 380 nm de TOMS [*Torres et al.*, 1998] puis modifié pour la sonde OMI en utilisant les longueurs d'onde 354 et 388 nm [*Torres et al.*, 1998].

$$AI = -100 \log_{10}(I^{354}obs/I^{354}cal)$$

Avec  $I^{354}$  obs l'intensité du rayonnement solaire observé au sommet de l'atmosphère et  $I^{354}$  cal l'intensité estimée à partir de  $I^{388}$  obs.

L'AI - TOMS est un des produits historiques pour l'étude des poussières désertiques [Chiapello et al., 1999; Prospero et al., 2002; Washington et al., 2003; Engelstaedter and Washington, 2007]. Il a déjà été utilisé pour l'étude de l'impact des aérosols sur la santé [Prospero et al., 2008], et pour l'étude des épidémies de méningites [Molesworth et al., 2003; Thomson et al., 2006]. A partir de l'AI, des études ont montré les tendances décennales du transport de poussières au dessus de l'Atlantique nord [Chiapello et al., 1999, 2005]. Malgré la sensibilité de l'AI à l'altitude de la couche d'aérosols [Torres et al., 1998; Ginoux and Torres, 2003], les capacités de l'AI pour détecter les événements de poussières sahariens ont été démontrées par comparaison avec des mesures de concentration en particules [Chiapello et al., 1999]. Les données que nous avons utilisées dans ce travail couvrent une période allant de 2006 à 2009. La Figure 2.3 présente les moyennes mensuelles des données de 2006 à 20009 à notre disposition. Nous remarquons globalement que le maximum de l'AI est enregistré pendant la saison des pluies et le minimum pendant les mois de la saison sèche. Les moyennes mensuelles des données de l'AI pour notre période d'étude (2006-2009) présentent des caractéristiques similaires avec celles présentées par Engelstaedter and Washington [2007] sur la période de 1980-1992. De la frontière entre le Niger et le Tchad à l'est jusqu'à l'océan à l'ouest, il y a une zone de fort AI. On remarque pour les deux périodes d'acquisition de données (1980-1992 Engelstaedter and Washington [2007] et 2006-2009), que cette zone semble s'élargir à mesure qu'on progresse dans le temps au cours de l'année.



Figure 2.3 – Moyennes mensuelles des données de l'AI obtenues entre 2006 et 2009.

Les différentes bases de données utilisées dans ce travail et les stations de mesures ont été présentées dans cette partie. Plusieurs types de données ont été présentés, chacun ayant une source et un échantillonnage temporel spécifique. Pour la suite de l'étude, plusieurs bases de données ont été constituées. Une base de données qui regroupe  $PM_{10}$  et météorologie locale, une base de données qui regroupe la météorologie à grande échelle, et une autre qui regroupe  $PM_{10}$ , AOT, météorologie locale, météorologie à grande échelle, et une autre qui regroupe uniquement l'AI. Nous analyserons d'abord la variabilité des concentrations d'aérosols à l'échelle saisonnière puis diurne en identifiant les conditions météorologiques locales qui contrôlent cette variabilité. Ensuite avec les données atmosphériques, nous réaliserons une analyse climatologique en types de temps sur les trois stations afin de mieux comprendre la relation  $AOT-PM_{10}$  pour chaque régime de temps. Cependant, vu l'origine multi-sources et l'échantillonnage temporel différent pour chacune des données à notre disposition, il est nécessaire de disposer d'une méthode statistique appropriée capable de prendre en compte ces difficultés. De nouveaux algorithmes ont été développés pour une étude non linéaire mieux adaptée à certaines structures de données. Ces algorithmes ont mis l'accent sur la possibilité de pouvoir traiter des bases de données de plus en plus importantes et complexes. Ils fournissent aussi des sorties graphiques et de tableaux sous une forme qui facilite leur interprétation et leur utilisation par d'autres utilisateurs non spécialistes. On peut citer les cartes topologiques auto-organisatrices de Kohonen qui font partie de la famille des réseaux de neurones à apprentissage non supervisé. Plus récemment des approches de « subspace clustering » (méthodes permettent d'identifier simultanément les clusters et le sous espace spécifique dans lequel chaque cluster est défini) ont fait leur apparition. Ces dernières prennent en compte la taille et la nature multi-blocs des données. Ainsi pour cette étude, nous développons une approche de classification avec sélection de variables basées sur les méthodes de subspace clustering. L'ensemble des approches et méthodes statistiques utilisées dans le travail est présenté dans le chapitre suivant.

# Chapitre 3

# Méthodes et approche statistique utilisées

# 3.1 Introduction

Le but de cette thèse est de développer une approche statistique permettant d'étudier la variabilité des aérosols désertiques en Afrique de l'Ouest et de comprendre les mécanismes météorologiques qui la contrôlent. Ainsi deux grandes parties ont été retenues. La première consiste à analyser le cycle saisonnier et le cycle diurne des aérosols sur la période de 2006 à 2010, des données disponibles pour les trois stations sahéliennes du « Sahelian Dust Transect » (SDT) afin de comprendre les phénomènes météorologiques locaux qui les contrôlent. La deuxième consiste à réaliser une analyse climatologique en types de temps en saison sèche sur les trois stations afin de trouver à l'échelle de l'Afrique de l'Ouest la relation entre la mesure de la concentration massique au sol  $PM_{10}$  et la mesure de l'épaisseur optique des aérosols. Cette relation permettra de pouvoir modéliser et prévoir les concentrations massiques en  $PM_{10}$  à partir des mesures d'épaisseurs optiques. Nous décrirons dans cette partie les différents développements statistiques permettant de réaliser notre étude. Nous présenterons d'abord la méthode de filtrage des données. Un cycle saisonnier très fort est connu dans notre zone d'étude capable de masquer certains phénomènes. Il est alors intéressant de le filtrer afin de pouvoir étudier des phénomènes à plus petite échelle. Nous présenterons aussi l'analyse Factorielle des Correspondances (A.F.C)utilisée dans le chapitre 5 afin de caractériser les types de temps obtenus. Nous détaillerons enfin une méthode, non linéaire de Subspace clustering utilisée avec des tests statistiques pour développer une approche de classification avec sélection de variable utilisée dans le chapitre 5.

# 3.2 Méthode de filtrage de données

Si l'on veut éliminer la composante saisonnière d'un paramètre météorologique représenté sous la forme d'une série temporelle, on peut effectuer une dessaisonalisation. Elle permettra de disposer d'une série temporelle (communément appelée série anomalie) où les signaux de faible fréquence ont été éliminés en grande partie. Dans le cadre de cette thèse, on travaillera avec un type de données dessaisonnalisées, celle dont on a éliminé le cycle saisonnier moyen sur les données brutes.

# 😰 Elimination du cycle saisonnier moyen.

Pour cela, on élimine, pour chaque année, le cycle saisonnier moyen estimé sur les 5 années de données brutes ; il se présente de la sorte : Notons :

 $S^{an}$ , la série temporelle d'une variable S de l'année an,  $CS^{an}$  la composante saisonnière moyenne sur les 5 années de la série S et  $A^{an}$  la série temporelle désaisonnalisée recherchée. Alors on estimera la série temporelle  $A^{an}$  par la formule :

$$\mathcal{A}^{an} = \mathcal{S}^{an} - \mathcal{C}\mathcal{S}^{an} \tag{3.1}$$

# **3.3** Analyse Factorielle des Correspondances

L'Analyse Factorielle des Correspondances (A.F.C) est une méthode factorielle de la Statistique Descriptive Multidimensionnelle.

Son objectif est d'analyser la liaison existant entre deux variables qualitatives (si on dispose de plus de deux variables qualitatives, on aura recours à l'Analyse des Correspondances Multiples. Il est utilisé dans ce travail afin de caractériser les régimes de temps par rapport aux mois de l'année afin de trouver une liaison entre les différents régimes de temps des trois stations. Ainsi, avant de mettre en Œuvre une A.F.C, il faut s'assurer que cette liaison existe bien. Pour cela, il existe des graphiques et des caractéristiques numériques (indice khi-deux et ses dérivés) permettant de mettre en évidence une telle liaison lorsqu'elle existe. On notera qu'on dispose aussi d'un test statistique, le test du khi-deux d'indépendance, basé sur l'indice khi-deux, permettant de tester s'il existe ou non une liaison significative entre deux variables qualitatives.

L'A.F.C est, en fait, une Analyse en Composantes Principales (A.C.P) particulière, réalisée sur les profils associés à la table de contingence croisant les deux variables considérées. Plus précisément, l'A.F.C consiste à réaliser une A.C.P sur les profils-lignes et une autre sur les profils-colonnes. Les résultats graphiques de ces deux analyses sont ensuite superposés pour produire un graphique de type nuage de points, dans lequel sont réunies les modalités des deux variables considérées, ce qui permet d'étudier les correspondances entre ces modalités, autrement dit la liaison entre les deux variables.

L'A.F.C étant une A.C.P particulière, nous n'en donnons que les grandes lignes.

## 3.3.1 Les données

L'AFC s'applique essentiellement à des tableaux de contingence (voir Tableau 3.1). C'est un tableau d'effectifs qui contient à l'intersection de la ligne i et de la colonne j des individus  $z_{ij}$ . Il s'agit de la distribution d'une population totale M selon deux caractères quelconques X en ligne et Y en colonne. Ce sont donc des caractères qualitatifs nominaux et/ou ordinaux. Dans un tableau de contingence, les modalités des caractères sont exclusives les unes par rapport aux autres et exhaustives. Il en résulte que les sommes en ligne et en colonne du tableau ont un sens. Considérons deux variables qualitatives : X à r modalités notées  $(x_1, \ldots, x_l, \ldots, x_r)$ ;Y à c modalités notées  $(y_1, \ldots, y_h, \ldots, y_c)$ . On les observe simultanément sur n individus (ayant ici obligatoirement tous le même poids 1/n). On sait que ces données peuvent être présentées sous la forme d'une table de contingence, ou tableau à double entrée présenté comme suit :

	У1	 $y_h$	 $y_c$	somme
$x_1$	$n_{11}$	 $n_{1h}$	 $n_{1c}$	$n_{1+}$
:	:	:	:	:
$x_l$	$n_{l1}$	 $n_{lh}$	 $n_{lc}$	$n_{l+}$
:	:	••••	:	:
$x_r$	$n_{r1}$	 $n_{rh}$	 $n_{rc}$	$n_{r+}$
sommes	$n_{+1}$	 $n_{+h}$	 $n_{+c}$	n

Tableau 3.1 – Le tableau de contingence croisant les caractères X et Y

# 3.3.2 Le problème et la méthode

On suppose qu'il existe une liaison entre X et Y, et on cherche à décrire, à expliciter, cette liaison. Pour cela, on se base sur l'étude des profils-lignes et des profils-colonnes. La définition du  $l^{ieme}$  profil-ligne est :

$$(\frac{n_{l1}}{n_{l+}}, \dots, \frac{n_{lh}}{n_{l+}}, \dots, \frac{n_{lc}}{n_{l+}})$$

et celle du  $h^{ieme}$  profil-colonne :

$$\left(\frac{n_{1h}}{n_{+h}}, \ldots, \frac{n_{lh}}{n_{+h}}, \ldots, \frac{n_{rh}}{n_{+h}}\right)$$

Rappelons que la liaison entre les deux variables est d'autant plus grande que les profils (lignes ou colonnes) sont différents. C'est alors par l'étude des ressemblances et des différences entre profils-lignes et entre profils-colonnes que l'on peut analyser la liaison entre les deux variables.

La méthode consiste à réaliser l'A.C.P du tableau des profils-lignes (les individus de cette A.C.P. sont les lignes de la table de contingence, c'est à dire les modalités de X) et l'on fait la représentation graphique des individus, donc des modalités de X. On a un seul graphique si on ne conserve que deux dimensions, plusieurs dans le cas contraire. On réalise d'autre part l'A.C.P du tableau des profils-colonnes (les individus de cette A.C.P sont maintenant les colonnes de la table de contingence, c'est à dire les modalités de Y) et l'on fait la représentation graphique des individus, donc des modalités de Y. On montre que ces deux A.C.P se correspondent et qu'il est donc légitime de superposer les deux représentations graphiques. On obtient ainsi un graphique de type nuage de points (ou un ensemble de graphiques si on conserve plus de deux dimensions), représentant à la fois les modalités de X et celles de Y. C'est l'interprétation de ces graphiques, pour lesquels on dispose d'un certain nombre d'indicateurs, qui permet d'expliciter la liaison entre les deux variables considérées. En particulier, on s'attache à étudier les correspondances entre les modalités de X et celles de Y.

# 3.4 Méthodes de Classification

# 3.4.1 Généralités sur les méthodes de classification

Les méthodes de classification non-supervisées (ou clustering) permettent d'explorer des données non-labélisées dans le but de trouver des groupes d'observations homogènes et bien séparés. Les récentes avancées technologiques en termes de capacité de stockage d'informations d'une part, et la multiplication des sources d'informations d'autre part, contribuent à la mise en place de bases de données complexes et de grande dimension. Dans des domaines tels que la génétique, la finance et les études météorologiques par exemple, on rencontre des données de très grande dimension. De plus ces données peuvent présenter une structure en plusieurs blocs de variables caractérisant chacune une vue particulière sur les données recueillies selon une thématique spécifique, on parle de données multi-vues ou multi-blocs. Cependant la majorité des mesures de distance perdent leur pouvoir discriminant au fur et à mesure que la dimension augmente; les observations étant pratiquement tout équidistantes les unes par rapport aux autres [*Parsons et al.*, 2004].

En outre, en l'absence d'une structure globale de corrélation entre les variables (ce qui est souvent le cas en grande dimension à cause de la présence possible de variables souvent distribuées uniformément), la similarité entre deux observations est souvent portée par un nombre limité de variables. Les classes sont donc à rechercher dans des sous-espaces de l'espace initial, on parle alors de méthode de subspace clustering. Le principe des méthodes de subspace clustering reposent sur la recherche de sous espaces de l'espace initial permettant une meilleure détection et interprétation des groupes d'individus [Agrawal et al., 1998].

Ces méthodes peuvent être classées en deux grandes familles : Les méthodes de hard subspace clustering qui déterminent le sous-espace exact permettant de déterminer les classes d'une partition et les méthodes de soft subspace clustering qui affectent des poids aux variables.

### 3.4.1.1 Les méthodes de hard subspace clustering

Elles utilisent en général les méthodes de classification basées sur un maillage de l'espace des observations en définissant pour chaque dimension un histogramme. Puis, les intervalles ayant une densité d'observations supérieure à un seuil fixé a priori définissent des classes pour chaque variable. Les auteurs font ensuite l'hypothèse que si un espace de q dimensions présente une forte densité d'observations alors tout espace composé de q-1 dimensions de cet espace est aussi dense. Ce principe conduit à des classes qui se chevauchent. Cette famille de méthodes comprend entre autres les méthodes *CLIQUE*, *ENCLUS* [Agrawal et al., 1998].

On peut citer aussi les méthodes de Bi-partitionnement ou de Co-clustering qui recherchent une partition des observations et une partition des variables dont les classes caractérisent les observations. Cette famille de méthodes a suscité beaucoup d'intérêt dans les domaines tels que l'analyse de données textuelles et la génétique où l'objectif est de définir des classes de documents (par des classes de mots) ou de gènes. Elles visent à obtenir des blocs individus/variables ou lignes/colonnes les plus homogènes selon des critères métriques ou probabilistes. Il existe plusieurs algorithmes de bi-partitionnement [Charrad and Ahmed, 2011]. On retrouve d'une part, les méthodes de bi-partitionnement simple qui appliquent un algorithme de classification simple sur les lignes et sur les colonnes séparément. Les bi-classes sont construites à partir des classes obtenues sur les lignes et sur les colonnes. Cette famille inclut les méthodes Croeuc, Croki2, Crobin et Cromul présentées par Govaert [1983, 1984] dont la difficulté réside dans le choix du nombre de classes en ligne et en colonne. Nadif and Govaert [1993], présentent un formalisme du problème de la classification croisée sous l'approche modèle de mélange pour mieux appréhender les résultats fournis par ces algorithmes, Jollois [2003] présente ensuite une extension des algorithmes Croeuc et Crobin aux modèles de mélange permettant ainsi de surmonter les difficultés liées au choix du nombre de classes.

D'autre part, les méthodes divisives procèdent par découpage itératif afin d'aboutir à des classes. Cette famille inclut les méthodes one-way splitting et two-way splitting de Hartigan [1975]. Dans le cas des cartes topologiques, Allab et al. [2011] proposent Bi-SOM un algorithme de bi partitionnement basé sur les cartes auto-organisatrices (SOM, Kohonen [1998]), qui traite simultanément les individus et les variables dans une seule carte et qui fournit un outil de visualisation des classes dans une structure topologique. Le principal inconvénient de l'approche Bi - SOM concerne le chevauchement des classes d'individus qui engendre des difficultés supplémentaires d'interprétation des classes d'individus.

## 3.4.1.2 Méthode de soft subspace clustering

La deuxième famille concerne les méthodes de soft subspace clustering qui affectent des poids aux variables, puis les sous espaces sont déterminés par les variables ayant les plus forts poids. Huang et al. [2005] dans W - K - Means puis Jing et al. [2007] dans Entropy weighting K-Means (EWKM) proposent de définir un système de pondération par modification de la fonction de coût associée à l'algorithme des K-Moyennes en y introduisant des poids. Dans la méthode EWKM, les auteurs minimisent simultanément, l'inertie intra-classe et maximise un terme d'entropie négatif dans le processus d'apprentissage. EWKM calcule pour chaque variable des poids inversement proportionnels à leur variance dans chaque classe. Le sous-espace de variables pertinentes pour chaque classe est défini en se basant sur ces poids, facilitant ainsi l'interprétation des classes.

Dans le cas particulier des données initialement structurées en blocs de variables où chaque bloc à sa spécificité, la perturbation apportée par les variables de bruit d'un bloc peut être atténuée par l'information globale apportée par le *bloc* d'où la nécessité de définir des poids sur les blocs initiaux. Cependant, peu de méthodes de subspace clustering prennent en compte la structure initialement multi-blocs des données en classification sans perdre la notion de spécificité des blocs.

Des approches récentes basées sur l'introduction dans la méthode des K – moyennes d'une pondération des variables ou des blocs permettent de prendre en compte en plus de la grande dimension, la structure multi-blocs [*Chen et al.*, 2012]. Plus récemment, ces approches ont été étendues, à travers une nouvelle méthode 2S – *SOM Ouattara* [2014], aux cartes auto-organisées ou self organizing maps(*SOM*) [*Kohonen*, 1998] permettant ainsi d'exploiter les propriétés de visualisation de *SOM*. La méthode 2S – *SOM* repose sur l'introduction d'un système de pondération sur les variables et sur les blocs par modification de la fonction de coût de *SOM*, constituant ainsi une méthode de soft subspace clustering.

Par rapport à notre problématique dans ce travail et la structure de nos données, nous nous somme intéressé à cette nouvelle approche de subspace clustering. Ainsi, nous avons proposé une approche hiérarchique de classification avec sélection de variables fondée sur une double utilisation de 2S - SOM Kaly et al. [2014 b].

# 3.4.2 Présentation de l'Algorithme (2S-SOM)

#### Notation :

Nous disposons de N observations  $z_i$  décrites par p variables divisées en B blocs. On recherche une partition des observations en K classes.

Les notations suivantes seront utilisées :

- $\mathcal{Z}$  la matrice de N observations  $z_i \in \mathbb{R}^p$  avec  $i = 1, \ldots, N$ .
- $\mathcal{V} = \{z^j, j = 1, \dots, p\}$  l'ensemble des variables divisé en *B* blocs de  $p_b$  variables tels que  $p_1 + \ldots + p_b + \ldots + p_B = p$ .
- $\alpha$  est une matrice  $K \times B$  où K désigne le nombre de classes c dans Z,  $\alpha_{cb}$  est le poids du bloc b dans la classe c de Z.
- $\beta = [\beta_1, \ldots, \beta_B]$  est une matrice  $K \times p$  où  $\beta_b$  est une matrice de dimension  $K \times p_b$  définissant les poids  $\beta_{cbj}(j = 1, \ldots, p_b)$  sur les variables du bloc b pour chaque c de Z.

### 3.4.2.1 La méthode des cartes auto-organisatrices de Kohonen (SOM)

Les cartes topologiques ou auto-organisatrices font partie de la famille des méthodes neuronales, à « apprentissage non supervisé ». Cela signifie que, dans une première approche, ces méthodes seront utilisées dans un but descriptif. Les données à analyser sont constituées d'observations dont on cherche à comprendre la structure : il n'y a pas de but précis à atteindre, ni de réponse souhaitée [Kohonen, 1998]. Le but de ces cartes est de représenter des observations multidimensionnelles sur un espace discret de faible dimension (en général 1D ou 2D) qui est communément appelé la carte topologique. Dans ces méthodes, chaque classe est représentée par un neurone qui est caractérisé par un vecteur dit référent. La présentation de l'algorithme de Kohonen permet de comprendre la méthode utilisée dans ce travail.

# L'algorithme de Kohonen

La Figure 3.1 suivante présente une carte topologique en une grille 2D. Un réseau de Kohonen est composé de 2 couches; la première appelée couche d'entrée sert à la présentation des observations et contient exactement n cellule(s), n étant la dimension de l'espace des observations; la deuxième appelé couche d'adaptation est formée du treillis des cellules formant la carte topologique.

Notons que l'algorithme cherche une partition de l'ensemble z en K sous ensembles de sorte que toute cellule c de la carte soit connectée à une cellule (ou observation) j de la couche d'entrée. Cette connexion est quantifiée par  $w_{ij}$  appelé poids synaptique. Chaque sous-ensemble, noté  $i \in C$ , est associé un vecteur dit référent ou représentant  $w_c$  défini dans le même espace que les données de l'ensemble Z. Soit  $W=\{w_c; c=1,\ldots,K\}$  l'ensemble des vecteurs référents. Dans le cas des méthodes de quantification vectorielle, la partition C est souvent définie par une fonction d'affectation  $\chi$  permettant de définir les sous-ensembles c de la partition C tel que  $c=\{z_i \in Z \not/ \chi \ (z_i = c)\}$ . L'ensemble C est constitué d'un ensemble de neurones interconnectés et le lien entre les neurones se fait par l'intermédiaire d'une structure de graphe non-orienté (Figure 3.1). La structure de la carte C est induite par une distance souvent discrète  $\sigma surC$  définie comme étant la longueur du plus court chemin. Pour chaque neurone  $c \in C$ , la distance  $\delta$  permet de définir la notion de voisinage d'ordre d de c:

$$V_c(d) = \{r \in C, \delta(r, c) \le d\}$$

Plus précisément, le lien entre deux neurones r et c de la carte C est introduit par une fonction noyau  $\kappa$  positive et symétrique telle que lim  $\kappa(x) = 0$ . Cette fonction noyau définit une zone d'influence autour de chaque neurone c de la carte :{ $r, \kappa(\delta(r, c)) < \alpha$ } où  $\alpha$  est le seuil d'activation d'un neurone comme faisant partir du voisinage de c. Dans la littérature, il existe plusieurs manières de définir la fonction  $\kappa$ .



Figure 3.1 – Carte topologique en 2 - D. Le réseau est constitué de deux couches : une couche d'entrée qui sert à la présentation des observations et une couche d'adaptation pour laquelle est défini un système de voisinage.

Les fonctions noyaux utilisées dans la littérature se présentent comme suit :

 Les fonctions de voisinage à seuil : les cellules du voisinage ont la même influence, en dehors ils n'en ont aucune :

$$\kappa^T(\delta) = \begin{cases} 1si\delta < T\\ 0sinon \end{cases}$$

 La fonction de voisinage gaussienne, l'influence entre deux neurones dépend de la distance entre ces neurones :

$$\kappa^T(\delta) = exp(-\frac{|\delta^2|}{T^2})$$

A la fin de l'apprentissage, chaque cellule c de la carte topologique est caractérisée par un neurone défini par un ensemble composé du vecteur référent  $w_c$  et du groupement d'observations qu'il représente.

L'algorithme des cartes topologiques cherche à minimiser la fonction de coût généralisée suivante :

$$\mathcal{J}_{SOM}^T(\chi, \mathcal{W}) = \sum_{z_i \in Z}^N \sum_{c \in C} \kappa^T(\delta(c, \chi(z_i)))(||z_i - w_c||^2)$$
(3.2)

La fonction  $\mathcal{J}_{SOM}^T$  est une extension de la fonction de cout des K-moyennes dans laquelle la distance euclidienne entre une observation  $z_i$  et son référent  $w_{\chi(z_i)}$  est remplacée par une distance pondérée  $d_T$  représentant la somme pondérée des distances euclidiennes de  $z_i$  chacun des vecteurs référents  $w_i$  du voisinage d'influence du neurone  $\chi(z)$ 

$$d_T(z_i, w_{\chi(z_i)}) = \sum_{c \in C} \kappa^T(\delta(c, \chi(z_i)))(||z_i - w_c||^2)$$
(3.3)

La décomposition de la fonction objective  $\mathcal{J}_{SOM}^T$  (3.4) dépendant de T permet de mettre son expression sous la forme :

$$\mathcal{J}_{SOM}^{T}(\chi, \mathcal{W}) = (\sum_{c} \sum_{r \neq c} \sum_{\chi(z_i) = r} \kappa^{T}(\delta(c, \chi(z_i)))(||z_i - w_c||^2)) + \kappa^{T}(\delta(c, c)) \sum_{c} \sum_{z_i \in C} ||z_i - w_c||^2$$

$$= \frac{1}{2} \sum_{c} \sum_{r \neq c} \kappa^{T}(\delta(c,c)) \sum_{\chi(z_{i})=r} ||z_{i} - w_{c}||^{2} + \sum_{\chi(z_{i})=r} ||z_{i} - w_{c}||^{2} + \kappa^{T}(\delta(c,c)) \sum_{c} \sum_{r \neq c} I_{c} \quad (3.4)$$

Cette décomposition fait apparaitre deux termes dont il faut minimiser la somme. Le premier terme introduit la contrainte de conservation de la topologie. En effet, si deux neurones c et r sont proches sur la carte, $\kappa^T(\delta(c, r))$  est grand, car  $\delta(c, r)$  est petit; la minimisation de ce terme rapproche les deux sous ensembles c et r liés aux cellules c et r. Les proximités sur la carte entrainent des proximités dans l'espace des données.

Le deuxième terme de l'expression correspond à la fonction C, utilisée par l'algorithme des K – moyennes, pondérée par  $\kappa^T(\delta(c, r)) = \kappa(0)$ . Son importance relative dépend du paramètre T: plus T est petit, plus ce terme est pris en considération durant la minimisation. Ce terme a tendance à faire une partition dont les sous ensembles sont compacts, et pour laquelle les vecteurs référents deviennent les centres de gravité des différents sous ensembles de la partition.

Algorithme : SOM

Entrée : l'ensemble Z des observations.

- 1. Phase d'initialisation : choisir les référents initiaux (en générale de manière aléatoire), la structure et la taille de la carte C; fixer les valeurs de  $T_{max}$ ,  $T_{min}$  et le nombre d'itérations  $N_i$ ;
- 2. Etape itérative t. L'ensemble des référents  $W_{t-1}$  de l'étape t-1 est connu,
  - a) Choisir une observation
  - b) Calculer la nouvelle valeur de la température T en appliquant la formule

$$T = T_{max} \times \left(\frac{T_{min}}{T_{max}}\right)^{\frac{1}{N_i - 1}} \tag{3.5}$$

Pour cette valeur du paramètre, effectuer les deux phases suivantes :

c) Phase d'affectation : on suppose que  $W^{t-1}$  connu; on affecte l'observation  $Z_i$  au neurone  $\chi^T(Z_i)$  défini à partir de la fonction d'affectation :

$$\chi^T(Z_i) = \underset{c \in \mathcal{C}}{\operatorname{argmin}} ||z_i - w_c||^2$$
(3.6)

d) Phase de minimisation : calcul de l'ensemble des nouveaux référents ; les vecteurs référents sont modifiés selon la formule 3.9 en fonction de leur distance au neurone sélectionné à l'étape d'affectation.

$$w_c^t = w_c^{t-1} - \mu^t \kappa^t (\delta(c, \chi(c, z_i)))(z_i - w_c)$$
(3.7)

3. Répéter l'étape itérative en faisant décroitre la valeur de T jusqu'à ce que l'on atteigne  $t = N_i$ .

#### 3.4.2.2 2S-SOM

2S - SOM permet de tenir compte de la dimension élevée des données, de la structure multi-bloc, de la présence de données manquantes ou aberrantes et de faciliter la comparaison des cellules et la visualisation des données. C'est une méthode qui repose sur une modification de la fonction de coût de SOM en introduisant un double système de poids  $\alpha_{cb}(b = 1, \ldots, B)$  et  $\beta_{cbj}(j = 1, \ldots, p_b)$  définis respectivement sur les blocs et sur les variables pour chaque cellule  $c \in C$ . La classification et les poids relatifs à la pertinence des blocs et des variables sont donc obtenus par minimisation de la fonction objective  $\mathcal{J}_{2S-SOM}^T$  définie par la relation suivante (équation 3.8) :

$$\mathcal{J}_{2S-SOM}^{T}(\mathcal{X}, \mathcal{W}, \alpha, \beta) = \sum_{c \in \mathcal{C}} \left( \sum_{b=1}^{B} \left( \sum_{i=1}^{N} \alpha_{cb} \mathcal{K}^{T}(\sigma(\mathcal{X}(z_{i}), c)) d_{\beta_{cb}} + J_{cb} \right) + I_{c} \right)$$
(3.8)

avec  $d_{\beta_{cb}} = \sum_{j=1}^{p_b} \beta_{cbj} (z_{ibj} - \omega_{ckj})^2$  et sous les contraintes :

$$\begin{cases} \sum_{j=1}^{p_b} \beta_{cbj} = 1, \ \beta_{cbj} \in [0,1], \forall \ c \in \mathcal{C}, \forall b \\ \\ \sum_{b=1}^{B} \alpha_{cb} = 1, \ \alpha_{cb} \in [0,1], \ \forall \ c \in \mathcal{C} \end{cases}$$
(3.9)

 $I_c = \lambda \sum_{b=1}^{B} \alpha_{cb} log(\alpha_{cb})$  et  $J_{cb} = \eta \sum_{j=1}^{p_b} \beta_{cbj} log(\beta_{cbj})$  représentent des termes entropies négatives pondérées et associées aux vecteurs poids relatifs aux blocs et aux variables pour une cellule c. La minimisation de la fonction de coût  $J_{2S-SOM}^T$  s'effectue de façon alternée en quatre étapes dont les deux premières phases d'affectation des observations aux classes et d'actualisation des vecteurs référents sont identiques à celles de la méthode SOM. Les valeurs des poids sont supposées connues et fixées à leur valeur courante, on a alors :

— étape 1 : Les référents  $\mathcal{W}$  sont connus et fixés, les observations sont affectées aux cellules en respectant l'équation (3.10) :

$$\mathcal{X}(z_i) = \underset{c \in \mathcal{C}}{\operatorname{argmin}} \left( \sum_{r \in \mathcal{C}} \left( \sum_{b=1}^B \alpha_{cb} \mathcal{K}^T(\sigma(r, c)) d_{\beta_{cb}} \right) \right)$$
(3.10)

— étape 2 : Actualisation des centres de classe à l'aide de : (3.11)

$$\omega_c^T = \frac{\sum_{i=1}^N \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c)) z_i}{\sum_{i=1}^N \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c))}$$
(3.11)

A l'aide du lagrangien de la fonction  $J_{2S-SOM}^T$  relativement aux quantités  $\alpha$  et  $\beta$  on détermine les poids  $\alpha$  et  $\beta$  associés respectivement aux blocs et aux variables. Ainsi, on obtient :

— Etape 3 : les paramètres  $\mathcal{X}$ ,  $\omega$  et  $\beta$  connus et fixés à leurs valeurs courantes alors on a :

$$\alpha_{cb} = \frac{exp(\frac{-\Psi_{cb}}{\lambda})}{\sum_{b=1}^{B} exp(\frac{-\Psi_{cb}}{\lambda})}$$
(3.12)

avec

$$\Psi_{cb} = \sum_{z_i \in r, r \neq c} \mathcal{K}^T(r, c) d_{\beta_{cb}} + \mathcal{K}^T(c, c) \sum_{z_i \in c} d_{\beta_{cb}}$$
(3.13)

— Etape 4 : de manière identique, si les paramètres  $\mathcal{X}$ ,  $\omega$  et  $\alpha$  sont connus et fixés à leurs valeurs courantes alors , on a :

$$\beta_{cbj} = \frac{\exp(\frac{-\Phi_{cbj}}{\eta})}{\sum_{j=1}^{p_b} \exp(\frac{-\Phi_{cbj}}{\eta})}$$
(3.14)

avec

$$\Phi_{cbj} = \sum_{z_i \in r, r \neq c} \alpha_{cb} \mathcal{K}^T(r, c) (z_{ibj} - \omega_{cbj})^2 + \mathcal{K}^T(c, c) \sum_{z_i \in c} \alpha_{cb} (z_{ibj} - \omega_{cbj})^2$$
(3.15)

Le poids d'une variable ou d'un bloc sera donc d'autant plus important qu'il minimise simultanément la somme des écarts entre les référents  $w_c$  et les observations appartenant à la cellule c et aux cellules r du voisinage T de la cellule c. Les coefficients de pondération  $\alpha_{cb}$  et  $\beta_{cbj}$  définis par 2S - SOM indiquent respectivement l'importance relative des blocs et des variables dans les classes. Ainsi, plus le poids d'un bloc b ou d'une variable j est important, plus le bloc ou la variable contribue à la définition de la classe au sens où elle permet de réduire la variabilité des observations dans la cellule et dans son voisinage proche. Finalement, à la convergence, 2S - SOM fournit d'une part une carte topologique permettant de visualiser les données et d'autre part des systèmes de poids pour les classes de la classification.

L'algorithme : 2S - SOM

Entrée : La matrice Z

- 1. Initialisation : choisir la dimension de la carte, le voisinage initial T0 et final Tf des cellules, définir l'ensemble des centres de classe  $W^0$ , initialiser les poids  $\alpha_{cb}^0$  sur les blocs et les poids  $\beta_{cbi}^0$  sur les variables et le couple de paramètres  $\lambda$  et  $\eta$ .
- 2. Affectation : utiliser la formule (3.10) pour affecter chaque observation sa cellule d'appartenance.
- 3. Actualisation des centres de classe : utiliser la formule (3.11) pour actualiser les référents des cellules.
- 4. Actualisation des poids sur les blocs : utiliser les formules (3.12 et 3.13) pour actualiser les poids sur les blocs.
- 5. Actualisation des poids sur les variables : utiliser les formules (3.14 et 3.15) pour actualiser les poids sur les variables.
- 6. Répéter les étapes 2 à 5 jusqu'à la convergence de l'algorithme vers un minimum.

Sortie : Une carte topologique, les poids  $\alpha_{cb}$  sur les blocs et  $\beta_{cbj}$  sur les variables.

En définitive, l'algorithme 2S - SOM qui est une extension de la méthode SOM par ajout des deux étapes d'actualisation des poids sur les blocs et sur les variables hérite de la même complexité  $o(N \times p \times K \times N_A)$  où K est le nombre de cellules de la carte et  $N_A$ le nombre d'apprentissages.

L'initialisation de la taille de la carte est basée sur les heuristiques faisant intervenir le nombre d'observations dans la détermination du nombre de cellules de la carte telles que présentées par Vesanto et al. [2000]. Ceci peut conduire à des cartes de grandes dimensions (pour N grand). Dans ce cas, en vue de faciliter  $PM_{10}$  interprétation des classes, il est possible d'appliquer un algorithme de classification ascendante hiérarchique (CAH) sous contrainte de voisinage sur la matrice composée des vecteurs référents pour réduire ce grand nombre de cellules en un nombre restreint de classes [Gordon, 1996; Yacoub et al., 2001] contenant les cellules de la carte. La contrainte de voisinage dans la CAH permet alors de conserver la topologie des observations fournie par la carte 2S - SOM.

# 3.5 Evaluation des variables pertinentes pour chaque classe de la CAH

La CAH est une méthode de classification appartenant à la famille des méthodes hiérarchiques. C'est une méthode de classification itérative dont le principe consiste en un ensemble de partitions de K (c'est-à-dire des groupements de données de disjoints et d'union égal à K ) de n éléments en classes de moins en moins fines obtenues par regroupements successifs de parties. Il s'agit d'une méthode qui dans cette étude permet à partir de la carte topologique de calculer une hiérarchie de partitions, chaque partition permettant de regrouper d'une manière différente les neurones de la carte (Figure 3.2). Les différentes partitions de la hiérarchie sont déterminées d'une manière itérative, en commençant par la partition la plus fine qui est composée de  $PM_{10}$  ensemble des neurones de la carte. La classification hiérarchique utilise cette partition initiale et procède à des regroupements successifs en fusionnant à chaque itération deux sous-ensembles de neurones. Le choix des deux sous-ensembles qui vont fusionner à une étape donnée est effectué à  $PM_{10}$  aide d'une mesure de similarité définie entre deux sous-ensembles. On choisit, parmi tous les couples de sous-ensembles qui constituent la partition à cette étape, les deux sous-ensembles de neurones les plus semblables, au sens de la mesure de distance choisie. Ces regroupements successifs produisent un arbre binaire de classification appelé dendrogramme (Figure 3.3), dont la racine correspond à la classe regroupant l'ensemble des individus. Ce dendrogramme représente une hiérarchie de partitions. On peut alors choisir une partition en tronquant l'arbre à un niveau donné, le niveau dépendant soit des contraintes de l'utilisateur (l'utilisateur sait combien de classes il veut obtenir), soit de critères plus objectifs.



Figure 3.2 – Recherche d'une partition adaptée aux classes recherchées. La méthode consiste à regrouper les neurones de la carte par classification ascendante hiérarchique (CAH) et à tester les différentes partitions obtenues en fonction des données expertisées.



Figure 3.3 – Regroupement des neurones de la carte par classification ascendante hiérarchique : les feuilles de l'arborescence représentent les neurones (6 neurones pour cet exemple); l'axe des ordonnées donne, pour chaque regroupement ou palier agrégatif, l'indice d'agrégation pour la similarité choisie.

L'algorithme de classification hiérarchique se résume de la manière suivante :

Algorithme de classification hiérarchique

- 1. Phase d'initialisation : On considère au départ la partition formée par les neurones ; chaque neurone est affecté à un sous-ensemble distinct. Choisir le nombre P de groupe que l'on souhaite constituer.
- 2. Pour une partition donnée, trouver les deux sous-ensembles les plus proches au sens du critère de similitude choisi, et les fusionner de manière à former un seul sousensemble.
- 3. Si le nombre de groupement de la partition en cours est inférieur à P, revenir à l'étape (2), sinon fin de l'algorithme.

Différentes mesures de similarité sont proposées dans la littérature [Jain and Dubes, 1988]. La mesure de similitude la plus connue est celle de Ward, qui consiste à opérer des regroupements de façon à ce que la somme des inerties des groupements obtenus reste la plus petite possible : ceci revient à favoriser des groupements les plus compacts possible dans l'espace (euclidien) des données. En retenant le critère de Ward pour effectuer des groupements de neurones de la carte, on se place dans l'espace des observations; le regroupement se fait alors par l'intermédiaire des vecteurs poids  $W_c$ . Mais les neurones appartiennent à la carte qui a une structure topologique discrète définie par le graphe : il est alors possible de favoriser des groupements en tenant compte de cette structure discrète. On sera amené à favoriser des groupements de neurones représentant des régions connexes sur la carte.

# Sélection des variables pertinentes

Au niveau des cellules de la carte, la pertinence d'une variable ou d'un bloc est fournie directement par son poids défini par 2S - SOM.

Les poids  $\alpha_{cb}$  et  $\beta_{cbj}$  sont définis pour un *bloc b* et pour une cellule *c* sous les contraintes  $\sum_{j}^{p_b} \beta_{cbj} = 1$  et  $\sum_{b}^{B} \alpha_{cb} = 1$ . La contribution moyenne d'une variable à une cellule de la carte est donc  $\frac{1}{p_b}$  que l'on peut utiliser comme un seuil de sélection des variables pertinentes.

Une variable telle que  $\beta_{cbj} < \frac{1}{p_b} \forall c \in C$  sera non- pertinente donc éliminée. Dans certains cas, la taille de la carte peut conduire à un grand nombre de cellules. Il est alors possible d'appliquer un algorithme de classification ascendante hiérarchique (CAH) sous contrainte de voisinage sur la matrice composée des vecteurs référents pour réduire ce grand nombre de cellules en un nombre restreint K' de classes contenant  $n_{k'}$ cellules par classe k' [Gordon, 1996].

Ce regroupement des cellules engendre la nécessité de définir le poids et la pertinence d'une variable dans les classes. Nous proposons de prendre la moyenne des poids des cellules constituant la classe, soit  $\gamma_{k'bj} = \frac{1}{n_{k'}} \sum_{l=1}^{n_{k'}} \beta_{c_l^{k'}bj}$  pour la variable j du bloc b et la

classe k' composée des cellules  $c_l^{k'}$ . De même, on définit  $\delta_{k'b} = \frac{1}{n_{k'}} \sum_{l=1}^{n_{k'}} \alpha_{c_l^{k'}b}$  la moyenne des poids du bloc b sur les cellules de la classe k'.

Pour évaluer la pertinence d'une variable dans une classe de la CAH, nous proposons d'utiliser un test statistique basé sur le principe de la valeur-test proposée par *Lebart* et al. [1997]. On désigne par  $\gamma_{bj} = \frac{1}{N_{cell}} \sum_{c \in \mathcal{C}} \beta_{cbj}, s_{bj}^2, s_{k'bj}^2 = \frac{N_{cell} - n_{k'}}{N_{cell} - 1} \frac{s_{bj}^2}{n_{k'}}$  respectivement la moyenne, la variance des poids de la variable j du bloc b pour l'ensemble des cellules et la variance des poids de la variable j du bloc k' de la CAH. La valeur-test, pour les poids d'une variable j dans une classe k', se définit alors par :

$$t_{k'bj} = \frac{\gamma_{k'bj} - \gamma_{bj}}{s_{k'bj}}$$

La valeur test  $t_{k'bj}$  peut se lire comme la statistique d'un test de comparaison de moyennes où, sous l'hypothèse nulle de tirage au hasard de  $n_{k'}$  cellules parmi  $N_{cell}$ , elle suivrait de manière asymptotique la loi normale centrée réduite. Pour les niveaux de risque usuels (5%), on considérera donc que la différence entre le poids moyen dans la classe et le poids moyen sur la carte est significative lorsque la valeur absolue de la valeur test est supérieure à 2.

Ainsi, au niveau de la valeur test, la pertinence d'une variable ou d'un bloc dans une classe de la CAH est relative à la distribution des poids de 2S - SOM pour l'ensemble des cellules de la carte.

Pour une classe la contribution moyenne des variables est :

$$\frac{1}{p_b} \sum_{j=1}^{p_b} \gamma_{k'bj} = \frac{1}{p_b} \sum_{j=1}^{p_b} \frac{1}{n_{k'}} \sum_{l=1}^{n_{k'}} \beta_{c_l^{k'}bj} = \frac{1}{p_b}$$

Par ailleurs, compte tenu des propriétés de 2S - SOM [*Ouattara*, 2014], les variables de bruit qui ont en général des poids  $\gamma_{bj} < \frac{1}{p_b}$  peuvent également être sélectionnées par la valeur test. Ainsi, parmi les variables sélectionnées par la valeur test, seront éliminées celles dont le poids moyen  $\gamma_{bj} < \frac{1}{p_b}$ .

Il est alors possible de sélectionner l'ensemble des variables pertinentes pour les blocs à travers une première application de la méthode 2S - SOM.

L'approche hiérarchique que nous proposons ici, consiste ensuite, à appliquer de nouveau 2S - SOM sur les variables sélectionnées pour obtenir la partition recherchée. La réduction du nombre de variables permet de simplifier d'une part l'interprétation des classes obtenues, d'autre part elle fournit des partitions de meilleure qualité en terme d'indices de pureté et de NMI (présentés dans la partie 3.6.1) illustré sur des données labellisées sur un jeu de données réelles issus de l'UCI et sur un jeu de données simulées). Cet approche décrite ci dessus a fait l'objet d'un article (cf Kaly et al. [2014 b]).

# 3.6 Les critères d'évaluation d'une classification

L'évaluation de la qualité d'une classification est un aspect très important pour valider les classes obtenues. Plusieurs critères d'évaluation des classes sont présentés dans la littérature [Halkidi et al., 2002; Jain, 2008] et peuvent être regroupés en 2 grandes familles. La famille des critères non-supervisés ou internes qui utilise uniquement les informations internes aux données telles que la distance entre les observations. Puis la famille des critères externes qui s'utilise généralement pour la validation d'un algorithme. Une présentation plus complète de cette partie est dans la thèse de *Ouattara* [2014].

# 3.6.1 Critères d'évaluation interne

Les critères d'évaluation interne des classes sont basés sur la définition de mesures propres aux classes comme la distance entre les observations et leur centre de classe. Ils sont basés sur les propriétés voulant que :

- des individus d'une même classe partagent les mêmes propriétés (compacité).
- des individus appartenant à des classes différentes aient peu de propriétés en commun (séparabilité).

Pour évaluer le respect de ces deux notions, différentes mesures basées sur les distances entre les observations  $z_i$  et les centres de classe  $w_k$  ont été définies pour quantifier l'adéquation entre une partition et l'idée que l'on se fait d'une bonne classification. Cette section présente plus en détail les indices d'évaluation de la pertinence d'une classification.

 Somme des carrées des erreurs : La somme des carrées des erreurs (Mean Square Error, MSE) permet d'évaluer la compacité des classes d'une classification. Elle vaut :

$$MSE = \frac{1}{N} \sum_{k=1}^{K} \sum_{z_i \in c_k} ||z_i - w_k||^2$$
(3.16)

où K est le nombre de classes.

— Indice de Davies-Bouldin : L'indice de Davies and Bouldin [1979] évalue la qualité d'une classification en mesurant la compacité et la séparabilité des classes travers le calcul de la moyenne de la similarité entre les classes :

$$DB(C) = \frac{1}{K} \sum_{l=1}^{K} \max_{k=1,\dots,K,k \neq l} \left( \frac{S_{db}(ck) + S_{db}(cl)}{d(w_k, w_l)} \right)$$
(3.17)

où  $d(w_k, w_l)$  est la distance entre les centres des classes  $c_k$ ,  $c_l$  et la moyenne des distances entre les observations de  $c_k$  et les centroïdes  $w_k$  de chaque classe. Pour des groupes compacts, la moyenne  $S_{db}(ck)$  de la distance au référent vecteur

valeur faible de cet indice implique une classification de bonne qualité en termes de compacité et de séparabilité.

### Indices spécifiques aux cartes topologiques

Les cartes auto-organisées font partie des méthodes de quantification vectorielle qui ont des propriétés spécifiques, il semble donc naturel de les évaluer à l'aide de l'erreur de quantification moyenne que l'on définit ainsi :

— Erreur de quantification :

$$mqe = \frac{1}{N} \sum_{i=1}^{N} ||z_i - w_c i||^2$$
(3.18)

où  $c_i$  est l'indice du prototype le plus proche de  $z_i$ 

Le principe de conservation de la topologie des observations sur la carte implique d'évaluer la qualité de la topologie fournie par SOM. Le taux d'erreur topologique permet de quantifier la conservation de la topologie locale de l'espace des observations par la carte.

$$Tge = 1 - \frac{1}{N} \sum_{i=1}^{N} 1/N(c_i) \left( \underset{c \neq \chi(i)}{armin}(||z - w_c||^2) \right)$$
(3.19)

où  $1/N(c_i)$  est la fonction indicatrice de l'ensemble des voisins du prototype le plus proche de l'observation  $z_i$ .

La mesure de quantification vectorielle et l'erreur topologique peuvent être contradictoires puisqu'elles évaluent des propriétés différentes de la carte topologique. La mesure de distorsion présentée ci-dessous crée un compromis entre ces deux mesures. Mesure de distorsion : La mesure de distorsion prend en compte l'erreur de quantification vectorielle et la conservation de la topologie locale à travers l'introduction d'une pondération basée sur la fonction de voisinage définie dans SOM. Elle vaut l'erreur quadratique pondérée par la fonction de voisinage.

$$distorsion = \sum_{i=1}^{N} \sum_{c} K^{T}(c_{i}, c) ||z - w_{c}||^{2}$$
(3.20)

où  $K^T(c_i, c)$  est la fonction de voisinage.

# 3.6.2 Critères d'évaluation externe

La comparaison de deux partitions C et C' d'un même ensemble de données Z peut être réalisée à partir du tableau 3.2 de contingence $\tau = (n_{kl})$ , où  $(n_{kl})$  désigne le nombre d'objets appartenant simultanément aux classes k et l des partitions respectifs C et C'.

Tableau 3.2 – Table de contingence entre deux partitions C et C' contenant respectivement K et K' classes;  $(n_{kl})$  est l'effectif d'observations appartenant simultanément à la classe k de C et la classe l de C'.

	1	 l	 K'	$n_k$
1	$n_{11}$	 $n_{1l}$	 $n_{1K'}$	$\sum_{l} n_{1l}$
:	•	•	•	
k	$n_{k1}$	 $n_{kl}$	 $n_{kK'}$	$\sum_{l} n_{kl}$
:	•	•	:	:
K	$n_{K1}$	 $n_{Kl}$	 $n_{KK'}$	$\sum_{l} n_{Kl}$
$\overline{n}_{.l}$	$\sum_k n_{k1}$	 $\sum_k n_{kl}$	 $\sum_k n_{kK'}$	

Pour introduire les critères externes on définit, à partir du tableau de contingence, les Quantités  $N_{11}$ ,  $N_{10}$ ,  $N_{01}$  et  $N_{00}$ :

—  $N_{11}$ , le nombre de fois où deux observations sont dans la même classe dans C et dans C'(accords positifs)

$$N_{11} = \frac{1}{2} \left( \sum_{k=1}^{K} \sum_{l=1}^{K'} n_{kl} (n_{kl} - 1) \right)$$
(3.21)

 $- N_{10}$ , le nombre de fois où deux observations sont dans la même classe de C' et dans des classes différentes dans C.

$$N_{10} = \frac{1}{2} \left( \sum_{k=1}^{K'} n_{k.}^2 - \sum_{k=1}^{K} \sum_{l=1}^{K'} n_{kl}^2 \right)$$
(3.22)

-  $N_{01},$  le nombre de fois où deux observations sont dans la même classe de et des classes différentes

$$N_{10} = \frac{1}{2} \left( \sum_{k=1}^{K'} n_{.l}^2 - \sum_{k=1}^{K} \sum_{l=1}^{K'} n_{kl}^2 \right)$$
(3.23)

—  $N_{00}$ , le nombre de fois où deux observations sont dans des classes différentes de C et de C' (accords négatifs).

$$N_{00} = \frac{1}{2} \left( n^2 + \sum_{k=1}^{K} \sum_{l=1}^{K'} n_{kl}^2 - \left( \sum_{k=1}^{K} n_{k.}^2 + \sum_{k=1}^{K'} n_{.l}^2 \right) \right)$$
(3.24)

Les indices de précision et le coefficient de rappel sont des mesures asymétriques évaluant la similarité entre une partition C' fournit par un algorithme et les labels de référence C.

— Indice de précision : L'indice de précision indique la probabilité que deux objets soient regroupés dans la partition C' s'ils le sont dans la partition C :

$$prec(C, C') = \frac{N_{11}}{N_{11} + N_{01}}$$
(3.25)

— Indice de rappel : Le coefficient de rappel indique la probabilité que deux objets soient regroupés dans la partition C s'ils le sont dans la partition C' :

$$rapp(C, C') = \frac{N_{11}}{N_{11} + N_{10}}$$
(3.26)

Les critères de rappel et de précision prennent leurs valeurs dans l'intervalle [0;1], cependant une valeur proche de 1 n'implique pas que les partitions soient identiques. La combinaison de ces deux indices en prenant leur moyenne arithmétique, géométrique et harmonique, fournit respectivement le deuxième coefficient de Kulczynski, l'indice de Folkes et Mallows et la F-mesure. Ces trois critères sont symétriques et prennent leurs valeurs sur l'intervalle [0,1]. Ils valent 1 si et seulement si les partitions et sont identiques.

F-mesure : La F-mesure est la moyenne harmonique de l'indice de précision et du coefficient de rappel :

$$F(C, C') = \frac{prec(C, C') \times rapp(C, C')}{prec(C, C') + rapp(C, C')}$$
(3.27)

En utilisant une moyenne harmonique pondérée, on définit la  $F_{\alpha}$ -mesure comme suit :

$$F(C, C') = \frac{(1 - \alpha) \times prec(C, C') \times rapp(C, C')}{\alpha \times prec(C, C') + rapp(C, C')}$$
(3.28)

où  $\alpha$  est un coefficient de pondération strictement positif. Notons que N pour  $\alpha > 1$  cet indice est négatif.

— La pureté d'une partition : La pureté d'une partition s'évalue en quantifiant la cohérence d'une partition par rapport une autre. La manière la plus simple d'évaluer la pureté est de rechercher le label majoritaire de chaque classe et de sommer le nombre d'observations ayant le label majoritaire par classe. La pureté se définit alors simplement par l'expression suivante :

$$Pur(C, C') = \frac{1}{N} \sum_{k=1}^{K} argmax(n_{kl})$$
(3.29)

Cette valeur de la pureté est équivalente à l'estimation du pourcentage d'individus ayant le label majoritaire dans les classes de la partition C. Sa valeur est bornée dans [0,1]; 1 implique que les individus formant les classes ont des labels identiques. Une formulation probabiliste de la pureté d'une partition consiste à calculer la probabilité qu'étant donnée une classe de la partition C, deux individus tirés au hasard sans remise aient le même label. En définissant par  $\frac{n_{kl}}{n_k}$  la probabilité que le premier individu ait le label l et par  $(\frac{n_{kl}}{n_k})^2$  la probabilité que le deuxième appartienne la même classe. Ainsi on évalue la pureté d'une classe par :

$$Pur_{prob}(c_k) = \sum_{l=1}^{K'} (\frac{n_{kl}}{n_k})$$
(3.30)

Ce qui donne pour une partition :

$$Pur_{prob}(c) = \frac{1}{N} n_k Pur_{prob}(c_k)^2$$
(3.31)

Cette nouvelle mesure de pureté prend en compte la proportion des différents labels dans les classes en favorisant les classes ayant un nombre limité de labels.

— Indice de Rand : L'indice de Rand indique la proportion de paires d'observations pour lesquelles deux partitions sont en accord. Il correspond à la mesure de similarité binaire simple correspondance qui prend ses valeurs dans l'intervalle [0, 1] et est défini de la manière suivante :

$$R(C, C') = \frac{N_{11} + N_{00}}{N_{11} + N_{01} + N_{10} + N_{00}}$$
(3.32)

Cependant, pour deux partitions définies aléatoirement la valeur de l'indice de Rand n'est pas nulle, d'autre part lorsque ces deux partitions ont des nombres de classes différents, l'indice de Rand peut être proche de 1. Hubert and Arabie [1985] et Chavent et al. [2001] proposent plusieurs variantes de l'indice initial de Rand pour surmonter ces problèmes. Le plus connu, l'indice de Rand ajusté  $R_a$  prend la valeur 0 lorsque les deux partitions sont définies aléatoirement et la valeur 1 lorsqu'elles sont identiques. L'indice de Rand est souvent plus élevé que sa version corrigée qui se définit comme suit.

$$R_a(C, C') = \frac{n^2 \sum_{ij} n_{ij}^2 - \sum_i n_{i.}^2 \sum_j n_{.j}^2}{\frac{1}{2} n^2 (\sum_i n_{i.}^2 + \sum_j n_{.j}^2) - \sum_i n_{i.}^2 \sum_j n_{.j}^2}$$
(3.33)

— Entropie d'une partition : L'entropie associée à une partition  $C = \{c_1, \ldots, c_k\}$ mesure l'incertitude de la variable aléatoire C dont la valeur est l'indice de la classe d'un objet prélevé aléatoirement dans l'ensemble  $\{1, \ldots, K\}$ . Elle est définie par :

$$H(C) = -\sum_{k=1}^{K} P(k) log(p(k)) = -\sum_{k=1}^{K} \frac{n_{k.}}{N} log(\frac{n_{k.}}{N})$$
(3.34)

L'entropie d'une partition est toujours positive et prend la valeur 0 lorsqu'il n'y a aucune incertitude quant à l'appartenance d'un objet à une classe. Autrement dit, lorsque K = N.

— Information mutuelle entre deux partitions :

$$I(C, C') = \sum_{i=1}^{K} \sum_{i'=1}^{K'} P(k, l) log\left(\frac{P(k, l)}{P(k)P(l)}\right) = \sum_{i=1}^{K} \sum_{i'=1}^{K'} \frac{n_{kl}}{N} log(\frac{n_{kl}N}{n_{k.}n_{.l}}$$
(3.35)

L'information mutuelle I est symétrique et positive, ainsi lorsque deux partitions C et C' sont égales, on a : I(C, C') = H(C) = H(C') L'information mutuelle I entre deux partitions C et C' quantifie l'information apportée par la variable aléatoire v associée à C sur la variable v' associée C' et réciproquement. Il est aussi possible de définir la version normalisée de l'information mutuelle NMI. L'indice NMI est indépendant du nombre de classes :

$$NMI(C, C') = \frac{I(C, C')}{\sqrt{H(C)H(C')}}$$
(3.36)

# Chapitre 4

# Variabilité temporelle des aérosols mesurés au sol au Sahel

# 4.1 Introduction

La première étape de notre travail est largement basée sur l'analyse statistique des observations disponibles pour les trois stations sahéliennes du « Sahelian Dust Transect » (SDT). Ainsi nous avons réalisé une analyse statistique des observations disponibles de 2006 à 2010 pour les trois stations sahéliennes du SDT. Dans ce chapitre, nous présentons d'abord un article publié dans Atmospheric Research, intitulé « Variability of mineral dust concentrations over West Africa monitored by the Sahelian Dust Transect », et ensuite une analyse comparée des cycles saisonniers de la concentration en  $PM_{10}$  et de l'épaisseur optique en aérosols.

# 4.2 Variabilité des concentrations en poussières minérales

L'article présenté ci dessus analyse la variabilité saisonnière et diurne des concentrations de poussières sur la période 2006-2010 et sur les trois stations du Sahelian Dust Transect : M'Bour (Sénégal ; 14.39° N, 16.96° W), Cinzana (Mali, 13.28°N, 5.93°W) et Banizoumbou (Niger,  $13.54^{\circ}N$ ,  $2.66^{\circ}E$ ). L'analyse des concentrations de poussières mesurées entre 2006 et 2010 a permis de confirmer un cycle saisonnier régional mis en évidence sur 3 années de mesures par Marticorena et al. [2010]. Ce cycle saisonnier est caractérisé par un maximum de concentration en saison sèche et un minimum pendant la saison des pluies sur 'ensemble de trois stations. La mise en relation entre les cycles saisonnier de la concentration en  $PM_{10}$  et la vitesse du vent montre un décalage dans le temps des fortes valeurs des deux paramètres sur les stations de Cinzana et Banizoumbou. Les  $PM_{10}$  les plus forts sont enregistrés en mars alors que les vents les plus forts sont enregistrés en début de saison des pluies, au mois de juin. Ce décalage montre que le vent local n'est pas responsable des maximums de  $PM_{10}$  en saison sèche, ce qui conforte l'idée que les transports d'aérosols sahariens sont responsables des maxima de concentration observés en saison sèche au Sahel [Marticorena et al., 2010]. Par ailleurs le gradient d'Est en Ouest c'est à dire de Banizoumbou à Mbour des concentrations médianes mensuelles apparaît être lié à la position géographique de chaque station par rapport aux différentes zones sources Sahariennes. Banizoumbou ou les plus fortes concentrations ont été enregistrées est beaucoup plus proche des zones reconnues comme étant parmi les plus grandes productrices en aérosols désertiques en Afrique telles que le sud de Algérie, en Libye et au Tchad (dépression de Bodélé) [Prospero et al., 2002]. Les données recueillies sur les cinq années ont permis de quantifier la variabilité des concentrations moyennes mensuelles. Il existe une très forte variabilité des concentrations mensuelles sur les cinq années de mesure qui peut être en partie liée à la variabilité interannuelle. Des concentrations médianes mensuelles très élevées ou faibles peuvent enregistrés à n'importe quelle période de 'année. C'est le cas du mois de janvier 2007, où la concentration médiane est significativement plus élevée que les autres années pour l'ensemble des trois stations. L'année 2009 apparaît comme l'année la moins poussiéreuse sur la période des cinq années dans les trois stations. Elle est également parmi les années les plus sèches à Cinzana et M'Bour sur la période des cinq années. Les plus faibles concentrations sont enregistrées pendants la saison des pluies coïncidant avec la période de maximum de précipitation obtenu en août à Cinzana et Banizoumbou et en septembre à Mbour. Les concentrations médianes mensuelles en saison des pluies sont deux fois plus élevées à M'Bour (~ 30  $\mu g.m^{-3}$ ) qu'à Banizoumbou et Cinzana (~ 10  $\mu g.m^{-3}$ ) malgré le fait que les précipitations annuelles sont dans le même ordre de grandeur qu'à Banizoumbou. Cette analyse nous a également permis de mettre en évidence un cycle diurne clair et distinct entre la saison des pluies et la saison sèche sur les trois stations. Pendant la saison sèche, le cycle diurne de la concentration en  $PM_{10}$  est caractérisé par un maximum à 10 h et minimum pendant la nuit, coïncidant avec le cycle diurne de la vitesse du vent. Au Sahel, le cycle diurne des vents de surface est fortement liée à la dynamique du jet NLLJ [Lothon et al., 2008; Guichard et al., 2009]. Le phasage des pics de vitesse de vent et de concentration en  $PM_{10}$  suggère que le NLLJ saharien module le cycle diurne des concentrations en particules mesurées au Sahel et qu'il est un agent de transport efficace d'aérosols sahariens vers le Sahel en saison sèche. Pendant la saison des pluies, le cycle diurne de la concentration en  $PM_{10}$  apparaît diffèrent de celui de la saison sèche. La concentration en  $PM_{10}$  présente des maxima à partir de la fin d'après midi, entre 16h et 23h, sur les stations de Banizoumbou et Cinzana. Ces maxima coïncident avec un renforcement des directions de vent du flux de mousson (180-270°). Pour la station de Mbour, le cycle diurne est différent des autres stations avec un pic en fin d'après midi (19h) et un pic en début de matinée (8h). Le premier coïncide avec un renforcement des vents de secteur ouest nord-ouest. Le cycle diurne de la concentration en  $PM_{10}$  est en déphasage avec celui de la vitesse de vent qui présente un maximum entre 15 et 17 h. Ces résultats suggèrent que la variabilité des concentrations de poussière en saison des pluies est modulée par l'occurrence des systèmes convectifs à l'échelle régionale.

Atmospheric Research 164-165 (2015) 226-241

Contents lists available at ScienceDirect



Atmospheric Research

journal homepage: www.elsevier.com/locate/atmos

# Variability of mineral dust concentrations over West Africa monitored by the Sahelian Dust Transect



F. Kaly <sup>a,b,c,\*</sup>, B. Marticorena <sup>b</sup>, B. Chatenet <sup>b</sup>, J.L. Rajot <sup>b,d</sup>, S. Janicot <sup>c</sup>, A. Niang <sup>a</sup>, H. Yahi <sup>b,c</sup>, S. Thiria <sup>c,h</sup>, A. Maman <sup>e</sup>, A. Zakou <sup>e</sup>, B.S. Coulibaly <sup>f</sup>, M. Coulibaly <sup>f</sup>, I. Koné <sup>f</sup>, S. Traoré <sup>f</sup>, A. Diallo <sup>g</sup>, T. Ndiaye <sup>g</sup>

<sup>a</sup> LTI, UCAD, Dakar, Senegal

- <sup>b</sup> LISA, UMR CNRS 7583; UPEC; UPD IPSL, Créteil, France
- <sup>c</sup> LOCEAN, IPSL, UPMC, Sorbonne Universités, Paris, France
- <sup>d</sup> iEES Paris, UMR IRD 242; UPMC; CNRS; INRA; UPEC; UPD, Bondy, France

<sup>e</sup> IRD, Niamey, Niger

<sup>f</sup> IER, Cinzana, Mali

<sup>g</sup> IMAGO, US 191, IRD, Dakar, Senegal

<sup>h</sup> UVSQ, F-78035 Versailles, France

#### ARTICLE INFO

Article history: Received 5 November 2014 Received in revised form 10 April 2015 Accepted 18 May 2015 Available online 30 May 2015

Keywords: Aerosol Mineral dust aerosol PM<sub>10</sub> Spatiotemporal variability Sahel West Africa

#### ABSTRACT

The "Sahelian belt" is known as a region where mineral dust content is among the highest in the world. In the framework of the AMMA international Program, a transect of three ground based stations, the "Sahelian Dust Transect" (SDT), has been deployed in order to obtain quantitative information on the mineral dust content over the Sahel. These three stations: Banizoumbou (Niger), Cinzana (Mali) and M'Bour (Senegal) are aligned at 13°N along the east–west main pathway of the Saharan and Sahelian dust toward the Atlantic Ocean. The SDT provides a set of aerosol measurements and local meteorological parameters to describe and understand the mechanisms that control the temporal and regional variability of mineral dust content in these regions.

In this work we analyze the seasonal and diurnal variability of the dust concentrations over the period 2006–2010. The analysis of the dust concentrations measured between 2006 and 2010 confirmed a regional seasonal cycle characterized by a maximum in the dry season, with median concentration ranging from 205  $\mu$ g m<sup>-3</sup> at Banizoumbou to 144  $\mu$ g m<sup>-3</sup> at M'Bour, and a minimum (11–32  $\mu$ g m<sup>-3</sup>) in the wet season. The five year data set allowed the quantification of the variability of the monthly concentrations. The range between the percentiles 75 and 25 varies linearly with the median concentration: it is of the same order than the median value in M'Bour, 17% slightly higher in Cinzana and 50% higher in Banizoumbou. The range between the accepted maximum and minimum is also correlated with the median value, with slopes ranging from 14 in Banizoumbou to 7 in M'Bour. Part of the variability of the concentration at the monthly scale is due to interannual variability. Extremely high or low monthly concentration can be recorded that significantly impacts the five year median concentration and its range. Compared to the 3-year data set analyzed by Marticorena et al. (2010), the two additional years used in this work appear as the less dusty year (2009) and one of the dustier years (2010).

The sampling time step and the high recovery rates of the measurement stations allowed to investigate the diurnal cycle of the dust concentration for the first time. During the dry season, the hourly median concentrations range from 80 to 100  $\mu$ g m<sup>-3</sup> during the night to 100–160  $\mu$ g m<sup>-3</sup> during the day-time maximum. The diurnal cycle of the PM<sub>10</sub> concentrations is phased with the diurnal cycle of the surface wind speed and thus modulated by the interactions between the nocturnal lower level jet (NLLJ) and the surface boundary layer. The NLLJ appears as a major agent to transport Saharan dust toward the Sahel. During the wet season, the median PM<sub>10</sub> concentrations are maximum at night-time (<50  $\mu$ g m<sup>-3</sup>). The night-time concentrations are associated with a large range of variability and coincide with the periods of higher occurrence of meso-scale convective systems. The amplitude of the diurnal cycle is of the order of 60  $\mu$ g m<sup>-3</sup> in the dry season and 20  $\mu$ g m<sup>-3</sup> in the wet season, a typical diurnal pattern has been established suggesting that this temporal pattern is mainly driven by local meteorological conditions.

© 2015 Elsevier B.V. All rights reserved.

\* Corresponding author at: Laboratoire de Traitement de l'Information/ESP, Université Cheikh Anta Diop, BP 5085, Dakar Fann, Senegal. *E-mail address:* fak\_01@hotmail.com (F. Kaly).

http://dx.doi.org/10.1016/j.atmosres.2015.05.011 0169-8095/© 2015 Elsevier B.V. All rights reserved.

#### 1. Introduction

The Sahara and Sahel regions are the largest arid to semi-arid areas of the world and the most important sources of mineral dust with annual emissions estimated between 400 and 2200 Tg (e.g. Huneeus et al., 2011).

These aerosols influence the Earth's radiative budget directly by scattering and absorbing solar radiation in the atmosphere and indirectly by affecting cloud formation and, thus, cloud albedo (Sokolik and Toon, 1999). Mineral dust radiative forcing represents therefore a significant uncertainty in predicting climate change. Their radiative effects also have an impact on atmospheric dynamics at the local and regional scales and mineral dust may be responsible for the inhibition of hurricane activity on the tropical North Atlantic (Evan et al., 2006). This interaction between mineral dust and radiations is such that it affects the quality of forecasts of West Africa (Tompkins et al., 2005). Mineral dust transported over long distances can play a decisive role in the availability of nutrients (such as iron and phosphorus) for large oceanic and some continental ecosystems (Swap et al., 1992; Jickells et al., 2005).

Finally, mineral dust is suspected to play a role in the occurrence of meningitis epidemics in the Sahelian "meningitis belt". Indeed, meningitis outbreaks occur during the dry season, a few weeks after the maximum levels of mineral dust that can act as irritant and thus favor the development of the disease (Martiny and Chiapello, 2013; Deroubaix et al., 2013). Thomson et al. (2006) also suggest that the high dust concentrations anomaly in late autumn could increase the incidence of meningitis epidemics.

The evaluation of the different impacts of mineral dust requires the quantification of its atmospheric content, in particular near source areas. Such quantification must take into account the strong variability of the atmospheric concentrations induced by weather conditions, which affect emissions, transport and deposition of these aerosols.

Seasonality and variability of mineral dust occurrence over the North tropical Atlantic Ocean have been extensively studied in recent decades. In the Sahel and the Sahara, aerosol optical depths (AODs) measured by sun photometers from the AERONET network (http://aeronet.gsfc.nasa.gov/) have allowed to investigate the seasonal cycle of the vertically integrated atmospheric dust content in different stations of West Africa (Holben et al., 2001; Ogunjobi et al., 2008). Based on aerosol indexes (AIs) from the TOMS (Total Ozone Mapping Spectrometer) spatial instrument, Engelstaedter et al. (2006) showed a well-marked annual cycle of the mineral dust load in North Africa. On a longer time-scale, the analysis of 11 years (1982–1997) of AOD derived from the Meteosat satellites revealed a relationship between the transport of aerosols from the Saharan desert and the Sahel toward the Mediterranean Sea and the Atlantic Ocean and the North Atlantic Oscillation (NAO) (Moulin et al., 1997).

The aerosol optical depth may not always be a good indicator of the variability of mineral dust concentration at the surface due to the variability in the altitude of dust transport. The relationship between the surface concentrations and the AOD can vary significantly depending on the meso-scale meteorological conditions over the Sahel (Yahi et al., 2013). As a result, the seasonal cycles of the surface concentration and the AODs are not phased, with maximum dust concentrations observed several weeks (3 to 8) before the maximum of AOD (Deroubaix et al., 2013). Another proxy of the dust concentration is the horizontal visibility, with high concentrations producing severe reductions in the horizontal visibility. N'Tchayi M'Bourou et al. (1994, 1997) or Ozer (2001), for example, has used the horizontal visibility from weather reports to quantify the spatial and temporal variability of surface dust atmospheric content. Such measurements are thus mainly used to describe the occurrence of dust storm and dust events (i.e., Goudie, 1983), whereas visibility measurements by local operators are not totally reliable from a quantitative point of view, and can vary from one station to the other or from one operator to the other (Middleton, 1986). The analysis of weather reports from meteorological stations showed that from 1983 to 2008 dust events in the Sahel were mainly reported as "dust in suspension" and were due to Saharan dust transport rather than to local dust emissions (Klose et al., 2010). The synoptic patterns associated to the major dust events recorded during the studied period have been analyzed to identify the meteorological conditions responsible for the annual maximum of the dust event intensity in the spring (Klose et al., 2010). Cowie et al. (2014) also used the dust synoptic observations and wind speed measurements from meteorological stations over the Sahara and the Sahel (1984-2012) to propose a climatology of dust emission frequency and an estimation of the dust emission potential. They showed a maximum in the dust occurrence at the end of winter and early spring in the central Sahel, but a more complex pattern in the west Sahel. They also discuss the seasonal variation of the diurnal cycle of the dust emission frequencies. Despite a limited number of night-time observations in the Sahelian stations, their results suggest a strong influence of the nocturnal lower level jet (NLLJ) on the diurnal cycle of dust event occurrence. Low level jets (LLJs) are most commonly observed at night-time and can occur over all desert surfaces worldwide. They are characterized by horizontal wind speed maximum in the lowest few kilometers of the atmosphere (e.g., Blackadar, 1957; Holton, 1967; Banta et al., 2006). Over the Sahara, the NLLJ is suspected to play an important role for dust emission (Washington et al., 2006; Bou Karam et al., 2008; Knippertz and Todd, 2012; Heinold et al., 2013). After sunrise, surface heating causes the atmospheric boundary layer to grow in depth and mixes momentum from the LLJ down to the surface creating the distinctive peak of surface wind speeds from morning to midday observed in many Saharan locations (Todd et al., 2008; Schepanski et al., 2009; Marsham et al., 2011, 2013).

A limitation of the studies based on synoptic observations is that they allow to investigate the dust event occurrence only. Dust emissions or dust uplift potential can be computed from the measured surface winds which are available with a 3 h time step and not always measured during the night in the Sahelian region (e.g. Cowie et al., 2014).

In the framework of the AMMA (African Monsoon Multidisciplinary Analysis) international program, three stations, the "Sahelian Dust Transect", have been deployed in the Sahelian zone to document more precisely the variability of mineral dust and in particular the surface dust concentrations. Based on a three year data set of daily concentrations (2006–2008), Marticorena et al. (2010) showed that the dust concentrations at the three stations exhibit a marked seasonal cycle. This seasonal cycle is characterized by a monthly maximum during the dry season (February to April) and a minimum occurring during the rainy season (August-September). The results indicate that the general pattern of dust concentration is similar at regional scale. A decreasing gradient of the dust concentration is observed from Niger to Senegal, that they explained by the dilution of the most intense dust plumes moving from east to west. The seasonal cycle of the dust concentrations is not phased with the seasonal cycle of the local surface wind speed. This suggests that the maximum dust concentrations observed in the spring are mainly controlled by dust transported by the Harmattan flow from remote Saharan sources, in agreement with the analysis of Klose et al. (2010). Marticorena et al. (2010) also showed that local dust emissions generated by strong surface winds associated with meso-scale convective systems are responsible for the occurrence of extremely high daily concentrations observed at the beginning of the rainy season in Banizoumbou (Niger) and Cinzana (Mali). Indeed, strong dust-emitting winds occur at the leading edge of the cold-pool outflow generated by evaporating precipitation from convective systems, with scale ranging from local microburst to well organized squall lines of several hundreds of kilometers (e.g., Knippertz and Todd, 2012; Marsham et al., 2008). Moist convection also plays a critical role on the precipitation regime in these regions, the annual rainfall amount being provided by a limited number of organized convective systems (Mathon et al., 2002). Precipitation can efficiently scavenge local dust emissions

and dust transported from remote Saharan sources which explain the minimum of concentrations in the wet season (Marticorena et al., 2010).

The purpose of this paper is to investigate and quantify the variability of the dust concentrations from the seasonal to the daily scale over a five year period and to identify the role of the local meteorological parameters in this variability. The experimental data are described in Section 2. Section 3 presents the annual mean concentrations at the three stations and their variability. Section 4 discusses the variability of the dust concentrations at the seasonal scale. Section 5 investigates the diurnal cycle of the dust concentrations are summarized in the final section.

#### 2. Experimental data

#### 2.1. The Sahelian Dust Transect

In the framework of the AMMA (African Monsoon Multidisciplinary Analysis) international project (Redelsperger et al., 2006), a set of three stations the so-called "Sahelian Dust Transect" has been deployed in 2006 to monitor the mineral dust content over West Africa. As described in Marticorena et al. (2010), the three stations M'Bour (Senegal, 14.39°N, 16.96°W), Cinzana (Mali, 13.28°N, 5.93°W) and Banizoumbou (Niger, 13.54°N, 2.66°E) are almost aligned around 13–14° north, on the main pathway of Saharan and Sahelian dust toward the Atlantic Ocean (Fig. 1). They are located in the semi-arid Sahel, and the annual mean precipitation for the period 2006–2010 is, respectively, 496 mm in Banizoumbou, 715 mm in Cinzana and 511 mm in M'Bour.

In Senegal, the measurement site is located in the "Station de Géophysique" of the Institut de Recherche pour le Développement (IRD), south of the city of M'Bour, about 85 km from Dakar. The instruments are installed at ~10 m height, on the roof of a building close to the seaside. In Mali, the instruments have been deployed in an agronomical research station (Station de Recherche Agronomique de Cinzana, SRAC) of the Institut d'Economie Rurale (IER), 40 km east south-east of the town of Ségou. In Niger, the station is located in a fallow, 2.5 km from the village of Banizoumbou, about 60 km east of the capital Niamey. The instrumentation of the stations is detailed in Marticorena et al. (2010). In this study, two types of in-situ data are used: the dust surface concentration and local meteorological parameters.

#### 2.2. Measurements

#### 2.2.1. Atmospheric dust concentrations

Atmospheric concentrations of Particulate Matter smaller than 10 µm (PM<sub>10</sub>) are measured using a Tapered Element Oscillating Microbalance (TEOM 1400A from Thermo Scientific) equipped with a PM<sub>10</sub> inlet. The inlet is located at ~6.5 m height in Mali and Niger and ~10 m in Senegal. This cut off diameter does not cover the entire size spectrum of mineral dust, but it allows homogeneous measurements on all sites, comparable to data from the literature and to air quality standards. In Banizoumbou and Cinzana, mineral dust largely controls the mass concentrations, so the PM<sub>10</sub> concentrations can be considered as representative of the mineral dust concentration (Marticorena et al., 2010). The station of M'Bour being located close to the sea side, the PM<sub>10</sub> concentrations can be impacted by sea-salt aerosol brought in the westerly winds and by sea breezes. In M'Bour, 40% of the daily mean PM<sub>10</sub> concentrations associated with mineral dust are higher than 100  $\mu$ g m<sup>-3</sup> and less than 8% of the daily means are lower than 20  $\mu$ g m<sup>-3</sup>. On the Canary Islands, Viana et al. (2002) measured sea-salt concentrations ranging from 3 to 20  $\mu$ g m<sup>-3</sup> over a two year period. This concentration range suggests that, compared to mineral dust, sea-salt aerosol is not a major fraction of the daily aerosol mass concentration measured in coastal location such as M'Bour.

#### 2.2.2. Meteorological parameters

The basic meteorological parameters (wind speed and direction, air temperature, relative humidity) were measured with Campbell Scientific Instruments. Wind speed and wind direction are measured with a 2DWindSonic, temperature and relative humidity using 50Y or HMP50 sensors and rainfall with an ARG100 tipping bucket raingauge. The data acquisition is made using data loggers CR200. Meteorological measurements are made at ~10 m height in Senegal, ~6.5 m height in Niger and ~2.3 m height in Mali.

#### 2.2.3. Temporal representativity

The data have been recorded continuously from January 2006 to December 2010. They are recorded with a 5 min time step and have been averaged at an hourly time step. The average temporal representativity of the data set for the  $PM_{10}$  concentrations is 95% in Banizoumbou, 92% in Cinzana and 98% in M'Bour. The lowest temporal representativity in Cinzana is due to a computer problem in July 2007 (monthly recovery



Fig. 1. Average daily OMI (ozone monitoring instrument) UV aerosol index from the 1st January 2006 to the 31st December 2010 over West Africa. The black stars correspond to the location of the stations of the Sahelian Dust Transect: Banizoumbou (Niger, 13.54°N, 2.66°E), Cinzana (Mali, 13.28°N, 5.93°W), and M'Bour (Senegal; 14.39°N, 16.96°W).

rate of 26%) and to a pump break down from the end of April 2009 to the end of June 2009 (monthly recovery rates of 0 and 45% respectively). A pump break down also occurred in Banizoumbou in June 2009, decreasing the monthly recovery rate to 36%. Except these specific months, the monthly recovery rates vary from 66 to 100% in Cinzana, from 64 to 100% in Banizoumbou and from 85 to 100% in M'Bour. Among the 65 sampled months, 80% have recovery rates higher than 90% in Banizoumbou, 77% in Cinzana and 85% in M'Bour. Such a sampling rate is extremely remarkable considering the location of the stations and their environmental and technical conditions.

#### 3. Annual means and variability

A first global analysis of the data is given at the annual scale. For the PM<sub>10</sub> concentration, the wind speed, the air temperature and the relative humidity, the annual mean and median values, the 75th and 25th percentiles (or Q<sub>3</sub> and Q<sub>1</sub> quartiles), the accepted maximum  $(Q_3 + 1.5 (Q_3 - Q_1))$  and minimum  $(Q_1 - 1.5 (Q_3 - Q_1))$  and the extreme values are displayed as "box-plots" in Fig. 2. The mean air temperature exhibits an East to West gradient, since it ranges from 29° in Banizoumbou to 26° in M'Bour (Fig. 2a). It appears as the less dispersed meteorological parameter, with a standard deviations of the order of 5° (5.6° in Banizoumbou to 60% in M'Bour (Fig. 2b). The standard deviation is quite high (~25%) and comparable at the three stations. The mean values are higher than the median value, suggesting a distribution biased toward high values.

The highest annual mean wind velocity is recorded in Banizoumbou and the lowest in Cinzana (Fig. 2c). However, since the height at which the wind velocity is measured is not the same for the three stations, these values cannot be compared directly. These differences in the height of measurements can be accounting for assuming a neutral logarithmic wind profile and a roughness length typical of a Sahelian field (Marticorena et al., 2010). The mean wind speed measured at 6.5 m in Banizoumbou (2.67 m s<sup>-1</sup>) corresponds to a 10 m wind speed of 2.90 m s<sup>-1</sup>. Similarly, the mean wind speed measured at 2.3 height in Cinzana ( $1.57 \text{ m s}^{-1}$ ) corresponds to a 10 m height wind speed of 2.17 m s<sup>-1</sup>. Compared to the 10 m mean speed measured in M'Bour (2.26 m s<sup>-1</sup>), Banizoumbou appears as the station with the highest mean wind speed and Cinzana the lowest. At the three stations, the standard deviation of the mean wind is high, about 50% of the annual mean and extremely high values are recorded, as clearly seen in Fig. 2c.

The hourly mean PM<sub>10</sub> concentrations span over several orders of magnitude: 90% of the hourly concentrations measured at the three stations range between 10 and 1000  $\mu$ g m<sup>-3</sup>, but extreme values higher than 10,000  $\mu$ g m<sup>-3</sup> and lower than 0.01  $\mu$ g m<sup>-3</sup> are recorded in Banizoumbou and Cinzana. The statistical parameters of the concentration distributions have been computed on a decimal logarithmic base (Fig. 2d). The mean value is almost equal to the median values, suggesting a log-normal distribution of the hourly PM<sub>10</sub> concentrations. When converted the annual mean Log<sub>10</sub> of the concentration into concentrations, an annual mean value of  $69 \,\mu g \, m^{-3}$  is obtained in Banizoumbou, 59  $\mu$ g m<sup>-3</sup> in Cinzana and 76  $\mu$ g m<sup>-3</sup> in M'Bour. Marticorena et al. (2010) indicate an East to West gradient in the arithmetic mean PM<sub>10</sub> concentrations, with higher mean concentrations in Banizoumbou than in M'Bour. When computing the same arithmetic means, we obtain a annual mean concentration of 163  $\mu g \; m^{-3}$  in Banizoumbou, 119  $\mu$ g m<sup>-3</sup> in Cinzana and 112  $\mu$ g m<sup>-3</sup> in M'Bour, i.e. a similar East to West gradient than Marticorena et al. (2010). The range of the Log<sub>10</sub> of the concentrations is larger in Banizoumbou (151  $\mu g \ m^{-3}$  and 29  $\mu g~m^{-3})$  than in Cinzana (120  $\mu g~m^{-3}$  and 28  $\mu g~m^{-3})$  and than in M'Bour (129  $\mu g~m^{-3}$  and 46  $\mu g~m^{-3})$ . The accepted maximum is much higher in Banizoumbou (1738  $\mu g \ m^{-3})$  than in Cinzana



**Fig. 2.** Median (red line), percentiles 25% and 75% (box), concentrations higher and lower than the accepted maximum and minimum (red crosses) and mean (black crosses) of the hourly air temperature (a), relative humidity (b), wind speed (c) and Log<sub>10</sub> of the PM<sub>10</sub> concentrations (d) measured at the three stations from 2006 to 2010. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(1069  $\mu$ g m<sup>-3</sup>) and M'Bour (602  $\mu$ g m<sup>-3</sup>). More precisely, in Banizoumbou 56% of the data ranges between 10 and 100  $\mu$ g m<sup>-3</sup> and 35% between 100 and 1000  $\mu$ g m<sup>-3</sup>. At the station of Cinzana, 61% of  $PM_{10}$  data range between 10 and 100  $\mu g \; m^{-3}$  and 30% between 100 and 1000  $\mu$ g m<sup>-3</sup>. For the M'Bour station, 62% of the data range between 10 and 100  $\mu g$  m  $^{-3}$  and 36% between 100 and 1000  $\mu g$  m  $^{-3}.$ As highlighted in Fig. 1, the stations are located in the vicinity of important dust sources of the Sahara. Marticorena et al. (2010) suggested that the west to east gradient in the PM<sub>10</sub> concentrations is linked to their geographical positions compared to the main dust sources. Banizoumbou is closer to zones recognized as being among the most active sources of mineral dust in the North of Africa such as Libya and Chad (depression of Bodélé) (Prospero et al., 2002). Dust plumes originating from these eastern sources are transported to the west and diluted when successively recorded at the stations of Cinzana and M'Bour. The distribution of the data also reveals extremely high concentrations (>10,000  $\mu g~m^{-3})$  occasionally recorded in Banizoumbou and Cinzana. Over the five years of measurements, a maximum hourly concentration of 28,951  $\mu$ g m<sup>-3</sup> has been recorded in Banizoumbou, 14,563  $\mu$ g m<sup>-3</sup> in Cinzana and 7590  $\mu$ g m<sup>-3</sup> in M'Bour. For all the stations, the hourly minimum of PM<sub>10</sub> concentrations (>0.4  $\mu$ g m<sup>-3</sup>) is 6 orders of magnitude lower than the maximum.

#### 4. Seasonal variability

The Sahelian climate is characterized by pronounced seasonal contrasts alternating between a dry season and a rainy season, regulated by the displacement of the inter-tropical discontinuity (ITD). North of the ITD, the main flow is a hot and dry north-easterly flow, the socalled "Harmattan". South of the ITD, the monsoon blowing from south-west is the dominant wind regime (Lele and Lamb, 2010). Within this monsoon layer, the highest precipitation is located in the intertropical convergence zone (ITCZ). Based on the analysis of long timeseries of meteorological measurements in Niamey, Slingo et al. (2008) observed that the establishment of the monsoon starts modestly in April–May. The monsoon rainfall increases markedly to a late-July to mid-August maximum, before declining more abruptly through September and October. The authors consider that this behavior prevails across the entire Sahel, from Senegal to Ethiopia. In agreement with these observations, in this work we define the dry season as the period for which no significant precipitation is recorded, i.e. from November to April, and the wet season as the period when precipitation is measured (May to October).

The available five year data set allows the establishment of a typical seasonal cycle of the mineral dust content at the three stations and to quantify the variability of the monthly concentration on a multiannual basis.

#### 4.1. PM<sub>10</sub> concentration

The succession of the Harmattan and the Monsoon flow is clearly highlighted by the monthly variations of the surface wind directions (Fig. 3). In the dry season, the Harmattan regime induces a N-NE  $(0-90^\circ)$  dominant wind direction in Cinzana and M'Bour and an Easterly direction  $(45-135^\circ)$  in Banizoumbou. During the monsoon season, the dominant wind directions are similar in Banizoumbou and Cinzana, i.e. S-SW (~225°) while the station of M'Bour is mainly influenced by surface winds originating from W and NW directions (225–315°).

Considering the lognormal distribution of the hourly  $PM_{10}$  concentrations, the median and 75th and 25th percentiles and the accepted maximum and minimum of the decimal logarithm of the hourly concentrations have been computed at the monthly scale over the five year period (Fig. 4). They are given for each of the five years in Appendix A. For the sake of clarity, in the following, the statistical parameters of the log-normal distribution are discussed in terms of  $PM_{10}$  concentrations.

A similar seasonal cycle of the PM<sub>10</sub> concentrations is observed in Banizoumbou and Cinzana, with a maximum by the end of the dry season (~March) and a minimum in August during the rainy season. The maximum median PM<sub>10</sub> concentration is higher in Banizoumbou  $(205 \ \mu g \ m^{-3})$  than in Cinzana  $(165 \ \mu g \ m^{-3})$ . This seasonal cycle of the dust concentration is exactly opposite to the seasonal cycle of the horizontal visibility measured in Niamey (Niger) (Slingo et al., 2008). Indeed, high (low) dust concentrations cause low (high) horizontal visibility. This seasonal cycle is consistent with Marticorena et al. (2010) who analyzed a 3-year time series (2006-2008). The seasonal cycle of the concentration is very similar in M'Bour but with slight differences in the timing of the maximum and minimum concentrations. The maximum concentration is recorded in January (144  $\mu$ g m<sup>-3</sup>) rather than in March (126  $\mu$ g m<sup>-3</sup>) and the minimum concentration in September  $(32 \ \mu g \ m^{-3})$  rather than in August  $(35 \ \mu g \ m^{-3})$ . In the wet season, the median concentrations in M'Bour are much higher than at the two other stations: in August, the concentration is then twice higher in M'Bour than in Banizoumbou (13  $\mu g m^{-3}$ ) and in Cinzana  $(11 \,\mu g \, m^{-3})$ . At the three stations, the 75th and 25th percentiles follow the same seasonal pattern than the median. In fact, the interquartile range (IQR:  $Q_3 - Q_1$ ) is linearly correlated with the median values (r<sup>2</sup> from 0.86 in Cinzana to 0.98 in M'Bour), with slopes exhibiting an East to West gradient, from 1.5 in Banizoumbou, to 1.02 in M'Bour, with an intermediate value of 1.17 in Cinzana. The maximum concentration associated with the 75th percentile ranges from 450  $\mu$ g m<sup>-3</sup> in Banizoumbou in March to 237  $\mu g$  m<sup>-3</sup> in M'Bour in January, with an intermediate value of 309  $\mu$ g m<sup>-3</sup> in Cinzana in March. The minimum concentrations associated with the 75th percentile range from 20  $\mu$ g m<sup>-3</sup> in Banizoumbou and Cinzana in August to 50  $\mu$ g m<sup>-3</sup> in M'Bour in September. Similar trends are observed for the concentrations associated with the 25th percentile. Their seasonal maxima range from 113  $\mu$ g m<sup>-3</sup> in Banizoumbou to 87  $\mu$ g m<sup>-3</sup> in M'Bour while their minima range from 9 to 7  $\mu g \ m^{-3}$  in Banizoumbou and Cinzana and are much higher in M'Bour (48  $\mu g \ m^{-3}$  in September). The standard deviation of the Log<sub>10</sub> of the concentrations does exhibit any clear seasonal cycle. But because of the seasonal variations of the concentrations, the highest coefficient of variations (CVs) are recorded in the wet season (0.32 in Banizoumbou in July, 0.31 in Cinzana in August and 0.25 in M'Bour in September) and the minimum CVs are recorded in the dry season in Banizoumbou (0.18 in April) and in Cinzana (0.15 in February). In M'Bour the lowest CVs are recorded in June (0.13) and November (0.14). The maximum and minimum accepted concentrations follow the same seasonal pattern than the monthly median concentrations. The range between the maximum and minimum concentration is very large at the three stations. It is linearly correlated with the median concentrations with slopes increasing from M'Bour (6.7) to Cinzana (8.8) and Banizoumbou (14). This means that in M'Bour where monthly median concentration varies from 32 to  $144 \,\mu g \, m^{-3}$  the range between the maximum and minimum concentration is from 265 to 1045  $\mu$ g m<sup>-3</sup>. In Banizoumbou where the median concentrations vary from 13 to 198, the range varies from 73 to 1934  $\mu g m^{-3}$ .

These large ranges of concentration in the dry season are in a large part related to the interannual variability. When examining the monthly median concentrations of the different years (Appendix A), despite a similar seasonal pattern, unusually high or low monthly concentrations can occur at any month of any year during the dry season. These high or low monthly concentrations significantly impact the 5-year statistics. The most extreme example is given by the monthly median concentrations recorded in January 2007. It is higher than the five-year median concentration by 54% in M'Bour, 156% in Cinzana and 220% in Banizoumbou. It corresponds to a very intense event of Saharan dust transport that lasted about 7 days at the three stations, with concentrations higher than during the continental dust storm affecting the whole Sahara and West Africa in March 2006 (Slingo et al., 2008; Marticorena et al., 2010). Excluding the data recorded during this specific month



Fig. 3. Five year average monthly frequency of wind directions at (a) Banizoumbou, (b) Cinzana and (c) M'Bour.

would lead to a multiannual median concentration of 89  $\mu g m^{-3}$ instead of 103  $\mu$ g m<sup>-3</sup> and would reduce the monthly CV to 0.18 instead of 0.2. But it would mainly affect the range of the concentrations, with a large differences in the 75th percentile (150  $\mu g m^{-3}$  instead of 205  $\mu g~m^{-3})$  and in the maximum concentration (665  $\mu g~m^{-3}$ instead of 1280  $\mu g\ m^{-3}).$  In Cinzana excluding the data of January 2007 would decrease the median concentration from 95 to 81  $\mu$ g m<sup>-3</sup>, the 75th percentile from 190 to 143  $\mu g$  m<sup>-3</sup> and the maximum concentration from 1047 to 603  $\mu$ g m<sup>-3</sup>. An example of abnormally low concentration is given by the month of March 2008 in Banizoumbou. The monthly median value for March 2008 is 47% lower than the fiveyear median concentration. Excluding these monthly data from the multiannual statistic would increase the median value to 254  $\mu g \ m^{-3}$ instead of 205  $\mu$ g m<sup>-3</sup>. The examination of the maximum and minimum monthly median concentrations over five years shows that it is not simple to state whether a year is specifically "dusty" or "non-dusty"

(Table 1). For example, in Banizoumbou, four of the highest monthly median concentrations are recorded in 2010 but also five of the lowest median concentrations. The year 2008 also has the same number of highest and lowest monthly median concentrations both in Banizoumbou and in Cinzana. In Banizoumbou and Cinzana, the year 2009 appears as the less dusty year, none of the maximum monthly median and 2 to 3 of the lowest median concentrations being recorded in 2009. In M'Bour, the year 2009 has the highest monthly concentration in November but also 5 of the lowest monthly median concentrations during the rest of the year. At this station 6 of the highest monthly median concentrations are recorded in 2007 and 4 in 2008 while 3 of the lowest values are recorded in 2006 and 3 in 2010. Considering the number of highest monthly median concentrations, the dustier year is 2010 in Banizoumbou and 2007 in Cinzana and M'Bour. The two additional years of observations used in this study compared to the data set used in Marticorena et al. (2010) finally increase the



**Fig. 4.** Monthly median (red line), percentiles 25% and 75% (box), concentrations higher and lower (red crosses) than the accepted maximum and minimum (vertical bar and black dash) of the  $Log_{10}$  of the hourly  $PM_{10}$  concentrations measured from 2006 to 2010 at Banizoumbou (a), Cinzana (b) and M'Bour (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

range of the recorded concentrations by adding a very dusty year and the less dusty year of the five year time series.

#### 4.2. Surface winds

The presence of dust in the atmosphere can be due either to transport from remote sources or to local dust emissions. Examining the seasonal cycle of the local surface winds can help to discriminate these two factors.

The average monthly wind speeds measured at the three stations and the monthly coefficients of variation are shown in Fig. 5. A clear and similar seasonal cycle of wind speed is observed at Banizoumbou

# Table 1

Annual precipitation	(mm) from 2	006 to 2010,	mean annua	l precipitation	and s	tandard
deviation at the three	stations of th	ne Sahelian D	ust Transect.			

	Banizoumbou	Cinzana	M'Bour
2006	504	713	570
2007	456	911	321
2008	699	696	568
2009	391	677	770
2010	316	847	775
Mean	473	769	601
σ	145	104	187

and Cinzana, with a maximum in June and a minimum in October-November. In M'Bour, a maximum wind speed is observed in February with a secondary maximum in July, the minimum wind speed being recorded in September. These seasonal cycles differ from the seasonal cycle of the wind direction (Fig. 3) and are not phased with the succession of the dry and wet seasons. In Banizoumbou and Cinzana, the maximum of monthly wind speed observed at the beginning of the wet season coincides to south-westerly directions. It is due to the strong winds associated with convective systems that start developing in the North Sahel at this period (Marticorena et al., 2010). A similar seasonal cycle of the surface wind speed was observed by Guichard et al. (2009) for the station of Agoufou in the Malian Gourma. In M'Bour, the maximum wind speed is observed during the dry season (February-March). It is associated with a North-East direction (Fig. 3) and typically corresponds to the Harmattan flow. A secondary maximum is observed in June-July, that can be associated with convective systems that occur later in the wet season over M'Bour than over Banizoumbou and Cinzana. The monthly standard deviations of the wind speed are of the same order of magnitude but lower (0.75 to 2 m s<sup>-1</sup>) compared to the monthly means (1 to 4 m s<sup>-1</sup>) (Table 2). They are higher in the dry season and lower in the wet season at Banizoumbou and Cinzana, except for the month of June which exhibits the highest standard deviation in Banizoumbou. On the opposite in M'Bour the highest standard deviations occur in the wet season and the lowest at the beginning of the dry season. The monthly CVs exhibit quite complex behaviors compared to the seasonal cycle of the mean wind speeds. In Banizoumbou, the lowest CVs (0.4-0.5) are observed in the wet season, a period that includes the maximum monthly wind speeds but also quite low monthly means. The highest CVs correspond to the dry season, during which the surface winds are lower than in the wet season. In Cinzana, the lowest CVs correspond to the dry season. They are slightly higher in the wet season. In M'Bour, the highest CVs are recorded in the dry season, like in Banizoumbou, and the lowest at the beginning of the wet season.



**Fig. 5.** Five year average monthly wind speed (continuous line) and coefficient of variations (dotted lines) at the three Sahelian stations (Banizoumbou: red; Cinzana: blue; M'Bour: black); the vertical gray lines correspond to the limits between the dry and wet seasons. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

#### Table 2

Monthly median, 75th and 25th percentiles of the  $PM_{10}$  concentrations (in  $\mu$ g m<sup>-3</sup>) over the period 2006–2010, maximum and minimum monthly median concentrations of five years and year of occurrence of these maximum and minimum values for the three stations of the Sahelian Dust Transect. (Note that the statistics are computed on the decimal logarithm of the  $PM_{10}$  concentrations but are given here in terms of  $PM_{10}$  concentrations).

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Banizoumbou (Niger)												
Median 2006-2010	103.0	129.5	205.3	151.2	77.7	63.2	23.0	13.4	20.7	51.7	77.5	88.4
Perc. 75 2006-2010	205.6	296.3	450.6	299.5	129.9	117.6	46.5	20.4	35.0	94.9	136.8	182.4
Perc. 25 2006-2010	60.7	76.7	112.8	86.0	48.0	36.2	13.5	8.6	11.8	29.1	42.8	54.8
Max. monthly med.	330.7	265.1	415.5	211.1	93.9	100.2	43.4	15.1	29.0	85.8	104.9	124.8
Year of max	2007	2008	2010	2010	2010	2006	2006	2010	2008	2008	2007	2006
Min. monthly med.	58.0	82.5	109.0	87.2	65.7	38.4	18.8	10.4	15.4	33.4	63.8	62.5
Year of min.	2010	2010	2008	2009	2008	2008	2007	2007	2010	2010	2010	2009
Cinzana (Mali)												
Med. 2006-2010	95.3	127.1	164.5	110.3	67.6	49.6	24.2	11.5	16.8	43.5	70.0	72.9
Perc. 75 2006-2010	189.7	213.2	308.9	204.1	110.7	89.7	41.8	18.5	29.9	67.1	108.9	123.5
Perc. 25 2006-2010	60.8	83.3	94.4	71.3	45.2	31.9	15.3	7.5	9.5	25.9	46.9	45.2
Max. monthly med.	244.3	168.1	329.3	175.8	87.4	55.8	33.1	15.0	23.4	65.2	85.8	97.2
Year of max	2007	2008	2010	2010	2007	2007	2007	2006	2008	2008	2007	2008
Min. monthly med.	62.7	77.6	113.9	84.0	54.0	43.4	18.8	9.7	11.6	27.8	54.5	36.6
Year of min.	2006	2006	2009	2007	2008	2008	2008	2008	2010	2010	2009	2009
M'Bour (Senegal)												
Median 2006-2010	143.9	115.5	126.4	92.7	85.5	63.3	53.4	35.3	31.9	63.8	87.9	92.8
Perc. 75 2006-2010	237.3	195.2	221.9	140.3	129.4	83.4	72.8	53.8	53.3	92.8	132.0	141.7
Perc. 25 2006-2010	87.3	64.4	72.5	62.5	58.1	47.1	34.1	21.5	18.1	40.3	57.0	57.0
Max. monthly med.	221.6	179.0	172.5	127.0	99.4	68.5	60.5	43.7	45.2	84.2	107.5	128.1
Year of max	2007	2007	2007	2008	2007	2010	2007	2007	2008	2008	2009	2008
Min. monthly med.	105.9	91.0	95.5	78.1	70.1	60.6	49.8	26.2	23.8	55.4	70.0	58.5
Year of min.	2010	2006	2009	2010	2006	2008	2009	2009	2009	2009	2006	2010

The seasonal cycle of PM<sub>10</sub> concentrations is not phased with the seasonal cycle of the wind speed. The strongest PM<sub>10</sub> concentrations are recorded in March, while the most intense winds are recorded at the beginning of the rainy season, in June. This difference shows that the local wind is not responsible for local dust emissions that could explain the maximum  $PM_{10}$  of each station. This supports the idea that the dust transported by the wind from the Sahara is responsible for these maximum PM<sub>10</sub> concentrations observed at the three stations (Marticorena et al., 2010). Assuming a minimal erosion threshold of 6 m s<sup>-1</sup> at 10 m, the frequency of wind exceeding this threshold can be roughly estimated for the different stations, by scaling this value assuming a logarithmic wind profile and an aerodynamic roughness length typical of a Sahelian fallow ( $Z_0 = 0.05$  m; Bielders et al. (2004)). This estimation has been done both for the hourly mean and the hourly maximum wind speed. In M'Bour, hourly maximum wind speeds higher than 6 m s<sup>-1</sup> are recorded 5% of the time during the 5 years of measurements, and 6% when restricting the estimation to the five wet seasons. The frequency of mean wind speed higher than this threshold is much lower, less than 1%, but slightly higher during the wet season. For the station of Cinzana, where the wind speed is measured at 2.3 m high, a 10 m threshold of 6 m s<sup>-1</sup> corresponds to a wind speed of about  $4 \text{ m s}^{-1}$ . A threshold of  $4 \text{ m s}^{-1}$  is exceeded by the hourly maximum wind speed 9% of the time over five years and 11% during the wet seasons (and 3% by the hourly mean wind speed over five years, and 4% for the wet seasons). For Banizoumbou, the threshold at 6.3 m high equivalent to a threshold of 6 m s<sup>-1</sup> at 10 m high is 5.5 m s<sup>-1</sup>. The frequency of wind speed higher than 5 m s<sup>-1</sup>, the erosion threshold, is much higher than at the two other stations: it is 10% over five years for the mean hourly wind speed and 21% for the maximum hourly wind speed. It is slightly higher in the dry season (10% for the hourly mean and 23% for the hourly maximum). Consistently, a maximum of the wind speed (June) is associated with a secondary maximum of the PM<sub>10</sub> concentrations in Banizoumbou and with a slight increase of the concentrations in Cinzana. This maximum in the wind speed is due to the very high wind speeds in the gust front or in the cold pools generated by convective systems that develop in the Sahel at the beginning of the wet season. These high surface winds produce local dust emissions responsible for extremely high dust concentrations (Marticorena et al., 2010). Indeed, at Banizoumbou, the number of hourly mean wind speeds higher than 8 m s<sup>-1</sup> is much higher than at the two other stations (135 instead of 2 and 3) and significantly higher in the wet season (94) than in the dry season (41). When progressing into the wet season, the wind speed associated with convective systems decreases, precipitations increase, inhibiting dust emissions and washing out dust in suspension (Marticorena et al., 2010). This explains the minimum dust concentrations recorded in the core of the wet season. At M'Bour, the seasonal cycle of wind speed appears different from the two others stations. The maximum wind speed is observed in the dry season in February–March, like the maximum of the PM<sub>10</sub> concentrations. On the opposite, the secondary maximum of the PM<sub>10</sub> concentrations.

#### 4.3. Precipitation

The five-year average monthly precipitation at the three stations is reported in Fig. 6. The highest annual precipitation is recorded in Cinzana (769  $\pm$  103 mm) with a relatively low standard deviation. The annual distribution in Banizoumbou is very similar but annual precipitation are lower (473 mm) with a higher standard deviation (145 mm, i.e. 31% of the mean). The annual precipitation in M'Bour is



Fig. 6. Five year average monthly precipitation at the three Sahelian stations (Banizoumbou: dark gray; Cinzana: black; M'Bour: light gray).

higher (601 mm) but with a comparable inter annual variability (standard deviation: 186 mm, i.e. 31% of the annual mean). Compared to the three years investigated by Marticorena et al. (2010) (2006-2008), the years 2009 and 2010 are among the driest years in Banizoumbou, the wettest in M'Bour (Table 1). As typically observed in the Sahel, precipitation is distributed over a few month period: 90% of the annual precipitation occurs between June and September. As already stated by Marticorena et al. (2010), the distribution of the maximum of precipitation in M'Bour appears as slightly shifted to September compared to the two other stations. The precipitation is September (229 mm) is comparable to the precipitation in August (237 mm) in M'Bour, but much lower than in August at the two other stations (76 mm against 188 mm in Banizoumbou; 166 mm against 284 mm in Cinzana). In addition, the fraction of precipitation recorded before July is of the order of 2–3% in M'Bour but of the order of 15% at the two other stations. The monthly occurrence of precipitation, computed as the number of 5-min data with precipitation recorded divided by the total number of data was found linearly correlated with the monthly precipitation ( $r^2 = 0.92$ ) with similar slopes at the three stations (0.0128 to 0.014). As a result, the monthly occurrence and the precipitation amount exhibit the same seasonal pattern, with the same differences between the stations. The lowest PM<sub>10</sub> concentrations are recorded at the period where precipitation is the highest. The minimum concentrations coincide with the maximum of precipitation: in August in Banizoumbou and Cinzana (except in 2009) and in August (in 2008 and 2009) or September in M'Bour. As already suggested by Marticorena et al. (2010) this difference in the distribution of the precipitation during the wet season, and in particular in the timing of the maximum of precipitation may thus explain the differences in the timing of the minimum concentration recorded in M'Bour. Despite a relatively large variability in the annual precipitation, it is not possible to establish a clear link between precipitation and dust concentrations. However, it can be noticed that in Banizoumbou and Cinzana, the highest monthly median concentrations are recorded during the dry season following the wet season 2009, i.e. the driest year in Cinzana and the second driest year in Banizoumbou. Similarly, in M'Bour the driest wet season (2007) is followed by monthly median concentrations in the spring among the highest recorded during the five year period. Regarding the variability of both the dust concentration and the precipitation, longer time series are required to detect a relationship between these two parameters.

### 5. Diurnal variability

The diurnal cycle of mineral dust in the Sahel is not well documented due to a lack of observations throughout the day. Indeed, most studies that have addressed the diurnal cycle of mineral dust have used synoptic visibility data available with a 3 h or 6 h time step (e.g. N'Tchayi M'Bourou et al., 1994). In addition to the low sapling rate, the analysis of the diurnal cycle in the Sahelian meteorological stations is limited by fewer night-time observations, introducing large random errors in the night-time values at these stations (Cowie et al., 2014). Aerosol optical depths derived from sunphotometer measurements provide information on the atmospheric dust content, but only during daytime. Our data has an hourly resolution and a very good sampling rate over the five years. It thus offers a unique opportunity to document the diurnal cycle of the dust atmospheric load over the Sahelian region. The purpose of this section is to investigate the diurnal variability of the PM<sub>10</sub> concentrations at the three stations and to examine the role of the local meteorological conditions on this diurnal cycle.

### 5.1. Monthly diurnal cycles

Fig. 7 represents the monthly variations of the hourly mean wind speed, the hourly mean  $Log_{10}$  of the  $PM_{10}$  concentration and the hourly precipitation at the three stations of the Sahelian Dust Transect. The hourly mean wind speeds and  $PM_{10}$  concentrations appear as very contrasted depending on the season and the hour of the day. From



Fig. 7. Monthly mean diurnal cycle of wind speed (m s<sup>-1</sup>), decimal logarithm of the PM<sub>10</sub> concentration (µg m<sup>-3</sup>) and precipitation (mm) in Banizoumbou (a, b, c), Cinzana (d, e, f) and M'Bour (g, h, i) over the period 2006–2010.

November to July, the wind speed exhibits a clear maximum in the morning (~10 h) at the three stations. It is particularly intense and persistent in Banizoumbou, the maximum wind speed being recorded in March. At this station, a second period of high wind is observed during the night (22 h-6 h) from May to July, with a maximum in June between 1 h and 2 h in the morning. The monthly evolution of the diurnal cycle of the PM<sub>10</sub> concentrations has some similarities with the one of the wind speeds. During the dry season, a maximum of the PM<sub>10</sub> concentrations is observed in the morning, coinciding with the maximum of the wind speed. Like for the wind speed, the highest PM<sub>10</sub> concentrations are observed in March. The lowest concentrations are observed from July to October, when the wind speeds are the lowest. It also corresponds to the maximum of precipitation. At the beginning of the wet season (May to June) in Banizoumbou, several spots of very high concentrations are recorded between 20 h and 2 h in the morning in correspondence with a secondary maximum of wind speed. Additional but less intense high concentration spots are also recorded in the early morning (5 h) and during daytime (13 h, 16 h). In Cinzana the wind speed and PM<sub>10</sub> concentration patterns are similar than in Banizoumbou but less contrasted. The morning wind speed maximum (~10 h) is persistent during the dry season, but less intense than in Banizoumbou. Consistently, the morning maximum of the PM<sub>10</sub> concentration observed from December to April is also less intense than in Banizoumbou. A secondary maximum of wind speed is observed in June from 22 h to 2 h and a spot of high concentrations is also observed between 21 h and 22 h at this period. In M'Bour the morning maximum in wind speed and PM<sub>10</sub> concentration is still visible but much lower than at the two other stations. It corresponds to relatively high PM<sub>10</sub> concentrations from November to April. At this station, no nocturnal high winds or high PM<sub>10</sub> concentrations are recorded in the wet season. A secondary maximum in the wind speed is observed around 16 h from April to July that is not associated with any increase in the PM<sub>10</sub> concentrations. High PM<sub>10</sub> concentrations are also recorded in the late afternoon, around 18 h, especially in March.

These monthly diurnal cycles show that the dry and wet seasons are characterized by very different diurnal cycles of the PM<sub>10</sub> concentrations and surface wind speed. In the following, they are examined separately.

#### 5.2. Dry season

Fig. 8 presents the average (2006–2010) diurnal cycle of the wind speed and the diurnal cycle corresponding to the median 2006–2010  $PM_{10}$  concentration and Fig. 9 the average diurnal occurrence of wind directions at the three stations in the dry season.

A similar and well-marked diurnal cycle of wind speed is observed in the dry season at the three stations. The wind speed starts to increase sharply around 7 h to reach a maximum at 10 h. After the maximum, the wind speed slowly decreases and this decrease becomes sharper after 16 h in Banizoumbou and Cinzana. These high daily wind speeds are associated with dominant wind directions from east and northeast (Fig. 8), i.e. the typical direction of the Harmattan flow. From 19 h to 6 h in the morning, the wind speeds are very low ( $<2 \text{ m s}^{-1}$ ) and the wind direction is distributed in almost all sectors. This is probably due to the fact that wind direction measurement is not precise when the wind speed is low. In M'Bour, the decrease of the wind speed at the end of the afternoon is not as marked as at the two other stations. There is a progressive change in the wind direction from noon until the end of the afternoon (19 h), with an increasing proportion of wind directions from the sea (270-360°). This change in the dominant wind direction suggests the progressive establishment of a sea breeze. The standard deviations of the hourly surface wind speed are low and almost constant in Cinzana (0.6) and M'Bour (0.8) and slightly higher in Banizoumbou, up to 1.8 during the day, at the maximum of the wind speed. As a result, the coefficients of variations have a diurnal cycle opposite to the diurnal cycle of the wind speeds: the high wind



**Fig. 8.** Five year average diurnal cycle of the wind speed (a) and of the  $PM_{10}$  concentrations (b) corresponding to the median of the  $Log_{10}$  of the concentration (b) during the dry season (November-April; 2006–2010) at the three Sahelian stations.

speeds of the day-time are associated with low CVs ( $\sim$ 0.25) and the low wind speeds of the night-time are associated with high CVs (0.3–0.4 in Cinzana, 0.5–0.7 in Banizoumbou,  $\sim$ 0.75 in M'Bour).

The daily evolution of the median PM<sub>10</sub> concentrations in the dry season (Fig. 8b) has some similarities with the daily evolution of the wind speed (Fig. 8a). It is characterized by higher concentrations during the day than during the night. At the three stations, the morning increase of the wind speed coincides with an increase in the surface concentration. This maximum concentration is reached at 10 h in Banizoumbou and Cinzana and at 9 h in M'Bour, with a maximum concentration of 159  $\mu$ g m<sup>-3</sup> in Banizoumbou, 135  $\mu$ g m<sup>-3</sup> in M'Bour and 106  $\mu$ g m<sup>-3</sup> in Cinzana. A second increase of the concentration occurs at the three stations between 16 h and 22 h, with higher concentrations than in the morning in M'Bour (153  $\mu$ g m<sup>-3</sup>) and Cinzana  $(128 \ \mu g \ m^{-3})$ . In M'Bour, this evening maximum is observed when the dominant wind direction corresponds to the sea breeze. In such conditions, sea-salt aerosol can be transported to the station producing high PM<sub>10</sub> concentrations. In Cinzana, the evening maximum occurs when the wind speed is the lowest but corresponds to a slight reinforcement of the south-east, south and south-westerly wind sectors that can correspond to local sources of aerosols. During the night, from 22 h to 6 h in the morning, the hourly median concentrations are of the order of 100  $\mu g\,m^{-3}$  in Banizoumbou, 90  $\mu g\,m^{-3}$  in Cinzana and 80 in M'Bour, with night-time concentrations slightly larger in M'Bour (81 to  $107 \,\mu g \,m^{-3}$ ) that at the other stations (69 to 102  $\mu g \,m^{-3}$  in Cinzana and 96 to 109  $\mu$ g m<sup>-3</sup> in Banizoumbou). Surprisingly, during the day, the hourly median concentrations are higher in M'Bour than in Cinzana. However, the standard deviation of the Log<sub>10</sub> of the concentration, the CV and the range of the hourly concentration are higher in Cinzana than in M'Bour. In M'Bour, the IQR varies from 94 to 123  $\mu g \; m^{-3}$  but in Cinzana it varies from 102 to 160  $\mu$ g m<sup>-3</sup>. Similarly the range between the accepted maximum and minimum does not exceed 933  $\mu$ g m<sup>-3</sup> in M'Bour  $\mu$ g m<sup>-3</sup> but it reaches 1543  $\mu$ g m<sup>-3</sup> in Cinzana. At both stations, the largest range is associated with the morning


Fig. 9. Five year average diurnal cycle of the frequency of hourly wind directions in (a) Banizoumbou, (b) Cinzana and (c) M'Bour during the dry season (November-April; 2006 to 2010).

maximum. The range of concentration is even larger in Banizoumbou where the difference IQR ranges from 146 to 281  $\mu$ g m<sup>-3</sup> and the range between the maximum and the minimum from 1083 to 3652  $\mu$ g m<sup>-3</sup>. In Banizoumbou the IQR is linearly correlated with the median value, in particular for the day-time concentrations (slope: 1.6, r<sup>2</sup> = 0.83). With median concentrations from 97 to 159  $\mu$ g m<sup>-3</sup>, the IQR varies from 146 to 281  $\mu$ g m<sup>-3</sup> This relationship is not observed in Cinzana and in M'Bour. In M'Bour, it is almost independent from the median value and ranges from 93 to 123  $\mu$ g m<sup>-3</sup>. At the three stations, the lowest CVs are associated with the evening maximum of concentration (16 h–22 h), and the highest CVs (>0.2) are recorded during the night (0 h–6 h).

When the hourly median concentrations are normalized to the cumulated daily concentration, the monthly diurnal cycles exhibit the same pattern whatever the month during the dry season. This is the case at the three stations and illustrated in Fig. 10 for the station of Banizoumbou. In Cinzana and M'Bour, the shape of the normalized



Fig. 10. Five year monthly average diurnal cycle of the normalized median hourly  $PM_{10}$  concentrations in the dry season (November to April, 2006–2010) in Banizoumbou (Niger).

diurnal profile is similar all along the season, but with variations in the relative intensity of the morning and evening maxima. The morning maximum is higher than the evening maximum from January to March in Cinzana and slightly earlier in the dry season in M'Bour (December, January, February). This suggests that the diurnal cycle of the PM<sub>10</sub> concentration is modulated by local meteorological processes, while the daily concentration is controlled by synoptic factors that determine the intensity of the dust transport events.

The day-time maximum of the PM<sub>10</sub> concentrations observed at the three stations is in agreement with previous studies of the diurnal cycle of desert dust based on horizontal visibility measurements in the Sahel. Ozer (2001) showed that the occurrence of dusty episodes is maximum during the day, between 9 h and 15 h, and minimum at night-time, between 21 h and 3 h. N'Tchayi M'Bourou et al. (1994) also observed minimum occurrences of dust events at night and maximum occurrences between 9 h and 12 h at the Sahelian stations of Tombouctou (Mali) and Agadez (Niger). Cowie et al. (2014) noted a morning maximum of the dust event occurrence occurring at 12 h in the West Sahel and at 9 h in the Central Sahel, suggesting a strong influence of the NLLJ.

Indeed, the morning maximum of the PM<sub>10</sub> concentration is phased with the morning maximum of the wind speed. The diurnal cycle of the wind speed can be related to the presence of a LLJ over the Sahara and the Sahel. In the dry season, the NLLJ develops in the south of the Sahara and north of the Sahel (Fiedler et al., 2013). In Niamey, wind profiler and radio sounding measurements performed in 2006 revealed the persistence of a NLLJ all through the year, that comes from the north or the east in the dry season, and from the south or west during the wet season (Lothon et al., 2008). During the night, the NLLI is isolated from the surface by the nocturnal boundary layer. The day-time convective mixing of the high winds from the NLLJ also explains the morning maximum of the wind speed in the Sahel (Parker et al., 2005; Guichard et al., 2009). The morning maximum wind speed in Banizoumbou and Cinzana is actually associated with a reinforcement of easterly dominated wind directions (Fig. 9). However, at the latitude of Banizoumbou, vegetation cover is sufficient to prevent local erosion and thus dust emissions during the dry season (Abdourhamane Touré et al., 2011). As a result, the coincidence of the diurnal cycles of the wind speeds and of the PM<sub>10</sub> concentrations in Banizoumbou and Cinzana cannot be considered as due to local dust emissions (Rajot et al., 2008). On the opposite, these results suggest that the NLLJ is an efficient vector to transport dust to the Sahel, dust emitted further north or east in the Sahara. The downward convective mixing of the NLLJ in the morning brings both wind momentum and mineral dust to the surface level, causing the morning maximum of wind speed and PM<sub>10</sub> concentrations. After this morning maximum, the  $PM_{10}$  concentration decreases by dilution of this initial input or due to less efficient advection by lower wind speed in the boundary layer. Such a behavior was evidenced by lidar measurements performed in Banizoumbou on January 2006 (Heese and Wiegner, 2008). In particular, lidar depolarization measurements performed from 8 h until noon on a dusty day (15th January 2006) showed the presence of pronounced dust layer 0.5 and 0.9 km above a thin non-dusty surface layer. This dust layer progressively dispersed into the developing boundary layer by turbulent mixing during the morning. From about 10 h 30, the dust layer was well mixed and extended up to 1.2 km and a less pronounced dust layer with a constant vertical extent of 1.5 km remained visible throughout the morning. For this specific day (15 January 2006), the hourly PM<sub>10</sub> concentration measured in our data set started to increase at 7 h from less than 100  $\mu g~m^{-3}$  to reach a maximum of 567  $\mu g~m^{-3}$  at 11 h and progressively decrease from 11 h to 23 h to less than 200 µg m<sup>-1</sup>

The diurnal pattern observed on this specific day is similar to the normalized diurnal cycle of the  $PM_{10}$  concentrations over the five dry seasons 2006–2010. This suggests that the scenario of interaction between the NLLJ and the surface dust concentration is a typical feature of the Sahelian region.

#### 5.3. Wet season

The diurnal cycles of the wind speed and the  $PM_{10}$  concentrations (derived from the median  $Log_{10}$  of the  $PM_{10}$  concentrations) during the wet season are presented in Fig. 11 and the average diurnal occurrence of wind directions in Fig. 12. A similar diurnal cycle of the wind speed is observed at Banizoumbou and Cinzana, with a maximum around 10 h associated with wind directions typical of the monsoon flow (180–270°). Compared to the dry season (Fig. 8), the diurnal amplitude of the wind speed is reduced, with night-time wind speeds significantly higher than in the dry season. Guichard et al. (2009) showed that in Mali, the maximum of the surface wind speed observed in June was due to the higher night-time winds. According to these authors, this feature is valid until September beyond the overall weakening of the wind speed during the monsoon season.

This morning maximum is not observed in M'Bour where the maximum wind speed is measured in the afternoon (15–17 h), with wind directions from the sea (225–315°). In addition, the wind speeds recorded during the night are lower than in the dry season, unlike the two other stations. The diurnal cycle of the occurrences of wind directions in the rainy season (Fig. 11a) shows that the dominant wind direction in Banizoumbou and Cinzana is south-westerly, i.e. the typical direction of the monsoon flow, while in M'Bour, the dominant wind direction is from north-west to west, i.e. the direction of the sea. But in the morning and early afternoon, winds from the northeast have a significant contribution.

The diurnal cycle of the  $PM_{10}$  concentrations in the wet season (Fig. 11b) is very different from the one observed in the dry season and the hourly concentrations are much lower. The highest concentrations are observed in the late afternoon and during the night: from 16 h to 23 h in Cinzana and from 20 h to 2 h in Banizoumbou. The daily maximum median concentrations do not exceed 50  $\mu g m^{-3}$  in



**Fig. 11.** Five year average diurnal cycle of the wind speed (a) and of the  $PM_{10}$  concentrations (b) during the wet season (May–October; 2006–2010) at the three Sahelian stations.



Fig. 12. Five year average diurnal cycle of the frequency of hourly wind directions in (a) Banizoumbou, (b) Cinzana and (c) M'Bour during the wet season (May–October; 2006 to 2010).

Banizoumbou (47.8  $\mu$ g m<sup>-3</sup>) and Cinzana (44.4  $\mu$ g m<sup>-3</sup>). The minimum concentrations are of the order of 20 to 25  $\mu$ g m<sup>-3</sup> and are recorded between 1 h and 12 h in the morning. The diurnal cycle is different in M'Bour, with a maximum in the late afternoon (19 h) and a secondary maximum in the early morning (8 h). The morning maximum occurs at the time where the north-eastern winds significantly contribute to the hourly mean occurrence of wind direction. This may reflect the transport of mineral dust from the Sahel and the Sahara. On the opposite, the progressive increase of the PM<sub>10</sub> concentration that reaches a maximum at 19 h coincide to the reinforcement of the west to north-west wind sectors. This suggests that, like in the dry season, the sea breeze brings some sea-salt aerosol to the station.

At the three stations, the diurnal cycle of  $PM_{10}$  is not phased with the diurnal cycle of the wind speed. Like in the dry season, the morning peak in the wind speed and the reinforcement of the south-westerly direction can be associated with the progressive day-time mixing of the high winds of the Sahelian NLLJ. In the dry season, a NLLJ develops in the south of the Sahara and north of the Sahel but in the wet season,

a NLLI originates from southern Sahel (Fiedler et al., 2013). At this period of the year, no significant sources of aerosol (mineral dust or biomass burning) are active in the southern Sahel. Consequently, the morning maximum wind speed in the summer is not associated with high PM<sub>10</sub> concentrations. In Banizoumbou and Cinzana, the highest PM<sub>10</sub> concentrations are measured during the night. The surface wind speeds are lower in the night than during the day, but they are significantly higher than during the dry season. At the three stations, the variability of the wind speed is much more pronounced in the night-time and less pronounced in the morning (6 h to 12 h). The CVs of the Log<sub>10</sub> of the concentrations do not follow the same diurnal pattern. In Banizoumbou and Cinzana, high CVs (0.3) are observed between 10 h and 15 h, i.e. at the period of the day where the median concentrations are increasing to reach the evening maximum. High CVs are also observed during the night while the median hourly concentrations are relatively low (20–30  $\mu$ g m<sup>-3</sup>). In particular in Banizoumbou, the highest CV (0.35-0.4) are observed between 23 h and 3 h in the morning. In M'Bour, the CVs do not exhibit any clear seasonal cycle: they range between 0.17 and 0.2, the lowest values being associated with the evening maximum. In Banizoumbou and Cinzana, the range of the hourly concentrations is in the same order of magnitude that the one associated with the monthly concentrations: the IQR is linearly correlated with the median concentration, with a slope of 1.6 in Banizoumbou ( $r^2 = 0.86$ ) and 1.4 in Cinzana ( $r^2 = 0.97$ ). In M'Bour, the correlation is not as high  $(r^2 = 0.67)$  and the slope is lower than 1 (0.85). The correlation between the median concentration and the difference between the accepted maximum and minimum concentration is also lower ( $r^2 = 0.53$  in Banizoumbou, 0.76 in Cinzana and 0.32 in M'Bour). The low correlation in Banizoumbou is due to the fact that large ranges (>700  $\mu$ g m<sup>-3</sup>) are recorded during the night (23 h-3 h in the morning) when concentrations are relatively low  $(20-30 \ \mu g \ m^{-3})$ . The slope of the correlation is 18 in Banizoumbou and 15 in Cinzana, i.e. comparable to the range of the monthly values, but only 5 in M'Bour.

At the three stations, the highest concentrations are observed in May, with daily evening maximum of 114  $\mu$ g m<sup>-3</sup> in Banizoumbou, 94 in Cinzana and 118  $\mu$ g m<sup>-3</sup> in M'Bour. The lowest concentrations are recorded in August in Banizoumbou and Cinzana, with daily minimum of 9.6  $\mu$ g m<sup>-3</sup> and 8.8  $\mu$ g m<sup>-3</sup> and in September in M'Bour, with a minimum of 23  $\mu$ g m<sup>-3</sup>. However, like in the dry season, when normalizing the hourly mean concentrations by the daily cumulated concentration, the diurnal pattern is almost similar for all the months of the wet season (Fig. 13).

Convection over the Sahel develops during the wet season and exhibits a typical diurnal cycle with a maximum number of Mesoscale Convective Systems around 19 h and a maximum cloud cover between 20 h and 24 h (Mathon and Laurent, 2001). The diurnal cycle of the cloud frequency measured in Niamey from June to September 2006 shows that precipitating convection occurs almost exclusively in the evening (after 18 h) in June, due to the predominance of small-sized minimally propagating convective events (Bouniol et al., 2012). The diurnal cycle of the occurrence of convective clouds in June is thus consistent with the diurnal cycle of the PM<sub>10</sub> concentrations. Further in the season (July-August), the frequency of occurrence of convective clouds exhibits two peaks, one in the afternoon due to locally generated convection and the other in the morning associated with propagating systems initiated over the relief east of Niamey the day before (Bouniol et al., 2012). This timing of the occurrence of convective systems and its evolution during the wet season is consistent with the evolution of the diurnal cycle of the PM<sub>10</sub> concentrations measured in Banizoumbou. Local aeolian erosion and thus dust emissions have been effectively measured at the beginning of the wet season in Banizoumbou close to the SDT station (Abdourhamane Touré et al., 2011). These results highlight the role of the convective activity on the diurnal cycle of the dust concentrations in the wet season.



Fig. 13. Five year monthly average diurnal cycle of the  $PM_{10}$  concentrations normalized to the mean daily concentration in Banizoumbou (Niger).

#### 6. Conclusion

In the framework of the AMMA program, three stations (Banizoumbou (Niger), Cinzana (Mali) and M'Bour (Senegal)) dedicated to the monitoring of mineral dust in West Africa have been deployed in 2006. In this study, the seasonal and diurnal variability of the dust concentrations was analyzed based on a five year data set (2006–2010).

The analysis of the five year data set of  $PM_{10}$  concentrations recorded by the stations of the SDT reveals a log-normal distribution of the hourly concentrations. Statistics have thus been computed on a logarithmic base. The monthly median concentrations range from ten to a hundred  $\mu$ g m<sup>-3</sup> with IQRs and ranges between the maximum and the minimum that are proportional to the median values. The slopes range from 1.02 to 1.5 for the percentiles and from 6 to 14 for the maximum and minimum. The variability of the PM<sub>10</sub> concentration at the monthly scale is thus extremely large and it increase West to East.

This study confirms the seasonal pattern of the dust concentrations proposed by Marticorena et al. (2010). At the three stations the monthly concentrations are higher in the dry season, with maximum monthly concentrations ranging from 140 to 200  $\mu$ g m<sup>-3</sup>. They are not due to local dust emissions, since the 5-min surface wind velocities do not significantly exceed the erosion thresholds estimated for each station. In the dry season, the monthly median concentrations exhibit an increasing West to East gradient, consistent with the location of the stations compared to the main Saharan dust sources (Fig. 1), as suggested by Marticorena et al. (2010). The lowest monthly concentrations are recorded in the wet season and are phased with the maximum of the precipitation occurring in August in Banizoumbou and Cinzana and slightly later (September) in M'Bour. The monthly median concentrations in the wet season are twice higher in M'Bour (~30  $\mu g~m^{-3})$  than in Banizoumbou and Cinzana (~10  $\mu g m^{-3}$ ) despite the fact that annual precipitation is in the same order of magnitude than in Banizoumbou.

The variability of the monthly concentrations over the five years is very large and part of this variability is related to the year-to year variability. Extremely high or low monthly median concentrations are recorded at any period of any year. This is the case for the month of January 2007 during which the median and the range of concentrations is significantly larger than the other years at the three stations. The year 2009 appears as the less dusty year of the five year period at the three stations. It also among the driest year over the five year period in Cinzana and M'Bour. It is followed by a dry season where monthly median concentrations are among the highest measured during the period. However, regarding the variability of the concentrations and the precipitation over the five years, longer time series are required to establish a clear link between the precipitation regime and the atmospheric dust load.

For the first time, the diurnal cycle of the PM<sub>10</sub> concentrations measured in the Sahel is analyzed thanks to a relevant measurement time step and to an exceptional data recovery rates. Contrasted diurnal cycles are observed in the dry and in the wet season. During the dry season, the diurnal cycle of the PM<sub>10</sub> concentrations is phased with the diurnal cycle of the surface wind speeds. They are both characterized with a maximum around 10 h in the morning and low values during the night. In the Sahel, the diurnal cycle of the surface winds is strongly related to the dynamics of the NLLJ (Lothon et al., 2008; Guichard et al., 2009). Our results suggest that the Saharan NLLJ also modulates the diurnal cycle of the dust concentrations measured in the Sahel. The NLLJ appears as playing a significant role in the transport of mineral dust emitted from remote Saharan dust sources toward the Sahel. While the NLLJ controls the diurnal cycle of the PM<sub>10</sub> concentrations, the magnitude of the concentrations appears as controlled by the intensity of the dust transport events, which is determined by synoptic scale processes. During the wet season, the diurnal cycle of the PM<sub>10</sub> concentrations is characterized by a night-time maximum. The night-time concentrations are also characterized by a higher variability. They coincide with the period where convective systems are the most active in the Sahel.

They occur preferentially during the late evening at the beginning of the wet season (June) and the early morning or in the afternoon further in the season (July–August). The variability of the dust concentrations in the wet season appears as modulated by the evolution of the convective activity during this period and at the regional scale. However, local emissions by convective systems are short duration events that are not fully retrieved in the hourly data set used in this study. A future exploitation of the SDT measurements at their nominal temporal resolution (5 min) could be the investigation of the link between dust concentration and the convective activity during the wet season and their variations from year to year.

#### Acknowledgments

Visualization of the OMI aerosol index in Fig. 1 was produced with the Giovanni online data system, developed and maintained by the NASA GES DISC (Acker and Leptoukh, 2007). The authors sincerely acknowledge the local representations of the Institut de Recherche pour le Développement in Niger, Mali and Senegal for their administrative support and logistical help. The data from the SDT have been acquired in the Framework of the AMMA (African Monsoon Multidisciplinary Analysis) international program. Based on a French initiative, AMMA was built by an international scientific group and is currently funded by a large number of agencies, especially from France, U.K., U.S., and Africa. It has been the beneficiary of a major financial contribution from the European Community's Sixth Framework Research Programme.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx. doi.org/10.1016/j.atmosres.2015.05.011.

#### References

- Abdourhamane Touré, A., Rajot, J.L., Garba, Z., Marticorena, B., Petit, C., Sebag, D., 2011. Impact of very low crop residues cover on wind erosion in the Sahel. Catena 85, 205–214.
- Acker, J.G., Leptoukh, G., 2007. Online analysis enhances use of NASA Earth Science Data. Eos. Trans. AGU 88 (2), 14–17.
- Banta, R.M., Pichugina, Y.L., Brewer, W.A., 2006. Turbulent velocity variance profiles in the stable boundary layer generated by a nocturnal low-level jet. J. Atmos. Sci. 63, 2700–2719.
- Bielders, C.L., Rajot, J.L., Karlheinz, M., 2004. L'érosion éolienne dans le Sahel nigérien: influence des pratiques culturales actuelles et méthodes de lutte. Sécheresse 15, 19–32.
- Blackadar, A.K., 1957. Boundary layer wind maxima and their significance for the growth of nocturnal inversion. Bull. Am. Meteorol. Soc. 38 (5), 283–290.
- BouKaram, D., Flamant, C., Knippertz, P., Reitebuch, O., Pelon, J., Chong, M., Dabas, A., 2008. Dust emission over the Sahel associated with the West African monsoon intertropical discontinuity region: a representative case study. Q. J. R. Meteorol. Soc. 621–634 (134(632)).
- Bouniol, D., Couvreux, F., Kamsu-Tamo, P.-H., Leplay, M., Guichard, F., Favot, F., O'Connor, E.J., 2012. Diurnal and seasonal cycles of cloud occurrences, types, and radiative impact over West Africa. J. Appl. Meteorol. Climatol. 51 (3), 534–553.
- Cowie, S.M., Knippertz, P., Marsham, J.H., 2014. A climatology of dust emission events from northern Africa using long-term surface observations. Atmos. Chem. Phys. 14, 8579–8597. http://dx.doi.org/10.5194/acp-14-8579-2014.
- Deroubaix, A., Martiny, N., Chiapello, I., Marticorena, B., 2013. Suitability of OMI aerosol index to reflect mineral dust surface conditions: preliminary application for studying the link with meningitis epidemics in the Sahel. Remote Sens. Environ. 133, 116–127. Engelstaedter, S., Tegen, I., Washington, R., 2006. North African dust emissions and
- transport. Earth Sci. Rev. 79, 73–100. Evan, A., Heidinger, A.K., Knippertz, P., 2006. Analysis of winter dust activity of the coast of
- West Africa using a new 24-year over-water advanced very high-resolution radiometer satellite dust climatology. J. Geophys. Res. http://dx.doi.org/10.1029/2005jd006336, 111,d12, d12210.
- Fiedler, S., Schepanski, K., Heinold, B., Knippertz, P., Tegen, I., 2013. Climatology of nocturnal low-level jets over North Africa and implications for modeling mineral dust emission. J. Geophys. Res. 118. http://dx.doi.org/10.1002/jgrd.50394.
- Goudie, A.S., 1983. Dust storms in space and time. Prog. Phys. Geogr. 7, 502-530.
- Guichard, F., Kergoat, L., Mougin, E., Timouk, F., Baup, F., Hiernaux, P., Lavenu, F., 2009. Surface thermodynamics and radiative budget in the Sahelian Gourma: seasonal and diurnal cycles. J. Hydrol. 375, 161–177. http://dx.doi.org/10.1016/j.jhydrol.2008. 09.007.

- Heese, B., Wiegner, M., 2008. Vertical aerosol profiles from Raman polarization lidar observations during the dry season AMMA field campaign. J. Geophys. Res. http://dx.doi. org/10.1029/2007JD009487 113 (D00C11).
- Heinold, B., Knippertz, P., Marsham, J.H., Fiedler, S., Dixon, N.S., Schepanski, K., Laurent, B., Tegen, I., 2013. The role of deep convection and nocturnal low-level jets for dust emission in summertime West Africa: estimates from convection-permitting simulations. J. Geophys. Res. 118, 4385–4400. http://dx.doi.org/10.1002/jgrd.50402.
- Holben, B., Tanre, D., Smirnov, A., Eck, T., Slutsker, I., Abuhassan, I., Newcomb, W., Schafer, J., Chatenet, B., Lavenu, F., Kaufman, Y., Van de Castle, J., Setzer, A., Markham, A., Clark, D., Frouin, R., Halthore, R., Karnieli, A., O'Neill, N., Pietras, C., Pinker, R., Voss, K., Zibordi, G., 2001. An emerging ground-based aerosol climatology: aerosol optical depth from AERONET. J. Geophys. Res. 106, 12067–12098.
- Holton, J.R., 1967. The diurnal boundary layer wind oscillation above sloping terrain. Tellus 19 (2), 199–205.
- Huneeus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S., Baue, S., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Fillmore, D., Ghan, S., Ginoux, P., Grini, A., Horowitz, L., Koch, D., Krol, M.C., Landing, W., Liu, X., Mahowald, N., Miller, R., Morcrette, J.-J., Myhre, G., Penner, J., Perlwitz, J., Stier, P., Takemura, T., Zender, C.S., 2011. Global dust model intercomparison in Aerocom phase 1. Atmos. Chem. Phys. 11, 7781–7816.
- Jickells, T., An, Z., Andersen, K., Baker, A., Bergametti, G., Brooks, N., Cao, J., Boyd, P.W., Duce, R., Hunter, K., Kawahata, H., Kubilay, N., La Roche, J., Liss, P.S., M. N., Prospero, J., Ridgwell, A., Tegen, I., Torres, R., 2005. Global iron connections: between desert dust, ocean biogeochemistry and climate. Science 308 (5708), 67–71.
- Klose, M., Shao, Y., Karremann, M.K., Fink, A.H., 2010. Sahel dust zone and synoptic background. Geophys. Res. Lett. 37, L09802. http://dx.doi.org/10.1029/2010GL042816. Knippertz, P., Todd, M.C., 2012. Mineral dust aerosols over the Sahara: meteorological
- Knippertz, P., Todd, M.C., 2012. Mineral dust aerosols over the Sahara: meteorological controls on emission and transport and implications for modeling. Rev. Geophys. http://dx.doi.org/10.1029/2011RG000362.
- Lele, M.I., Lamb, P.J., 2010. Variability of the Intertropical Front (ITF) and rainfall over the West African Sudan–Sahel Zone. J. Clim. 23 (14), 3984–4004.
- Lothon, M., Saïd, F., Lohou, F., Campistron, B., 2008. Observation of the diurnal cycle in the low troposphere of West Africa. Mon. Weather Rev. 136 (9), 3477–3500. Marsham, J.H., Parker, D.J., Grams, C.M., Taylor, C.M., Haywood, J.M., 2008. Uplift of
- Marsham, J.H., Parker, D.J., Grams, C.M., Taylor, C.M., Haywood, J.M., 2008. Uplift of Saharan dust south of the intertropical discontinuity. J. Geophys. Res. 113, D21102. http://dx.doi.org/10.1029/2008JD009844.
- Marsham, J.H., Knippertz, P., Dixon, N.S., Parker, D.J., Lister, G.M.S., 2011. The importance of the representation of deep convection for modeled dust-generating winds over West Africa during summer. Geophys. Res. Lett. http://dx.doi.org/10.1029/ 2011gl048368, 38, 116803.
- Marsham, J., Hobby, M., Allen, C.J.B., Bart, M., Brooks, B., Cavazos-Guerra, C., Engelstaedter, S., Gascoyne, M., McQuaid, J., O'Learly, A., Ouchene, B., Ouladichir, A., Parker, D., Saci, A., Salah-Ferroudj, M., Todd, M., Washington, R., 2013. Meteorology and dust in the Central Sahara: observations from Fennec Superiste-1 during the June 2011 intensive observation period. J. Geophys. Res. 118 (1–21), 2013. http://dx.doi.org/10.1002/jgrd.50211. Marticorena, B., Chatenet, B., Rajot, J.L., Traoré, S., Coulibaly, M., Diallo, A., Koné, I., Maman,
- Marticorena, B., Chatenet, B., Rajot, J.L., Traoré, S., Coulibaly, M., Diallo, A., Koné, I., Maman, A., NDiaye, T., Zakou, A., 2010. Temporal variability of mineral dust content over West Africa: analyses of a pluriannual monitoring from the AMMA Sahelian Dust Transect. Atmos. Chem. Phys. 10, 8899–8915.
- Martiny, N., Chiapello, I., 2013. Assessments for the impact of mineral dust on the meningitis incidence in West Africa. Atmos. Environ. 70, 245–253.
- Mathon, V., Laurent, H., 2001. Life cycle of Sahelian mesoscale convective cloud systems. Q. J. Roy. Meteorol. Soc. 127 (572), 377–406. http://dx.doi.org/10.1002/qi.49712757208.
- Mathon, V., Laurent, H., Lebel, T., 2002. Mesoscale convective system rainfall in the Sahel. J. Appl. Meteorol. 41, 1081–1092. http://dx.doi.org/10.1175/1520-0450(2002)041<1081:MCSRIT>2.0.CO;2.
- Middleton, N.J., 1986. Dust storms in the Middle East. J. Arid Environ. 10, 83–96. Moulin, C., Lambert, C.E., Dulac, F., Dayan, U., 1997. Control of atmospheric export of dust
- from North Africa by the North Atlantic oscillation. Nature 387, 691–694. N'Tchayi M'Bourou, G., Bertrand, J., Legrand, M., Baudet, J., 1994. Temporal and spatial variations of the atmospheric dust loading throughout West Africa over the last thirty years. Ann. Geophys. 12, 265–273.
- Ogunjobi, K.O., He, Z., Simmer, C., 2008. Spectral aerosol optical properties from AERONET Sun-photometric measurements over West Africa. Atmos. Res. 88, 89–107.
- Ozer, P., 2001. Les lithométéores en région Sahélienne. Revue Internationale d'Ecologie et de Géographie Tropicales. 24 pp. 1–317.
- Parker, D.J., Burton, R., Diongue-Niang, A., Ellis, R.J., Felton, M., Taylor, C.M., Thorncroft, C.D., Bessemoulin, P., Tompkins, A.M., 2005. The diurnal cycle of the West African monsoon circulation. Q. J. Roy. Meteorol. Soc. 131, 2839–2860.
- Prospero, J.M., Ginoux, P., Torres, O., Nicholson, S.E., Gill, T.E., 2002. Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product. Rev. Geophys. 40 (1), 1–31.
- Rajot, J.L., Formenti, P., Alfaro, S., Desboeufs, K., Chevaillier, S., Chatenet, B., Gaudichet, A., Journet, E., Marticorena, B., Triquet, S., Maman, A., Mouget, N., Zakou, A., 2008. AMMA Dust Experiment: An Overview of Measurements Performed During the Dry Season Special Observation Period (SOP 0) at the Banizoumbou (Niger) Supersite. J G R113. http://dx.doi.org/10.1029/2008JD009906 (D00C14).
- Redelsperger, J.-L., Thorncroft, C.D., Diedhiou, A., Lebel, T., Parker, D.J., Polcher, J., 2006. African monsoon multidisciplinary analysis: an international research project and field campaign. Bull. Am. Meteorol. Soc. 87 (12), 1739. http://dx.doi.org/10.1175/ BAMS-87-12-1739.
- Schepanski, K., Tegen, I., Todd, M.C., Heinold, B., Bonisch, G., Laurent, B., Macke, A., 2009. Meteorological processes forcing Saharan dust emission inferred from MSG-SEVIRI observations of subdaily dust source activation and numerical models. J. Geophys. Res. 114, d10201. http://dx.doi.org/10.1029/2008jd010325.

- Slingo, A., Bharmal, N.A., Robinson, G.J., Settle, J.J., Allan, R.P., White, H.E., Lamb, P.J., Lele, M.I., Turner, D.D., McFarlane, S., Kassianov, E., Barnard, J., Flynn, C., Miller, M., 2008. Overview of observations from the RADACAST experiment in Niamey, Niger: meteorology and thermodynamic variables. J. Geophys. Res. 113, D00E01. http://dx.doi.org/ 10.1029/2008JD009909.
- Sokolik, I.N., Toon, O.B., 1999. Incorporation of mineralogical composition into models of the radiative properties of mineral aerosol from UV to IR wavelengths. J. Geophys. Res. 104, 9423–9444.
- Swap, R., Garstang, M., Greco, S., Talbot, R., Kållberg, P., 1992. Saharan dust in the Amazon basin. Tellus 44B, 133–149.
- Thomson, M.C., Molesworth, A.M., Djingarey, M.H., Yameogo, K.R., Belanger, F., Cuevas, L.E., 2006. Potential of environmental models to predict meningitis epidemics in Africa. Trop. Med. Int. Health 11 (6), 773–780.
- Africa. Trop. Med. Int. Health 11 (6), 773–780.
   Todd, M., Bou Karam, D., Cavazosand, C., Bouet, C., Heinold, B., Baldasano, J., Cautenet, G., Koren, I., Perez, C., Solmon, F., Tegenand, T., Tulet, P., Washington, R., Zakey, A., 2008.
   Quantifying uncertainty in estimates of mineral dust flux: an intercomparison of

model performance over the Bodélé depression, northern Chad. J. Geophys. Res. 113, d24107. http://dx.doi.org/10.1029/2008jd010476.

- Tompkins, A.M., Cardinali, C., Morcrette, J.-J., Rodwell, M., 2005. Influence of aerosol climatology on forecasts of the African easterly jet. Geophys. Res. Lett. 32, 11080. http://dx.doi.org/10.1029/2004gl022189.
- Viana, M., Querol, X., Alastuey, A., Cuevas, E., Rodriguez, S., 2002. Influence of African dust on the levels of atmospheric particulates in the Canary Islands air quality network. Atmos. Environ. 36 (38), 5861–5875.
- Washington, R., Todd, M., Lizcano, G., Tegen, I., Flamant, C., Koren, I., Ginoux, P., Engelstaeder, S., Bristow, C., Zender, C., Goudie, A., Warren, A., Prospero, J., 2006. Links between topography, wind, deflation, lakes and dust: the case of the Bodele depression, Chad. Geophys. Res. Lett. 33, I09401. http://dx.doi.org/10.1029/2006g1025827.
- Yahi, H., Marticorena, B., Thiria, S., Chatenet, B., Schmechtig, C., Rajot, J.L., Crepon, M., 2013. Statistical relationship between PM<sub>10</sub> surface concentration and aerosol optical depth over the Sahel as a function of weather types with the help of neural networks methods. J. Geophys. Res. http://dx.doi.org/10.1002/2013jd019465.

# 4.3 Analyse Comparative des cycles saisonniers des concentrations en $PM_{10}$ et de l'AOT

Pour compléter l'analyse précédente sur la saisonnalité des concentrations en  $PM_{10}$ , nous nous sommes intéressé au cycle saisonnier de l'épaisseur optique en aérosols. La Figure 4.1 présente le cycle saisonnier moyen de la concentration en  $PM_{10}$  et de l'AOT de 2006 à 2010 pour chaque station. Pour les stations de Banizoumbou et Cinzana, un cycle annuel apparaît clairement pour les deux paramètres, avec un maximum de concentration en  $PM_{10}$  en mars et un maximum d'AOT de mars à avril. Un pic secondaire est observé à Cinzana au mois de juin. Le cycle saisonnier de l'AOT de la station de Mbour apparait différent des autres stations avec trois pics, un au mois de mars, un autre en début de saison des pluies, au mois de juin, et un autre en fin de saison des pluies au mois de septembre. Le pic d'AOT du mois de juin a été mis en évidence par des travaux antérieurs, non seulement à la station de Mbour mais aussi au Cap-Vert [Léon et al., 2009; Oqunjobi et al., 2008]. Le pic du mois de mars est en cohérence avec celui observé aux deux autres stations et peut être associé au transport des poussières sahariennes mis en évidence par l'analyse des concentrations de surface [Marticorena et al., 2010]. Ce cycle saisonnier d'AOT a également été mis en évidence à partir d'observations satellitales. Kaufman et al. [2005] ont analysé les épaisseurs optiques dérivées des observations MODIS au dessus des lles du Cap Vert pour la période 2000 à 2003. Ils ont constaté la présence d'un cycle saisonnier régulier et bien marqué; les AOTs sont maximales en été, un autre maximum apparait au mois de mars et les valeurs les plus faibles sont observés en automne. Ce même cycle a été remarqué durant l'année 1987 à partir des épaisseurs optiques des aérosols issues de Météosat (Chiapello, 2006). Entre AOT et  $PM_{10}$ , nous observons une corrélation de 0.65 pour la station de Banizoumbou, 0.61 pour la station de Cinzana et de 0.40 pour la station de Mbour. Ce qui nous permet de dire que la relation  $AOT-PM_{10}$  est beaucoup plus forte sur les stations de Cinzana et de Banizoumbou qu'à Mbour. L'AOT renseigne le contenu atmosphérique en aérosols intégré verticalement et l'autre la teneur en aérosol en surface. Les différences de cycle saisonnier entre ces deux paramètres résultent de différences d'altitude de transport ou d'épaisseur des couches d'aérosols désertiques au cours de l'année.



Figure 4.1 – Cycle moyen saisonnier de la concentration en  $PM_{10}$  et de l'AOT de 2006 à 2010.

L'AOT est mesurée à plusieurs longueurs d'onde, ce qui permet d'accéder au coefficient d'Angström, $\alpha$ . Le coefficient d'Angström est sensible à la taille des particules : il est proche de 0 pour des particules grossières, telles que les aérosols désertiques, et généralement supérieur à 1 pour les particules fines constituant l'aérosol de pollution ou de feu. Ce coefficient est donc classiquement utilisé pour discriminer les différents types d'aérosols, en particulier en traçant le diagramme de dispersion entre l'AOT et $\alpha$ . La Figure 4.2 présente ce type de diagramme pour chaque station. Sur ces figures, nous pouvons remarquer que les AOT fortes sont associés aux  $\alpha$  faibles et qu'à l'inverse, les  $\alpha$  les plus élevés sont associés à des AOT faibles Cela suggère que les évènements de poussières ( $\alpha$  faibles) sont associés à de très fortes quantité d'aérosols alors que les aérosols de feux ou de pollution dont la présence est également détectée ( $\alpha$  forts) sont présent en trop faible quantité pour produire de fortes AOD. Ce comportement est typique des stations localisées dans des régions ou les aérosols désertiques dominent l'épaisseur optique en aérosols [*Dubovik et al.*, 2002; *Ogunjobi et al.*, 2008].



Figure 4.2 – Diagramme de dispersion des moyennes horaires d'AOT (440 nm) et du coefficient d'Angström

Le cycle annuel du coefficient d'Angström évolue en sens inverse de celui de l'AOT (Figure 4.3). Quand l'AOT est grande le coefficient d'Angström est petit On peut dire que les fortes AOT enregistrées à Banizoumbou et à Cinzana en saison sèche (Figure 4.3) et en juin à Mbour correspondant à de faibles coefficients d'angström sont liée à la présence de fortes quantité d'aérosols désertiques.



Figure 4.3 – Cycle saisonnier de l'AOT (pointillé) et du coefficient d'Angstrom (trait continu) pour chaque station (couleur rouge (Banizoumbou); bleu (Cinzana); noir (Mbour))

Afin de restreindre notre analyse à des situations clairement dominées par la présence d'aérosols désertiques, nous avons sélectionné les données en fonction de  $\alpha$ . Nous avons utilisé une valeur seuil de 0.4, appliquée aux moyennes horaires. Ce seuil a été utilisé dans des travaux de modélisation pour définir des conditions de comparaisons modèlesmesures limitées aux cas de poussières [Schmechtig et al., 2011] et dans des études d'impact sanitaire des aérosols désertiques [Deroubaix et al., 2013]. Après application de ce seuil, on a recalculé les moyennes mensuelles d'AOT et de concentration en  $PM_{10}$ . La Figure 4.4 présente le cycle saisonnier de l'AOT et de la concentration en  $PM_{10}$  après sélection en fonction du coefficient d'Angström. Les cycles saisonniers ne diffèrent pas radicalement de ceux obtenus sans sélection sur  $\alpha$ . On note cependant que les moyennes mensuelles de concentrations en hiver (décembre à février) sont plus fortes. Cela indique que lorsque l'on se limite aux évènements de poussières, les quantités de poussières présentes sont bien plus fortes que lorsque l'on intègre toutes les situations. La corrélation calculée entre AOT et  $PM_{10}$  pour les données correspondant à  $\alpha \ll 0.4$  reste plus élevée à Banizoumbou (0.67) et à Cinzana(0.60) qu'a Mbour (0.39).



Figure 4.4 – Cycle saisonnier de l'AOT et la concentration en  $PM_{10}$  pour Alpha  $\leq =0.4$ 

Cette étude nous a permis de voir que bien qu'il existe un cycle saisonnier aussi bien pour l'AOT et que pour la concentration en  $PM_{10}$ , avec un décalage temporel dans l'occurrence du pics qui est observé plus tôt pour la concentration de surface que pour l'AOT. Pour les stations de Cinzana et Banizoumbou, le maximum de la concentration en  $PM_{10}$ est enregistré au mois de mars alors que celui de l'AOT est obtenu au mois d'avril. De fortes AOT sont également observées en mars à Mbour. Ces pics correspondent à du transport d'aérosols sahariens dans les basses couches du flux d'harmattan. Le maximum d'AOT est observés en juin à Mbour, un pic secondaire est également enregistré à cette période à Cinzana, et selon la littérature, également au Cap-Vert. L'analyse des profils lidars mesurés à Mbour [Léon et al., 2009] montre que ce signal est lié au transport d'aérosols sahariens en altitude dans la Saharan Air Layer. Les différences de cycles saisonniers d'AOT et de concentration en  $PM_{10}$  sont donc vraisemblablement liées à des différences d'altitudes de transport ou d'épaisseur de couche d'aérosols. Comment les conditions météorologiques régionales influent ces caractéristiques du transport d'aérosols et donc les relations AOT- $PM_{10}$ , c'est une question, que nous allons aborder par le biais de des types de temps.

# Chapitre 5

# Relations AOT- $PM_{10}$ en fonction des régimes de temps d'Afrique de L'Ouest

# 5.1 Présentation de la méthodologie

## 5.1.1 Objectifs et stratégie

Dans le travail de Yahi et al. [2013], une analyse de la relation entre AOT et  $PM_{10}$  a été menée sur les stations de Banizoumbou et Cinzana durant la saison sèche (d'Octobre à Mai) en discriminant les situations météorologiques quotidiennes par une classification en 5 régimes de temps issues des données brutes de vent et de température sur 9 niveaux de pression. L'analyse de ces régimes de temps, de grande échelle, a montré qu'ils représentent différentes phases du cycle saisonnier durant la saison sèche. Yahi et al. ont montré ensuite que l'utilisation de cette information sur les régimes de temps permet de construire de meilleures régressions linéaires AOT- $PM_{10}$  pour une majorité de ces régimes et que la pente de ces régressions diffèrent suivant les régimes de temps. Dans une seconde étape, pour chaque régime de temps de grande échelle et à chaque station, l'information météorologique locale de la station permet d'améliorer la relation AOT- $PM_{10}$ .

L'approche développée dans ce chapitre vise à compléter le travail de Yahi et al. à plusieurs titres :

- La station de M'Bour, située le long de la petite côte, a été ajoutée à cette analyse sur les stations continentales de Cinzana et Banizoumbou. Ceci permettra de couvrir plus largement le transect sahélien et d'observer les contrastes possibles en longitude.
- Un nombre plus important de variables ont été inclus de façon à essayer de mieux caractériser les régimes de temps. D'une part nous avons ajouté la hauteur géopotentielle qui est apte à intégrer une grande part de l'information dérivée de la température et du vent sur la circulation de grande échelle. D'autre part nous avons intégré des variables décrivant mieux la composante thermodynamique de l'atmosphère, à savoir l'humidité spécifique, la vitesse verticale, et la hauteur de la couche limite. Ceci pour vérifier si ce type d'information est significatif ou pas dans la différentiation des régimes de temps.
- Nous avons aussi considéré dans les variables le coefficient d'Angstrom souvent utilisé pour caractériser les différents types d'aérosols, en particulier les poussières désertiques des aérosols de pollution.
- Du fait de ces deux types de variables, nous avons choisi de mettre en place et tester une méthode statistique spécifique de sélection de variables en classification multi bloc. Deux blocs ont été défini, un *bloc* constitué des variables de circulation à grande échelle et un autre *bloc* les variables thermodynamiques.

- Enfin nous avons développé la même approche que Yahi et al. mais sur deux échelles de temps différentes. En effet Yahi et al. notent que l'utilisation des données brutes (sans filtrage préalable) ne permet pas d'appréhender les origines d'évènements de soulèvement/transport de poussières importants mais sporadiques et d'échelle synoptique, qui sont observables sur le Sahel [Klose et al., 2010]. Nous avons donc appliqué cette approche, (i) aux données brutes comme dans Yahi et al.; (ii) aux données désaisonnalisées en éliminant le cycle saisonnier moyen calculé jour à jour, de manière à supprimer l'effet de saisonnalité qui domine largement dans les résultats de Yahi et al. et faire apparaître la variabilité résiduelle, particulièrement d'échelle interannuelle.
- Nous avons ainsi développé une approche similaire à celle de Yahi et al. pour ces deux échelles de temps. A noter que dans la partie (ii), nous avons aussi tenté de combiner les résultats des étapes (i) et (ii) pour évaluer la potentialité de mettre en place un système d'alerte statistique opérationnel aux évènements forts de poussière.

### 5.1.2 Présentation des données d'études

Pour cette étude nous avons utilisé deux types de données : les données météorologiques de grande échelle et les mesures in-situ. Les variables utilisées dans ce chapitre sont celles décrites au chapitre 2. Afin de rendre plus claire la lecture de ce chapitre, nous re-décrivons succinctement les différentes données utilisées. On rappelle également que l'étude porte sur la saison sèche (octobre à mai).

Les mesures in-situ utilisées sont :

- Les moyennes horaires des concentrations en  $PM_{10}$  mesurées avec un pas de temps de 5 min sur les trois stations d'Afrique de l'Ouest.
- Les moyennes horaires des mesures météorologiques de vitesse et direction de vent, température de l'air, humidité relative enregistrées avec un pas de temps de 5 min.
- Les moyennes horaires d'épaisseur optique en aérosols réalisées par les photomètres solaires AERONET/PHOTONS toutes les 15 min de jour et par ciel clair.

Concernant les données à grande échelle, nous avons 7 variables météorologiques c'està-dire la température (T), le vent zonal (U), le vent méridien (V), le géopotentiel (Z), l'humidité spécifique (Q), la vitesse verticale (W), ainsi que la variable hauteur de la couche limite (BLH) données sur 9 niveaux verticaux, à l'exception de la hauteur de la couche limite. Ces variables sont renseignées toutes les 3h les dix sept points de grilles entourant chaque station (cf Figure 2.2). Cela représente une base de données de dimension (9727 \* 935). Les 9727 observations représentent les données tri-horaires sur les saisons sèches (octobre à mai) des cinq années de mesure (2006-2010). Chaque observation de la base est donc associée à un vecteur de dimension  $(918 = 17 \times 9 \times 6 + 17)$ .

A partir de l'ensemble de ces données, la base de données de travail a été composée en ne retenant que les pas de temps pour lesquels on dispose des mesures coïncidentes des paramètres météorologiques locaux, de concentrations en  $PM_{10}$ , et d'AOT et des mesures météorologiques ECMWF.

### 5.1.3 Présentation de la méthodologie de filtrage

Dans ce chapitre, notre objectif consiste à déterminer et à caractériser des régimes de temps météorologiques en Afrique de l'Ouest en saison sèche. L'idée est de définir les types de temps qui pourraient avoir un impact sur la variabilité des aérosols pendant la saison sèche en Afrique de l'Ouest. Cette étude s'inspire de *Gueye et al.* [2010, 2012] qui ont défini les types de temps météorologiques pendant la mousson africaine au Sénégal et aux alentours à différentes échelles de temps. Ainsi nous avons défini dans ce travail deux types de régimes de temps en utilisant différentes approches. La première approche consiste à déterminer les régimes de temps en utilisant les données initiales sans aucun traitement préalable, ou régimes de temps bruts. L'idée ici est de déterminer les régimes de temps qui pourraient avoir un impact sur la variabilité saisonnière et interannuelle des aérosols pendant la saison sèche en Afrique de l'Ouest. Cette approche s'inspire des travaux de *Yahi et al.* [2013].

La seconde approche consiste à déterminer les régimes de temps en utilisant la base de données dont on a éliminé le cycle saisonnier moyen. L'objectif sera de déterminer les régimes de temps qui pourraient avoir un impact sur la variabilité interannuelle des aérosols.

Une analyse et une interprétation des régimes de temps du point de vue météorologique seront menées pour chaque approche ainsi qu'une analyse de la relation AOT- $PM_{10}$ .

# 5.1.4 Présentation de la méthodologie de classification par types de temps

Dans ce travail, nous avons structuré la base de données de chaque station en deux blocs de variables : l'un spécifique à la dynamique horizontale de l'atmosphère, il s'agit des paramètres T, U, V et Z composant un bloc de 612 variables et le deuxième correspond à l'instabilité verticale de l'atmosphère (Q, W, BLH) et est composé de 323 variables.

L'approche consiste à faire une classification des champs météorologiques ECMWFen types de temps c'est à dire regrouper les 9727 situations météorologiques en un nombre restreint de situations types. Pour ce faire, nous avons utilisé 2S - SOM (cf chapitre 3), une méthode de subspace clustering basée sur les cartes auto-organisatrices de Kohonen (SOM).

2S - SOM est un algorithme de classification multi-blocs basé sur la modification de l'algorithme SOM en introduisant un système de pondération à la fois sur les *blocs* et sur les variables permettant d'identifier les blocs et les variables les plus importants dans la classification.

La méthodologie développée dans cette thèse se résume à l'utilisation de 2S - SOM en deux étapes (voir Figure 5.1) sur les données de chacune des stations.

- La première étape consiste à appliquer 2S - SOM sur la base de données décrite cidessus, puis on applique la Classification Ascendante Hiérarchique (CAH) sur la carte obtenue. A partir des classes obtenues et des poids adaptatifs sur les blocs et les variables on identifie les variables et les blocs les plus pertinents dans la classification à l'aide des tests statistiques.



Figure 5.1 – Organigramme récapitulatif de l'approche développé dans ce travail

- La deuxième étape consiste à appliquer à nouveau 2S - SOM et la CAH sur les variables sélectionnées afin d'obtenir la partition finale qui représente les régimes de temps appelé WR (« weather regime »).

## 5.1.5 Sélection de Variables

La première étape de notre approche dans ce travail consiste à l'utilisation de 2S-SOM dans un but de sélection des variables. L'idée est d'identifier les variables les plus pertinentes, c'est à dire les variables qui ont le poids  $\beta$  le plus fort afin d'éviter la redondance. 2S - SOM permet de tenir compte de la dimension élevée des données, de la structure multi-blocs, de la présence de données manquantes ou aberrantes et de faciliter la comparaison des classes et la visualisation des données.

Ainsi nous avons formé pour chaque station une base de données répartie en deux blocs de variable caractérisant pour l'un la dynamique horizontale de l'atmosphère (T, U, V, Z) et pour l'autre l'instabilité verticale de l'atmosphère à travers les profils verticaux de vitesse verticale, d'humidité et la hauteur de la couche limite (Q, W, BLH). Pour chaque station, le bloc 1 a une dimension de [9727\*612] et le bloc 2 a une dimension de [9727\*323], soit une base de données de [9727\*935] par station. Plusieurs applications de 2S - SOM ont été réalisées en faisant varier les paramètres d'initialisation notamment les paramètres  $\lambda$  et  $\eta$  (voir chapitre 3), les dimensions et paramètres du voisinage de cellule de carte étant définis par le principe d'initialisation proposée par Vesanto (1999). La meilleure carte en termes de quantification vectorielle est retenue. A partir des poids adaptatifs sur les blocs et sur les variables de chaque bloc obtenus de la carte, on pourra identifier les variables et les blocs les plus importants dans la classification.

Ainsi pour chaque station nous avons obtenu une carte de [30\*16] neurones soit un total de 480 neurones, cette première carte est notée 2SSOM1. La représentation des variables n'a pas été faite à cause du nombre important de variables d'apprentissage, mais un travail préalable a été effectué pour s'assurer de leur bonne organisation avant de procéder à l'exploitation de la carte topologique de chacune des stations.

La Figure 5.2 ci dessous présente les poids Alpha de chaque bloc sur chaque nŒud de la carte dans chacune des trois stations. La majorité des poids tourne au tour de 0.5 et à part quelques neurones, le *bloc*1 domine légèrement le *bloc*2 sur les cellules de la carte dans les deux stations de Banizoumbou et de Cinzana. Pour la station de Mbour il n'y a pas de différence nette entre les blocs sauf pour quelques neurones.



Figure 5.2 – Représentation des poids Alpha de chaque bloc sur chaque cellule de la carte  $2S_SOM$ 1pour les stations de Banizoumbou, Cinzana et Mbour

Vu le grand nombre de neurones sur la carte obtenue, une CAH a été appliquée aux neurones de la carte de chaque station pour réduire le nombre de classes afin de mieux exploiter cette dernière. La Figure 5.3 présente la courbe des indices de Davies Bouldin et l'histogramme des niveaux d'agrégation pour 30 classes pour Banizoumbou, Cinzana et Mbour. L'indice de Davies Bouldin (DB) permet de mesurer la compacité et la séparabilité des groupes ou des classes, de petites valeurs de DB sont indicatives de la présence de groupes compacts et bien séparés. Afin de choisir le nombre de classes à retenir c'est à dire la meilleure classification, nous avons utilisé simultanément l'indice de Davies Bouldin (voir chapitre 3), le dendrogramme et l'histogramme des niveaux d'agrégation. Ainsi ces éléments de la Figure 5.3 nous ont permis de retenir 6 classes pour chaque station.



Figure 5.3 – Histogramme des niveaux d'agrégation pour 30 classes (a) et Indices de Davies Bouldin (b) pour les stations de Banizoumbou, Cinzana et Mbour.

Après avoir déterminé le nombre de classes, nous avons cherché à identifié les variables les plus importantes de chaque classe en s'appuyant sur les poids des variables. Pour chaque bloc, on considère qu'une variable a un poids significatif si sa valeur de poids est supérieure à 1/n, n étant le nombre total de variables dans le bloc (612 pour le *bloc* 1 et 323 pour le *bloc*2). Pour le *bloc*1, le géopotentiel (Z) est la variable la plus importante car ayant plus de poids fort dans chaque classe. Cependant le vent méridien (V) est la variable la moins importante car ayant peu de poids fort par rapport aux autres variables sur les trois stations. Pour le *bloc*2, l'humidité spécifique (Q) est la variable la plus importante tandis que la vitesse verticale (W) est la moins importante. Il faut signaler que la *BLH* ne compte que 17 variables. Ainsi nous pouvons dire que sur les deux stations de Banizoumbou et Cinzana, c'est le géopotentiel et l'humidité qui ont joué grand rôle dans la détermination des classes et c'est le vent méridien qui n'a pas joué un grand rôle dans la détermination des classes.

Pour la station de Mbour, les variables les plus importantes semblent être identiques à celles identifiées sur les autres stations à quelques différences près. L'humidité et le géopotentiel sont les variables les plus importantes à Mbour. Cependant l'humidité est un peu plus important et le géopotentiel un peu moins à Mbour que dans les autres stations. Les variables les moins importantes à Mbour sont le vent méridien et la vitesse verticale. Globalement, nous pouvons dire que c'est l'humidité qui a le plus joué dans la détermination des classes à la station de Mbour alors que pour les stations de Cinzana et Banizoumbou c'est le géopotentiel qui joue le plus.

Tableau 5.1 – Nombre de poids important de chaque variable de chaque bloc sur chaque classe pour les stations de Banizoumbou, Cinzana et Mbour.

Bani	T	U	V	Z	Q	W	BLH
Classe1	40	118	20	130	115	31	0
Classe2	32	136	34	149	76	51	16
Classe3	61	89	9	121	124	34	2
Classe4	49	72	13	148	101	14	0
Classe5	46	99	11	141	100	40	5
Classe 6	73	58	23	137	134	19	10
Cinza	T	U	V	Z	Q	W	BLH
Classe1	33	67	29	148	67	48	13
Classe2	48	92	0	101	116	0	0
Classe3	124	80	23	112	94	8	15
Classe4	73	57	0	126	116	5	3
Classe 5	35	64	21	150	80	45	9
Classe 6	46	62	0	150	104	2	0
Mbour	T	U	V	Z	Q	W	BLH
Classe1	81	104	0	105	134	5	8
Classe2	62	79	2	104	144	4	3
Classe3	82	95	0	102	118	17	9
Classe4	32	96	1	104	123	14	3
Classe5	104	106	0	99	126	15	10
Classe 6	109	83	13	95	141	23	5

Après avoir déterminé pour chaque classe les variables météorologiques les plus importantes (Tableau 5.2), nous avons pris l'union sans doublon de l'ensemble des variables significatives de chaque paramètre dans chaque classe de la classification précédente pour former une base, et cela pour chacune des stations (Tableau 5.2). Ainsi nous avons conservé la structure en deux blocs avec pour bloc1 les paramètres T, U, V et Z, et pour le bloc2 les paramètres Q, W et *BLH*. Concernant la station de Banizoumbou, pour le bloc 1, il y a 434 variables pertinentes sur 612 variables, soit environ 70% de variables pertinentes et 253variables sur 323, soit 78% de variables pertinentes pour le bloc2. Concernant la station de Cinzana, pour le bloc1, il y a 436 variables pertinente sur 612 soit 71 % de variables pertinentes et 214 variables sur 323, soit environ 66% de variables pertinentes pour le bloc 2. Pour la station de Mbour, on a 371 variables sur 612, soit 60% pour le bloc1, et 213 variables sur 323, soit 65% de variables pertinente pour le bloc 2. Globalement pour les deux blocs réunis de chaque station, on a 73% de variables pertinentes et 27% de variables non pertinentes pour la station de Banizoumbou, 69% de variables pertinentes et 31% de variables non pertinentes à la station de Cinzana, 62% de variables pertinentes et 38% de variables non pertinentes pour la station de Mbour.

	Bloc 1				Bloc 2				
	T	U	V	Z	Total	Q	W	BLH	Total
Banizoumbou	84	139	60	151	434	141	95	17	253
Cinzana	126	108	52	150	436	146	51	17	214
Mbour	126	123	16	106	371	153	49	11	213

Tableau 5.2 – Union des variables dominantes de chaque classe pour chacune des stations.

Ainsi, nous avons formé pour chaque station une base avec les variables issues de la réunion des variables des classes (Tableau 5.1). Pour la station de Banizoumbou, nous avons obtenu une base de données de 687 variables divisée en deux blocs, soit 434 variables pour le bloc 1 et 253 variables pour le bloc2. Pour Cinzana, nous avons obtenu une base de données de 650 variables divisée en deux blocs, soit 436 variables pour le bloc1 et 214 variables pour le bloc2. Pour la station de Mbour, nous avons obtenu une base de données de 584 variables divisée en deux blocs, soit 371 variables pour le bloc1 et 284 variables pour le bloc2.

Les variables obtenues dans ce processus de sélection de variables vont être utilisées dans la suite du travail afin de déterminer la partition finale.

# 5.2 Régimes de temps bruts

Les variables pertinentes obtenues dans la section précédente sont utilisées dans cette partie afin de déterminer les régimes de temps. 2S - SOM est à nouveau appliqué sur la nouvelle base de données de chaque station pour déterminer la partition finale.

### 5.2.1 Détermination du nombre de régimes de temps bruts

Après apprentissage, nous avons obtenu pour chaque station une carte 480 neurones appelée 2S - SOM2 La Figure 5.4 présente les poids de chaque *bloc* sur les neurones de la carte de chaque station obtenue sur les bases de données sélectionnées. Nous remarquons qu'en prenant l'union des variables les plus significatives dans chaque classe de la classification précédente pour faire une autre classification, que les poids varient aussi autour de 0.5. Nous constatons aussi que le *bloc*1 domine le *bloc*2 presque sur l'ensemble des nœuds de la carte topologique sur les trois stations. Pour la station de Mbour où il n'y avait pas de différence entre les deux blocs dans la première classification, on remarque ici une nette domination du *bloc*1, ce qui confirme que les variables supprimées n'apportaient pas vraiment d'informations.



Figure 5.4 – Représentation des poids Alpha de la carte 2S - SOM2 de chaque bloc sur chaque cellule de la carte pour les stations de (a : Banizoumbou , b :Cinzana et c : Mbour).

Pour déterminer le nombre de classes à retenir, on a procédé de la même manière que

précédemment, ce qui a permis de retenir 6 classes pour chaque station (voir Figure 5.5). La coupure présentée sur le dendrogramme de la Figure 5.5, permet de mieux comprendre le choix des six classes. Le trait horizontal rouge sur la figure montre effectivement que 6 classes peuvent être retenues sur cette classification. Un seul dendrogramme est présenté ici à titre d'exemple.



Figure 5.5 – Dendrogramme obtenu après la CAH des la base de données de la station de Banizoumbou.

Le Tableau 5.3 suivant présente pour chaque classe et pour chaque paramètre de chaque bloc les variables les plus importantes. C'est le géopotentiel, Z, qui est la variable la plus importante pour les stations de Cinzana et Banizoumbou alors qu'à Mbour c'est l'humidité specifique qui apparaît la plus importante dans la classification. Ainsi nous pouvons dire que pour le *bloc*1, la température (T), le vent zonal (U) et le géopotentiel (Z) sont les variables les plus importantes. Pour le *bloc*2, l'humidité spécifique est la variable la plus importante à la station de Mbour.

Bani	T	U	V	Z	Q	W	BLH
Classe1	54	51	1	117	94	8	10
Classe2	38	53	0	118	102	11	0
Classe3	53	11	0	127	90	15	17
Classe4	30	67	22	150	52	47	17
Classe5	33	60	9	139	96	18	10
Classe6	41	48	0	125	68	3	0
Cinzana	T	U	V	Z	Q	W	BLH
Classe1	33	51	0	115	62	3	8
Classe2	44	20	0	106	85	0	1
Classe3	30	45	0	127	68	0	0
Classe4	30	23	10	149	71	28	2
Classe5	29	27	14	141	63	26	5
Classe 6	31	0	0	145	97	0	0
Mbour	T	U	V	Z	Q	W	BLH
Classe1	58	52	0	82	114	1	8
Classe2	24	64	0	83	131	0	0
Classe3	85	66	0	75	92	2	4
Classe4	30	66	0	78	110	2	9
Classe5	59	38	0	80	96	4	5
Classe6	3	67	0	86	96	3	4

Tableau 5.3 – Nombre de variables significatives de chaque paramètre de bloc sur chaque classe pour les trois stations.

#### 5.2.2 Caractérisation des régimes de temps

A l'issu de ces traitements, 6 types de temps correspondant aux 6 classes ont donc été déterminés pour chaque station.

Le Tableau 5.4 suivant présente le pourcentage d'occurrence de ces régimes de temps sur l'ensemble des trois stations. Les régimes de temps les plus fréquents sont WR4 pour la station de Banizoumbou, WR5 pour la station de Cinzana et WR2 pour la station de Mbour. Les moins fréquents sont WR1 pour Banizoumbou et Mbour, et WR2 pour la station de Cinzana. Globalement, les régimes de temps des trois stations sont distribués de façon assez homogène. Le régime de temps le moins fréquent sur les trois stations a un pourcentage d'occurrence est d'environ 10% et le régime de temps le plus fréquent un pourcentage d'occurrence compris entre 20 et 24%.

L'équilibre des proportions entre les 6 types de temps de chaque station montre que les choix qui ont été faits (nombre de classe, critères de sélection des variables) étaient pertinents.

Tableau 5.4 – pourcentage d'occurrence de chaque type de temps pour les trois stations.

	WR1	WR2	WR3	WR4	WR 5	WR 6
Banizoumbou	$9{,}82\%$	18,77%	17,72%	$21,\!83\%$	16,79%	$15,\!07\%$
Cinzana	$17,\!53\%$	9,94%	18,59%	$18,\!39\%$	20,19%	15,36%
Mbour	10,50%	24,30%	$16,\!84\%$	$14,\!34\%$	$16,\!81\%$	17,21%

#### 5.2.2.1 Matrices de confusion

Les régimes de temps ont été déterminés de façon indépendante à chaque station, pour autant, on peut penser qu'elles puissent être affectées de façon similaire par des conditions météorologiques d'échelle synoptique. Nous avons donc quantifié le degré de confusion des types de temps des trois stations, qui nous renseigne ici sur leur degré de ressemblance an niveau régional (d'est en ouest). Les pourcentages de confusion des régimes de temps de la station de Banizoumbou et ceux de la station de Cinzana d'une part, et entre la station de Cinzana et Mbour d'autre part sont reportés dans Tableau 5.5. L'estimation globale de la concordance entre les classifications de Cinzana et les autres stations est très correcte (coefficient de Rand de l'ordre de 0.85 entre Banizoumbou et Cinzana et 0.8 entre Cinzana et Mbour).

Pour les stations de Banizoumbou /Cinzana, les maximums de données communes se trouvent sur la diagonale. Pour les stations de Mbour/Cinzana il existe une forte concordance de données sur la diagonale pour certaines classes, et plus de dispersion pour d'autres. Il y a donc plus de concordance entre les classes de Banizoumbou /Cinzana qu'entre les classes Mbour/Cinzana. Au vu de ces niveaux de concordances, nous avons désigné les types de temps par  $WR1 \cdots WR6$ , et le caractère B, C ou M désigne la station.

L'ensemble du processus de détermination des régimes de temps pour chaque station est résumé sur la Figure 5.1.

Tableau 5.5 – Pourcentage de confusion des classes de Banizoumbou et Cinzana (a) et des classes de Mbour et Cinzana (b)

a°)	WRB1	WRB2	WRB3	WRB4	WRB5	WRB6
WRC1	$66,\!18$	35,87	3,07	0,89	0,18	23,40
WRC2	$1,\!57$	$48,\!41$	0,75	0,05	0,00	$3,\!68$
WRC3	18,74	0,16	<b>56,84</b>	$13,\!90$	11,27	$11,\!39$
WRC4	$0,\!84$	$0,\!05$	$6,\!61$	<b>62,88</b>	$12,\!55$	$^{8,59}$
WRC5	$0,\!94$	0,00	19,72	17,62	$75,\!93$	$0,\!07$
WRC6	11,73	$15,\!50$	$12,\!99$	4,66	0,06	$52,\!86$
b°)	UUDM1		117 D 1 40	TTTDIA		IIIDIG
U)	WRM1	WRM2	WRM3	WRM4	WRM5	WRM6
WRC1	68,76	25,80	W RM 3 13,43	$\begin{array}{c} WRM4 \\ \hline 0,43 \end{array}$	$\begin{array}{r} WRM5 \\ 0,00 \end{array}$	<i>WRM6</i> 9,98
WRC1 WRC2	<b>68,76</b> 7,74	<i>WRM2</i> 25,80 <b>36,80</b>	$     \begin{array}{r} WRM3 \\     \hline             13,43 \\             0,12 \\         \end{array} $	$\begin{array}{r} WRM4\\ 0,43\\ 0,00 \end{array}$	$     \begin{array}{r} WRM5 \\             0,00 \\             0,00 \\           $	9,98 0,96
WRC1 WRC2 WRC3	<b>68,76</b> 7,74 10,58	W RM2           25,80           36,80           2,16	WRM3 13,43 0,12 13,00	$     \begin{array}{r} WRM4 \\             0,43 \\             0,00 \\             21,22 \\         \end{array} $	$     \begin{array}{r} WRM5 \\             0,00 \\             0,00 \\           $	$     \begin{array}{r} WRM6 \\             9,98 \\             0,96 \\             41,70 \\         \end{array} $
WRC1 WRC2 WRC3 WRC4	W RM1           68,76           7,74           10,58           4,21	W RM2           25,80           36,80           2,16           2,58	WRM3 13,43 0,12 <b>13,00</b> 40,23	W RM4           0,43           0,00           21,22           44,23	$\begin{array}{r} WRM5 \\ \hline 0,00 \\ 0,00 \\ 27,03 \\ 10,34 \end{array}$	WRM6      9,98     0,96     41,70     14,34
WRC1 WRC2 WRC3 WRC4 WRC5	W RM1           68,76           7,74           10,58           4,21           0,00	WRM2 25,80 36,80 2,16 2,58 0,00	WRM3 13,43 0,12 <b>13,00</b> 40,23 23,14	WRM4 0,43 0,00 21,22 <b>44,23</b> 30,75	WRM5           0,00           0,00           27,03           10,34 <b>62,20</b>	$\begin{array}{r} WRM6\\ \hline 9,98\\ 0,96\\ 41,70\\ 14,34\\ 8,30 \end{array}$

Les régimes de temps ayant été identifiés, nous allons les analyser en terme de variations temporelles et les interpréter en terme de situations météorologiques.

#### 5.2.2.2 Distribution diurne à inter-annuelle

L'analyse statistique des concentrations en  $PM_{10}$  nous a permis de mettre en évidence un cycle diurne de ces concentrations, phasé avec le cycle diurne de la vitesse de vent mesurée localement. Nous avons donc cherché à savoir si les régimes de temps étaient affectés par des processus diurnes. Les figures présentant le cycle diurne d'occurrence des régimes de temps sur les trois stations sont données en annexe (Figure A.1). Elles montrent clairement qu'il n'y a pas de tendance apparente sur le cycle diurne. A M'Bour, la répartition des régimes de temps est quasi-identique tout au long de la journée. Pour les deux autres stations, on note des variations mais elles n'apparaissent pas clairement organisées. Ceci nous emmène à dire que le cycle diurne n'a pas une influence notable sur les types de temps. Ce constat est confirmé par l'examen des cycles diurnes mensuels en annexe (Figure A.2) des différents régimes de temps.

A l'inverse, des tendances se dégagent clairement à l'échelle saisonnière (Figure 5.6). On note, en particulier des régimes de temps enregistrés en majorité pendant les mois où le régime d'harmattan domine très clairement (novembre à mars (WR3, 4, 5 pour la station de Banizoumbou; WR3, 4, 5 pour la station de Cinzana; WR4, 5et6 pour la station de Mbour) et des régimes de temps, WR1 et WR2, associés aux périodes de transition entre la saison sèche et la saison des pluies (avril, mai, octobre) pour l'ensemble des trois stations. Les régimes de temps WRB6, WRC6 et WRM3 correspondent en majorité à des mois d'harmattan (janvier, mars, avril) mais avec une contribution non négligeable des mois de transition entre la saison des pluies et la saison sèche (avril et octobre pour Banizoumbou, avril et mai pour Cinzana, octobre pour Mbour). Le type de temps WR3est enregistré à plus de 60 % entre janvier et mars, dont environ 32% en mars pour les stations de Cinzana et Banizoumbou. Pour l'ensemble des trois stations plus de 92 % de WR4 et WR5 est enregistré entre novembre et février). C'est également le cas de WRB3. Pour l'ensemble des trois stations, la majorité des données du régime de temps WR6 est enregistrée entre Février et Avril avec 59% à Banizoumbou, 85% à Cinzana et 72% à Mbour.

Les types de temps WR1 et WR2 sont typiquement des types de temps de fin et début de saison sèche, avec par exemple 56% de WRB2 en mai et 54% de WRB1 en octobre, 65.5% de WRC2 en mai et 62.5% de WRC1 en octobre et 49% de WRM2 en mai et 74% de WRM1 en octobre.

On peut conclure de ces résultats, que les régimes de temps dominants se succèdent au cours du temps dans l'ordre suivant : WR2, WR1, WR6, WR4, WR5 et WR3.



Figure 5.6 – Cycle saisonnier d'occurrence des types de temps en saison sèche pour les trois stations (a : Banizoumbou; b : Cinzan; c : Mbour).

La Figure 5.7 présente le cycle d'occurrence mensuelle des régimes de temps année par année pour les cinq années étudiées et sur chaque station. Les principales caractéristiques du cycle saisonnier établies à partir de la Figure 5.6 sont encore visibles. Cependant, cette Figure 5.7 met en évidence une forte variabilité interannuelle. La distribution des types de temps est différèrente d'une année à l'autre, et certaines anomalies peuvent être identifiées soit dans la fréquence relative d'un type de temps donné soit en termes de succession des différents types de temps sur la période étudiée.

La série temporelle des occurrences des régimes de temps identifiées dans chaque station confirme la prédominance de chaque régime de temps en saison sèche sur la période comprise entre octobre et mai comme on l'a déjà montré dans la figure 5.7.

Pour les types de temps caractéristiques de l'Harmattan, la variabilité interannuelle est très marquée sur l'ensemble des trois stations et semblent évoluer de la même façon à Banizoumbou et Cinzana, et différemment de Mbour. WR4 apparaît être le type de temps le plus important de Novembre à Février pour les stations de Banizoumbou et Cinzana mais avec des exceptions notables. Entre décembre 2007 et février 2008 et entre décembre 2008 et janvier 2009 c'est le régime de temps WR5 qui domine. On observe également une prédominance du régime de temps WR3 entre janvier et mars des années 2006 et 2009 et du régime de temps WR6 en mars 2010.

Pour la station de Mbour la prédominance du régime de WR4 entre Novembre et Février est moins intense par rapport aux deux autres stations. La période entre Novembre et Février de chaque année semble être simultanément dominée par les régimes de temps WR3 - WR4 - WR5 avec quelques exceptions. En février 2010, c'est le régime de temps WR3 qui domine. Entre janvier et mars des années 2006 et 2009, c'est le régime de temps WR5 qui domine.

	Octobre	Novembre	Decembr	e Janvier	Fevrier	Mars	Avril	Mai
					0	0	-	0
				2006	2006	2006	2006	2006
	0	0	0	0	0		0	0
	2006	2006	2006	2007	2007	2007	2007	2007
<b>B</b>		0	0	0	-		0	0
Banizoumbou	2007	2007	2007	2008	2008	2008	2008	2008
	0	-	9	3	0	0		0
	2008	2008	2008	2009	2009	2009	2009	2009
		0	0	0	-	-	•	0
	2009	2009	2009	2010	2010	2010	2010	2010
	0	-	0	WDB	1 WDB 3	WDB 3	WOR & WO	
	2010	2010	2010		T T T T T	THE		
	Octobre	Novembre	Decembre	Janvier	Fevrier	Mars	Avril	Mai
					0	0		
				2006	2006	2006	2006	2006
	0	-	$\sim$	0	0		-	0
	2006	2006	2006	2007	2007	2007	2007	2007
Cinzzono			0	0	0	-	-	
CIIIZZalla	2007	2007	2007	2008	2008	2008	2008	2008
	0	0	0	0	0	0	3	0
	2008	2008	2008	2009	2009	2009	2009	2009
		-	0	0	0	-	-	0
	2009	2009	2009	2010	2010	2010	2010	2010
	0	0	0	WRC 1	WRC 2	WRC 3 WR	C4 WRC 5	WRCS
	2010	2010	2010					
	Octobre	Novembre	Decembre	Janvier	Fevrier	Mars	Avril	Mai
				9	-		0	0
				2006	2006	2006	2006	2006
			0	2			-	0
	2006	2006	2006	2007	2007	2007	2007	2007
Mhour		0	3		9			0
Wibbul	2007	2007	2007	2008	2008	2008	2008	2008
	-	-	0		-	-	-	0
	2008	2008	2008	2009	2009	2009	2009	2009
	-	0	0	0			0	0
	2009	2009	2009	2010	2010	2010	2010	2010
		0	0	WRA	41 WRM 2	WRM 3	WRM 4	WRM 0
	2010	2010	2010					

Figure 5.7 – Cycle interannuel d'occurrence des types de temps de chacune des trois stations Persistances et transitions inter-classes.

### 5.2.2.3 Persistances et transitions inter-classes

La succession des régimes de temps au cours de la saison mise en évidence à partir des occurrences mensuelles moyennes suggère des transitions préférentielles d'un régime de temps à l'autre. Nous avons donc calculé les probabilités de passer d'un régime de temps à un autre sur un délai de quatre jours sur chacune des stations (Tableau 5.6). La lecture du tableau se fait par ligne. Sur 4 jours, la probabilité la plus forte est qu'un régime de temps n'évolue pas en un autre type de temps. Nous avons constaté que les régimes de temps restent plus persistants sur quatre jours, ce qui semble être lié au fort cycle saisonnier des données de la zone d'étude.

Tableau 5.6 – Probabilité de passer d'une classe à une autre en quatre jours (les valeurs en gras montrent la probabilité qu'un type de temps reste sur lui en quatre jours).

Baniz	WRB1	WRB2	WRB3	WRB4	WRB5	WRB6
WRB1	0,43	0,30	0,07	0,03	0,00	0,17
WRB2	0,19	$0,\!75$	0,02	$0,\!00$	$0,\!00$	$0,\!05$
WRB3	0,06	$0,\!04$	$0,\!54$	$0,\!18$	0,06	$0,\!12$
WRB4	0,00	$0,\!00$	$0,\!14$	$0,\!58$	$0,\!19$	0,10
WRB5	0,00	0,00	$0,\!14$	0,27	$0,\!59$	0,00
WRB6	0,05	$0,\!12$	$0,\!09$	$0,\!21$	$0,\!02$	$0,\!51$
Cinzana	WRC1	WRC2	WRC3	WRC4	WRC5	WRC6
WRC1	0,60	0,14	0,10	0,00	0,00	0,16
WRC2	0,31	$0,\!58$	0,00	$0,\!00$	$0,\!00$	$0,\!11$
WRC3	0,06	$0,\!00$	$0,\!43$	0,21	$0,\!13$	$0,\!17$
WRC4	0,00	$0,\!00$	$0,\!23$	$0,\!55$	0,21	0,02
WRC5	0,00	$0,\!00$	0,09	$0,\!30$	0,60	0,00
WRC6	0,12	$0,\!16$	$0,\!12$	$0,\!02$	$0,\!02$	$0,\!55$
Cinzana	WRC1	WRC2	WRC3	WRC4	WRC5	WRC6
WRC1	0,60	0,14	0,10	0,00	0,00	0,16
WRC2	0,31	$0,\!58$	0,00	$0,\!00$	$0,\!00$	0,11
WRC3	0,06	$0,\!00$	$0,\!43$	0,21	$0,\!13$	$0,\!17$
WRC4	0,00	$0,\!00$	$0,\!23$	$0,\!55$	0,21	0,02
WRC5	0,00	$0,\!00$	0,09	$0,\!30$	0,60	0,00
WRC6	0,12	0,16	$0,\!12$	0,02	0,02	$0,\!55$

La persistance de chaque régime de temps a été quantifiée par la probabilité de chaque régime de temps à rester inchangé en fonction du temps (Figure 5.8). La probabilité à t=0 correspond à la probabilité qu'un type de temps reste sur lui-même sur trois heures (base initiale). La persistance des régimes de temps diminue logiquement avec le nombre de jours, mais de façon différente selon le régime de temps et la station. Sur le pas de temps de trois heures les régimes de temps sont moins persistants sur les stations continentales de Cinzana et Banizoumbou que sur la station de Mbour. Les régimes de temps de la station de Mbour semblent être moins persistants que ceux des autres stations : la probabilité est inférieure à 0.5 au bout de trois jours pour les régimes WR3, WR4, WR5, WR6, alors qu'elle est supérieur à 0.5 au deux autres stations. Le régime de temps 2, caractéristique des mois de transition entre la saison sèche et la saison des pluies est la plus persistant. Il évolue de la même façon qu'à Banizoumbou. A Banizoumbou, les régimes de temps d'Harmattan évoluent de la même façon, avec des probabilités de rester inchangés qui varient de 0.5 à 0.6 au bout de 5 jours. A Banizoumbou, le régime WR1 est le moins persistant : sa probabilité de rester inchangé est inférieure à 0.5 après 2 jours. A Cinzana, les régimes WR1 et WR2 sont les plus persistants, mais ils diffèrent moins des autres régimes que sur les autres stations.

On observe sur certains régimes de temps que la décroissance n'est pas stricte, ce qui semble liée à la transformation des données. En effet notre base de données initiale (de réalisation des régimes de temps) est constituée sur un pas de temps de 3 heures. Ainsi pour avoir les régimes de temps sur un pas d'un ou deux jours, on prend le régime de temps qui apparaît le plus en un ou deux jours et ainsi de suite jusqu'à la fin pour constituer la nouvelle base. Une fois la base obtenue on détermine la succession des régimes de temps. Cependant le régime de temps dominant en deux jours peut ne pas l'être en 3 jours ce qui peut entrainer ces plateaux observés sur certaines courbes.



Figure 5.8 – Probabilité d'une classe à rester sur elle même (a :Banizoumbou; b :Cinzana; c : Mbour)

Pour synthétiser les résultats présentés ci-dessus et pour mieux appréhender la signification de ces régimes de temps, nous avons appliqué une analyse factorielle des correspondances (AFC) aux occurrences mensuelles, associant les fréquences d'occurrence de chaque régime de temps pour chaque mois de chaque année, et permettant de projeter sur le premier plan factoriel les régimes de temps et les observations (mois et années).

La Figure 5.9 suivante présente la projection de l'ensemble des types de temps des trois stations, année par année, sur le premier plan factoriel constitué des axes qui totalisent

62,2% d'inertie (Tableau 5.7). Ce pourcentage élevé indique que la projection sur ce plan est représentative de l'organisation générale des WR et fournit une illustration visuelle de synthèse intéressante. Le regroupement des différents régimes de temps, des trois stations (de WR1 à WR6) sur le premier plan factoriel permet de voir la forte dynamique régionale des régimes. Le premier panneau (en haut à gauche) présente les WR seuls. Il montre d'une part la forte cohérence spatiale des régimes de temps tels qu'ils ont été classés, puisqu'on trouve regroupés les régimes WRB1-WRC1-WRM1 et les régimes WRB2-WRC2-WRM2, situés à gauche sur l'axe 1 et opposés suivant l'axe 2, les régimes 4 et 5 des 3 stations à droite sur l'axe 1, enfin les régimes WRB3-WRC3-WRM3 et les régimes WRB6-WRC6-WRM6 proches de l'origine des axes. L'opposition suivant l'axe 1 décrit les contrastes entre les mois de Avril, Mai et Octobre de transition par rapport aux mois d'harmattan de Décembre, Janvier, Février, Mars et Novembre étant au centre du plan. L'axe 2 oppose le mois d'Octobre aux mois d'Avril et Mai. En comparant cette figure à la Figure 5.6, on retrouve le fait que les WR1 sont rencontrés majoritairement au mois d'Octobre, les WR2 au mois d'Avril et Mai, les WR4 et WR5 aux mois de Décembre, Janvier et Février, et les WR3 et WR6 aux mois de Mars et Novembre. De 2006 à 2010, les positionnements des mois sont relativement stables, montrant la dominance du cycle saisonnier sur la variabilité interannuelle. On peut noter cependant quelques variations comme pour les mois d'Avril et Mai, très proches des WR2 en 2010 et plus fluctuants les autres années. Ces résultats montrent que les régimes de temps obtenus décrivent bien la saisonnalité climatique en Afrique de l'Ouest.

Tableau 5.7 – Tableau des pour centages de variance expliquée des deux premiers axes et leur somme.

	Axe factoriel 1	Axefactoriel2	Somme
% devariance	41	21,2	62,2



Figure 5.9 – Caractérisation annuelle des régimes de temps par les mois a travers un plan factoriel.)

#### 5.2.2.4 Structures spatiales et interprétations météorologiques et climatiques

Les caractéristiques météorologiques et climatiques des types de temps des trois stations ont été étudiées en examinant les moyennes des variables météorologiques en surface (1000 hPa) sur l'ensemble de la zone d'étude. La Figure 5.10, présente les vents (m/s) et l'humidité spécifique (g/kg) par régime de temps, la Figure 5.11 présente les vents et la température (°C) par régime de temps et la Figure 5.12 présente les vents et le géopotentiel (mgp) par régime de temps dans chaque station. Le point sur la carte représente la station

de mesure considérée. De manière cohérente avec les résultats précédents, les 6 régimes de temps décrivent la saisonnalité climatique sur l'Afrique de l'ouest. Les structures de maximum de température, sur la moitié nord du domaine, associées aux structures de minimum de géopotentiel, décrivent la position et l'intensité de la dépression thermique saharienne. Les gradients de température et de pression associés entre le nord et le sud du domaine contrôlent l'extension en latitude des vents d'harmattan et de mousson. Ceci se traduit aussi par une extension méridienne plus ou moins forte des champs d'humidité de basses couches. Les 6 régimes de temps déclinent ainsi l'évolution de la circulation de basses couches d'octobre à mai. Pour chacune des stations de Banizoumbou à Mbour, les régimes de temps peuvent être regroupés principalement en deux groupes. Un groupe où les vents proviennent en majorité du nord-est est représenté par les types de temps WR3, WR4, WR5, WR6. Ce groupe peut aussi se scinder en deux, les régimes WR4 et WR5étant les plus secs avec la structure dynamique positionnée le plus au sud et représentant principalement les mois de novembre à janvier, et les régimes WR3 et WR6 présentant une situation intermédiaire caractéristique des mois de février à avril et octobre à novembre. Le deuxième groupe, WR1 et WR2, représente les structures dynamiques positionnées les plus au nord avec l'extension méridienne maximale des vents de sud et des fortes valeurs d'humidité. Nous remarquons de nombreuses similarités dans les caractéristiques météorologiques des régimes de temps des trois stations. Les régimes de temps WR1-WR2, caractéristiques des mois de transition entre la saison sèche et la saison des pluies (maioctobre), présentent autour des stations une forte humidité, une forte température et un géopotentiel faible sur la zone étudié. A l'inverse, les régimes de temps WR4-WR5 caractéristiques des mois de novembre à février présente un fort géopotentiel, une humidité et une température plus basse. Les régimes de temps WR3 et WR6 apparaissent moins humides et moins chauds que WR1-WR2, mais plus humides et plus chauds que WR4-WR5. Au niveau des stations à Banizoumbou et Cinzana, les régimes de temps WR1-WR2 sont caractéristiques des vents de sud-ouest (vents de mousson) alors que les régimes de temps WR3-WR4-WR5-WR6 se caractérisent par des vents de nord-est c'est à dire des vents d'harmattan. Ces résultats sont identiques à ceux présenté par Yahi et al. [2013] sur les stations de Banizoumbou et Cinzana. Bien qu'il existe des similitudes entre les différents régimes de temps des trois stations, quelques différences peuvent être notées entre les régimes de temps des stations continentales Banizoumbou, Cinzana et la station de côtière de Mbour. La station de Mbour présente une humidité plus élevée et des températures plus basses que les deux autres stations. Par ailleurs, on a vu dans le tableau des confusions de classes (Tableau 5.5) que les régimes de temps WR3, WR4 et WR6 de MBour se projettent sur plus d'une classe à Cinzana, ce qui peut être dû à des structures dynamiques liées au climat océanique de MBour différentes de celles plus continentales des deux autres stations.



101





Figure 5.10 – cartes moyennes des vents (m/s) obtenus et de l'humidité (g/kg) à 1000 hPa sur les stations de (a)Banizoumbou (b) Cinzana et (c) Mbour pour les six régimes de temps.





b : Cinzana



Figure 5.11 – cartes moyennes des vents (m/s) obtenus et la température (°C) à 1000 hPa sur les stations de (a)Banizoumbou (b) Cinzana et (c) Mbour pour les six régimes de temps.


a : Banizoumbou





Figure 5.12 – cartes moyennes de vents (m/s) et de géopotentiel (mgp) à 1000 hPa pour chaque classe sur les stations de (a) Banizoumbou, (b) Cinzana et (c) Mbour pour les six classes.

Pour caractériser un peu plus la structure dynamique des différents régimes de temps, les figures (Figure 5.13 et Figure 5.14) suivantes présentent les cartes moyennes des vents et de la vitesse verticale à 500 hPa et 850 hPa. Ces figures montrent le cycle saisonnier du déplacement du front intertropical (*FIT*; vitesse verticale à 850 hPa) et de la zone de convergence inter-tropicale (*ZCIT*; vitesse à 500 hPa). De janvier à mars les structures sont dans leur position la plus au sud (côte guinéenne). A partir du mois d'avril, elles commencent à remonter vers le nord pour atteindre leur position la plus septentrionale entre juillet et septembre. A partir d'octobre, elles commencent à redescendre vers le sud. On remarque aussi très nettement le décalage permanent en latitude entre le *FIT* et la *ZCIT*.

Sur les cartes des différents régimes de temps (Figure 5.15; Figure 5.16), on retrouve cette saisonnalité comme avec les champs dynamiques précédents. Ceci se détecte à 500 hPa par

l'extension plus ou moins importante de la ZCIT (zones en bleu caractérisant des vitesses ascendantes) dans le domaine, et par l'extension spatiale des zones de subsidence (orangerouge) correspondant à la branche descendante de la cellule de Hadley. A 850 hPa, les zones de vitesses verticales ascendantes signent la localisation du FIT. Pour l'ensemble des trois stations, c'est au niveau des régimes de temps 1 et 2 les plus humides que la localisation du FIT est plus septentrionale. Pour les régimes de temps les plus secs de l'ensemble des trois stations (WR4 et WR5), le FIT se localise tout à fait au Sud sur la côte (Golfe de Guinée). Le positionnement de la ZCIT suit la même évolution saisonnière que le FIT.



Figure 5.13 – Cartes moyennes du cycle saisonnier des vents (m/s) et de la vitesse verticale (hPa/s) à 500 hPa (valeurs négatives pour vitesses ascendantes et inversement).



Figure 5.14 – Cartes moyennes du cycle saisonnier des vents (vecteurs, m/s) de la vitesse verticale (hPa/s) à 850 hPa (couleur bleu pour valeurs négatives des vitesses verticales ascendantes et inversement).







Figure 5.15 – Cartes moyennes des vents (m/s, vecteurs) et de la vitesse verticale (hPa/s, couleur) à 500 hPa dans chaque classe (couleur bleu : valeurs négatives, vitesses ascendantes et couleur rouge : valeur positives de vitesse verticale).







Figure 5.16 – Cartes moyennes des vents (m/s, vecteurs) et de la vitesse verticale (hPa/s, couleur) à 850 hPa dans chaque classe (couleur bleu : valeurs négatives, vitesses ascendantes et couleur rouge : valeur positives de vitesse verticale).







Figure 5.17 – Rose des vents des mesures obtenues à la station d'étude (données in-situ) sur les stations de (a) Cinzana, (b) Banizoumbou et (c) Mbour pour les six classes.

La Figure 5.17 présente la rose des vents de réalisée avec la vitesse et la direction des vents mesurés in situ par type de temps pour chaque station. Les résultats sont très cohérents avec les directions indiquées par les champs météorologiques présentés dans les figures précédentes. Sur Banizoumbou et Cinzana, on retrouve très bien la dominance des vents de sud-ouest des régimes WR1 et WR2. Pour les autres régimes, on retrouve la dominance des vents de nord-est et des vents d'est sur Cinzana et Banizoumbou, ainsi qu'une orientation moins dispersée des directions de vents sur les WR4 et WR5. Sur M'Bour, la distribution des direction de vents est un peu plus partagée au sein de chaque régime de temps. Les WR4 et WR5 sont eux aussi bien structurés autour d'une direction de nord-est. Les régimes WR3 et WR6 intègrent une composante nord-ouest qui n'apparait pas pour les deux autres stations, avec une composante est-nord-est présente sur WR6 seulement. Enfin les WR1 et WR2 ont une composante d'ouest dominante au lieu de sud-ouest pour les deux autres stations. On retrouve là l'influence du climat océanique de cette station.

Les profils verticaux moyens de vent zonal et méridien des pixels où sont localisées les sont présentés sur la Figure 5.18. Pour Banizoumbou et Cinzana, les profils des classes associées à l'harmattan ont des allures identiques. De même pour les classes associées à la présence du flux de mousson (WR1 et WR2). Les profils des régimes de temps WR1 et WR2 se différentient fortement des autres profils par la présence de vent de sud-ouest dans les basses couches et d'un pic de vent zonal à 700 hPa traduisant la présence du JEA. Cependant le JET n'est pas encore très visible en haute troposphère pour les mois de ces classes (mai et octobre pour les plus proches de l'été). Banizoumbou se différentie plus que Cinzana pour ces 2 classes. A M'Bour, de par l'influence océanique, la différentiation des classes « de mousson » par rapport aux autres classes est moins nette, comme c'était le cas sur les roses des vents. On observe cependant le pic du JEA dans le profil de vent zonal (WR1 et WR2). Pour l'ensemble des trois stations, les régimes de temps assimilés au flux d'harmattan (WR3-WR4-WR5-WR6) se caractérisent par un pic de vent zonal à 925 hPa. On le retrouve également sur la composante de vent méridienne pour les stations de Cinzana et Banizoumbou. Ainsi ce maximum à 925 semble signaler la présence du jet nocturne de basses couches. En effet, un NLLJ est observé sur Sahara et le Sahel en provenance du nord pendant la saison sèche et du sud-ouest pendant la saison des pluies [Parker et al., 2005; Washington and Todd, 2005]. Ces résultats sont comparables avec ceux présentés par Yahi et al. [2013]. Cependant la classification de Yahi et al. [2013] ne permettait pas de mettre en évidence le pic à 700 hPa marquant la présence JEA. Cette différence peut être du au fait qu'on a utilisé plus de classes et plus de variables mais aussi au fait qu'un travail préliminaire de sélection des variables a été réalisé dans ce travail.



Figure 5.18 – Profils de vent obtenu avec les classes de la base de données pour chacune des trois stations

Enfin, pour faire le lien avec la partie suivante sur l'analyse des données AOT et  $PM_{10}$ , on analyse la répartition du contenu atmosphérique en poussières pour les 6 régimes de temps. De la même façon que l'on a caractérisé les conditions météorologiques moyennes associées à chaque régime de temps, on a également caractérisé le contenu moyen en aérosols désertiques sur la zone d'étude en utilisant l'AI - OMI. La Figure 5.19 présentent l'AI moyen pour chaque WR sur les trois stations, ainsi que le vent et le géopotentiel moyen à 925 hPa et . Quel que soit le régime, on retrouve des localisations communes des zones de fort AI. La zone où les AI sont les plus forts est localisée è l'extrême est du domaine et centrée à environ 16°N. Elle se situe en aval de la dépression de Bolélé, une zone reconnue comme la plus active du Sahara [*Prospero et al.*, 2002]. Une deuxième zone de forts AI est centrée à environ 19°N et 11°O. Une zone troisième zone apparait sur le régime WR2 et WRC6, centrée sur 18°N et 1°O.

décrites dans la littérature (cf page 23). Par contre, l'intensité relative de ces différentes zones varie selon les types de temps.

Pour les trois stations, les AI les plus forts correspondent aux régimes WR1 et WR2 caractéristique du flux de mousson. Ces forts AI sont localisés sur la partie nord du domaine, en zone saharienne, en accord avec *Engelstaedter and Washington* [2007] ont montré que les AI maximum sont maximum en été. Mais si l'on examine les directions de vent, on constate que les aérosols sahariens responsables de ces forts AI ne peuvent pas être transportés en surface sur les stations de mesure. Pour les autres régimes, la zone de plus forts AIs est la zone localisée à l'est du domaine, à l'ouest de la dépression de Bodélé. Les directions de vents de surface des régimes WR3, 6, 4et5 permettent le transport depuis cette source vers les trois stations. Mais la station de M'Bour peut être également affectée par la zone source la plus à l'ouest, pour les régimes WR3, voire WRM2. De façon similaire, la source située au centre pourrait affecter les mesures réalisées à la station de Cinzana dans le régime WRC6.







Figure 5.19 – Cartes moyennes des vents (m/s, vecteurs) à 925 hPa, de Géopotentiel (mgb, isolignes) à 925 hPa et l'Aerosol Index (couleur) (A :Banizoumbou; B :Cinzana; C :Mbour)

## 5.2.3 Relation AOT- $PM_{10}$ avec et sans régimes de temps

Dans cette partie, nous étudions la relation AOT- $PM_{10}$  avant et après la classification en régimes de temps afin d'évaluer si cette classification peut améliorer globalement cette relation et de voir quels sont les types de temps qui influencent positivement ou négativement la relation, ceci pour chacune des trois stations. Nous suivons l'approche développée par Yahi et al. [2013], mais nous l'utilisons à partir d'une classification en régimes de temps plus précise et nous examinons aussi la station de M'Bour et les contrastes possibles entre cette station et les deux autres stations plus continentales.

Afin d'avoir une base de données d'étude de la relation AOT- $PM_{10}$ , on a fait une intersection par rapport aux dates entre les données météorologiques à la station, les données de  $PM_{10}$  et l'AOT, puis entre cette dernière base formée et les données météorologiques sur les 9 niveaux de pression. On a ainsi formé une base des données météo et  $PM_{10}$ , c'est-à-dire une base de données qui correspond aux situations météorologiques à grande échelle pour lesquelles on a une mesure d'épaisseur optique et une mesure de $PM_{10}$ . La combinaison de ces données nous a permis d'avoir une base de données de 2960 lignes pour la station de Banizoumbou, 2291 lignes pour la station de Cinzana, 2549 lignes pour la station de Mbour. Pour une validation très rigoureuse de nos résultats, on a séparé l'ensemble des données en un ensemble d'apprentissage représentant 3/4 des 2960 situations de la station de Banizoumbou, sélectionnées d'une manière aléatoire et un ensemble de test comportant les 1/4 des situations restantes. La base d'apprentissage contient donc 2220 observations et la base de test 740 observations. Et de la même manière, on a séparé les données de la station de Cinzana et la station de Mbour, ce qui nous a donné sur les 2291 observations de la station de Cinzana, 1719 pour l'apprentissage et 573 observations pour le test. Sur les 2549 observations de la station de Mbour, 1912 observations sont disponibles pour la base d'apprentissage et 637 observations pour la base de test.

La Figure 5.20 suivante présente la relation AOT-  $PM_{10}$  avant la séparation en types de temps. Ainsi nous présentons le diagramme de dispersion, la droite de régression et le coefficient de corrélation linéaire entre AOT-  $PM_{10}$ . Avant la séparation en types de temps, le coefficient de corrélation linéaire entre la concentration en  $PM_{10}$  et l'épaisseur optique est de 0.72 pour la station de Cinzana, 0.69 pour la station de Banizoumbou et de 0.59 pour la station de Mbour. Ces figures et les corrélations calculées permettent de voir la complexité de la relation entre les deux paramètres. L'étude faite dans le chapitre précèdent a permis de voir que le cycle saisonnier de l'AOT apparaît être diffèrent de celui de la concentration en  $PM_{10}$  et que ceci peut être du à la distribution en altitude du transport des poussières. Ceci nous incite à utiliser l'information des types de temps afin de voir comment les situations météorologiques qui influencent la relation.

Afin de voir si les types de temps définis permettent d'améliorer la relation AOT- $PM_{10}$ , nous avons tout d'abord projeté les données in-situ de chaque station sur les types de temps correspondant à la station. Après avoir projeté les données in-situ sur les types de temps, nous avons jugé nécessaire de voir quelles sont les variables in-situ caractéristiques de chaque type de temps. Ainsi pour caractériser les types de temps en fonction des données in-situ, on a utilisé les valeurs test de *Lebart et al.* [1997].

Pour caractériser une classe par une variable in situ, on compare  $X_k$ , la moyenne d'une variable X dans la classe k, à la moyenne générale  $\overline{X}$  de la variable et on évalue l'écart en tenant compte de la variance  $S_k^2(X)$  de cette variable dans la classe. La valeur-test est ici simplement la quantité :

$$t_k(X) = \frac{\bar{X}_k}{S_k(X)}$$

avec :

$$S_k^2(X) = \frac{n - n_k S^2(X)}{(n - 1)n_k}$$

où  $S^2(X)$  est la variance empirique de la variable X.

la variance d'une moyenne dans le cas d'un tirage sans remise des k éléments concernés. Si la valeur test est positive (>2) (resp. négative  $\leq 2$ ), la classe est caractérisée  $S_k^2(X)$  par les valeurs significativement fortes (resp. faibles) de la variable.

Nous remarquons que sur l'ensemble des trois stations (Tableau 5.8), les types de temps WR1 et WR2 qu'on a considéré comme des types de temps humides sont caractérisés par une humidité locale significativement élevée et une  $PM_{10}$  locale significativement faible. Pour les stations de Cinzana et Banizoumbou, WR1 est caractérisé par une vitesse de vent et AOT faibles. La classe WR2 pour les deux stations de Banizoubou et Cinzana est principalement caractérisée par la température et l'AOT significativement élevées alors



Figure 5.20 – Nuage des points et droite de régression linéaire entre AOT et  $PM_{10}$  avant la séparation en types de temps

qu'à Mbour elle est caractérisée seulement par l'humidité et l'AOT élevée. Le type de temps WR3 est respectivement caractérisé par la température significativement élevée à Cinzana. En plus de la température WR3 est caractérisé par l'AOT significativement fort à Banizoumbou et la concentration en  $PM_{10}$  à Mbour, les autres variables étant faibles ou pas caractéristiques du type de temps WR3. Le type de temps WR4 est caractérisé par la vitesse du vent et la concentration en  $PM_{10}$  fortes sur les stations de Cinzana et Banizoumbou. Nous remarquons également que le type de temps WR5 de la station de Cinzana n'est caractérisé par aucune variable significativement forte alors qu'à Banizoumbou, et à Mbour, il est caractérisé par la vitesse du vent significativement forte. Le types de temps WR6 est principalement caractérisé par la température, la concentration en  $PM_{10}$ et l'AOT significativement forts sur les stations de Banizoumbou et Cinzana, et la vitesse de vent en plus pour la station de Cinzana. A Mbour, le type de temps WR6 est caractérisé par l'humidité et l'AOT significativement élevée. Ce tableau nous permet de voir que sur l'ensemble des trois stations, les types de temps WR1 et WR2 caractéristiques du flux de mousson sont caractérisés par une concentration locale en  $PM_{10}$  significativement faible et une humidité locale significativement forte, mais avec des valeurs locales d'AOT contrastées. Pour les autres régimes de temps, les  $PM_{10}$  sont souvent significativement forts et les valeurs d'AOT contrastées aussi.

Baniz	WRB1	WRB2	WRB3	WRB4	WRB5	WRB6
VVent	-8,84	2,15	-3,42	7,25	2,15	-2,15
Temp	-0,63	$15,\!43$	$9,\!54$	-19	-16,1	$11,\!59$
Hum	25,79	25,71	$-14,\!55$	-10,59	-9,91	-11,18
$PM_{10}$	-7,67	-6,38	$1,\!6$	$6,\!48$	$1,\!03$	$2,\!91$
AOT440	-3,05	$12,\!18$	$2,\!05$	-8,31	-7,88	4,44
Cinzana	WRC1	WRC2	WRC3	WRC4	WRC5	WRC6
VVent	-11,59	0,72	-1,28	$5,\!88$	0,05	$5,\!24$
Temp	0,50	$9,\!47$	$8,\!95$	-15,01	-18,32	$18,\!69$
Hum	31,31	$11,\!36$	-10,82	-7,14	-8,83	-8,81
$PM_{10}$	-8,84	-3,28	1,03	$3,\!19$	$0,\!45$	$5,\!57$
AOT440	-2,61	5,73	$1,\!05$	$-6,\!65$	-8,47	$13,\!17$
Mbour	WRM1	WRM2	WRM3	WRM4	WRM5	WRM6
VVent	-4,73	-1,58	$0,\!50$	$1,\!65$	$5,\!04$	-2,28
Temp	4,11	-1,58	$5,\!15$	1,96	-5,21	-1,94
Hum	8,98	$13,\!07$	-5,24	-11,93	-8,37	$3,\!47$
$PM_{10}$	-4,76	-3,18	$3,\!35$	2,31	$1,\!53$	-0,40
AOT440	-2,21	$10,\!23$	-4,17	-4,44	-5,94	$3,\!29$

Tableau 5.8 – Significativité des variables in-situ dans chaque classe

Après projection de la base de donnée in-situ sur les types de temps, nous avons tracé la Figure 5.21 qui présente la distribution de la concentration en  $PM_{10}$  (en log) et de l'AOT dans chaque type de temps sur chaque station. Nous remarquons que par rapport à la concentration en  $PM_{10}$ , sur l'ensemble des stations, les types de temps humides WR1et WR2 ont des médianes et des quartiles Q25 and Q75 faibles et les types de temps correspondant au flux de l'harmattan ont des valeurs plus élevées. Concernant l'AOT, les régimes de temps les plus secs (WR4 et WR5) sont associés à des médianes et des quartiles faibles sur l'ensemble des trois stations. Pour l'ensemble des trois stations, c'est le régime de temps WR1 qui regroupe l'AOT comme la concentration en  $PM_{10}$  la plus faible. Bien que sur chaque station des données atypiques semblent être plus présentes sur les types de temps plus secs, le type de temps WR2 caractérisé en majorité par les données du mois de Mai présente des données extrêmes comparativement à WR1, ce qui semble être lié aux systèmes convectifs qu'on observe dans cette zone à l'approche de la saison des pluies [Marticorena et al., 2010].



Figure 5.21 – Boxplot présentant la Médiane (lingne rouge), les quartiles 25% et 75% des concentrations en  $PM_{10}$  et de l'AOT pour chaque régime de temps et pour chaque station.

Nous analysons maintenant les relations  $AOT-PM_{10}$  plus précisément. La Figure 5.22 et le Tableau 5.9 présentent les nuages de points et les coefficients de corrélation entre  $PM_{10}$  et AOT pour chaque régime de temps et chaque station. Pour toutes les stations, nous remarquons que les types de temps correspondants aux mois d'harmattan ont des corrélations entre  $AOT-PM_{10}$  élevées par rapport à celles obtenues sur l'ensemble des données avec respectivement des corrélations (0.85 WRB3; 0.82 WRB4; 0.88 WRB5; 0.85 WRB6) pour la station de Banizoumbou et des corrélations de (0.82 WRC3; 0.80 WRC4; 0.81 WRC5; 0.84 WRC6) pour la station de Cinzana. Pour les deux stations, les corrélations correspondantes aux mois d'intersection entre la saison des pluies et saison sèche c'est à dire les mois de début ou fin de mousson ont des corrélations AOT- $PM_{10}$ inferieures à celles obtenues sur l'ensemble des bases de chaque station avec respectivement des corrélations de (0.42 WBC1; 0.44 WRB2) pour la station de Banizoumbou et des corrélations de  $(0.62 \ WRC1; 0.60 \ WRC2)$  pour la station de Cinzana. La station de Mbour présente des types de temps un peu différents par rapport aux autres en terme de corrélation. Le type de temps WRM2 caractéristique du flux de mousson a une corrélation supérieure à la corrélation sur toute la base et le type de temps WRM4 caractéristique du flux d'harmattan à une corrélation inférieure contrairement aux autres stations où on avait les corrélations correspondantes aux mois d'intersection entre la saison des pluies tous inférieures à la corrélation sur toute la base et les autres supérieures.

Ces résultats montrent l'apport des régimes de temps sur la relation AOT- $PM_{10}$ . Nous avons indiqué dans le Tableau 5.9 la corrélation générale calculée pour chaque station sans différentiation des régimes de temps. Cette corrélation est souvent inférieure à celle déterminée pour les régimes de temps individuels. Ainsi nous constatons que sur les stations continentales de Cinzana et Banizoumbou les types de temps caractéristiques du flux d'harmattan donnent une meilleure relation AOT- $PM_{10}$  par rapport à celles caractéristiques du flux de mousson. Le fait que la relation AOT- $PM_{10}$  soit meilleure en saison d'harmattan qu'en saison de mousson peut être du au changement de direction des vents. En effet les fortes concentrations obtenues au Sahel sont enregistrées au cœur de la saison sèche pendant la période d'harmattan et proviennent du Sahel [Marticorena et al., 2010]. Cependant en saison de mousson, le changement de la direction des vents, qui viennent de la mer ou des zones non désertiques entraine une baisse de la présence et de la teneur des aérosols dans l'atmosphère. Cette baisse semble être à l'origine de la mauvaise relation. Cela peut être du aussi au fait que en saison sèche le maximum de concentration est obtenu au sol alors qu'en saison des pluies, le transport de poussière se fait en général en altitude.

Les relations obtenues en terme de corrélations entre les différents régimes de temps étant différentes, nous avons voulu voir si des différences peuvent être observées sur les pentes des régressions en traçant la Figure 5.23. Pour la station de Banizoumbou, la pente de la régression pour les régimes de temps caractéristiques du flux d'harmattan les plus secs WRB4 et WRB5 (avec respectivement 718 et 600  $\mu g.m^{-3}$ ) est plus élevée que pour les régimes de temps WRB1, WRB2, WRB3, et WRB4 (respectivement 107, 232, 473 et  $406 \ \mu g.m^{-3}$ ). Cependant, la pente des régimes de temps semi-humides (WRB3 et WRB6) sont plus élevées que celles des régimes de temps les plus humides caractéristiques du flux de mousson (WRB1 et WRB2). De même, les types de temps caractéristiques du flux d'Harmattan identifiés à Cinzana (de WRC3 à WRC6) ont des pentes plus élevées que les types de temps associés au flux de mousson (WRC1 et WRC2).

La différence entre les pentes des régimes de temps caractéristiques du flux d'harmattan et ceux caractéristiques du flux de mousson peut être expliquée par le fait que pour une mesure d'AOT donnée, la concentration en  $PM_{10}$  apparaît plus de deux fois plus grande sur les régimes de temps les plus secs que sur les régimes de temps les plus humides. Par exemple pour la station de Banizoumbou, la valeur 2 de l'AOT correspond à une valeur  $PM_{10}$  de 463 g.m-3 pour le régime de temps WRB2 caractéristique du flux de mousson et une valeur, respectivement de 1460  $\mu g.m^{-3}$  pour WRB4 et 1200 pour WRB5caractéristique du flux d'harmattan. A Cinzana pour une valeur de l'AOT égale à 2, on obtient une valeur de concentration en  $PM_{10}$  de près de 400  $\mu g.m^{-3}$  pour WRC2caractéristique du flux de mousson et une valeur d'environ 1200  $\mu g.m^{-3}$  respectivement pour WRC4 et WRC5 caractéristique du flux d'harmattan. Les résultats obtenus ici sont en accord et peuvent être comparés avec celles de Yahi et al. [2013].

Contrairement aux stations continentales (Cinzana et Banizoumbou), les pentes des régimes de temps la station de Mbour apparaissent différentes. Bien que la pente du régime de temps WRM1 (103  $\mu g.m^{-3}$ ) caractéristique du flux de mousson est plus faible, la pente du régime de temps WRM2 (278  $\mu g.m^{-3}$ ) caractéristique aussi du flux de mousson est plus élevé que celles de deux types de temps caractéristiques du flux d'harmattan (WRM4 (200  $\mu g.m^{-3}$ ) et WRM4 (233  $\mu g.m^{-3}$ )).

Nous retenons ici que les régimes de temps caractéristiques du flux d'harmattan amé-

liorent la relation pour les stations continentales (Banizoumbou et Cinzana) tandis que les régimes de temps caractéristiques du flux de mousson la détériorent. Nous avons retenu aussi que les régimes de temps peuvent être différenciés selon les pentes obtenues par régression, donnant des proportionnalités entre AOT (et donc l'épaisseur de la couche de poussière) et  $PM_{10}$  différentes. Les pentes des régimes de temps caractéristiques du flux d'harmattan sont plus élevées que les pentes des régimes de temps caractéristiques du flux de mousson. Par contre pour la station de Mbour qui est une station côtière, bien qu'il y ait des régimes de temps de l'harmattan qui améliorent la relation avec des pentes élevées et d'autres qui la détériorent, il y a aussi des régimes de temps caractéristiques du flux de la mousson qui améliorent la relation avec des pentes plus élevées que celles de certains régimes de temps caractéristiques du flux d'harmattan. La différence entre les stations continentales et la station côtière peut faire penser qu'à Mbour les alizés maritimes ont une influence sur la relation AOT- $PM_{10}$  au sein des régimes de temps.





Figure 5.22 – Nuage des points et droite de régression linéaire entre  $PM_{10}$  et AOT pour chaque classe et pour les trois station.

Tableau 5.9 – Corrélations entre  $PM_{10}$  et AOT par régime de temps et avant la séparation en régimes de temps pour les trois stations .

	WR 1	WR $2$	WR 3	WR4	WR5	WR6	Sans WR
Bani	$0,\!42$	0,44	$0,\!85$	$0,\!82$	$0,\!88$	0,84	0,69
Cinzana	$0,\!62$	$0,\!60$	$0,\!82$	$0,\!80$	$0,\!81$	$0,\!85$	0,72
Mbour	$0,\!32$	$0,\!56$	$0,\!54$	$0,\!48$	0,70	$0,\!52$	$0,\!51$



Figure 5.23 – Droite de régression entre  $PM_{10}$  et AOT pour chaque régime de temps.

Afin de restreindre les comparaisons à des situations clairement dominées par les poussières et non par d'autres aérosols comme les aérosols carbonés issus de feux ou les aérosols de pollution, on a trié les données en fonction du coefficient d'Angström. Ainsi on considère les observations correspondantes au coefficient d'Angström (Alpha) <0.4. Le Tableau 5.10 suivant présente pour chaque station le nombre de données de chaque type de temps avant et après tri. Nous remarquons sur tous les types de temps de chaque station, qu'il reste plus de 50% des données correspondant aux particules sauf le type de temps 4 sur les stations de Cinzana et Banizoumbou et le type de temps 5 à Mbour où on a le nombre d'observations restants  $\leq 50\%$ . Le fait qu'il reste moins de données pour WRB4, WRC4 et WRM5 peut être lié par le fait qu'ils sont dominés par les mois de Décembre-Janvier, ce qui correspond à la saison des feux au sud du Sahel. Il est intéressant de noter aussi le très fort pourcentage pour le régime de temps de mousson WR2 signifiant une quasi-absence des aérosols autres que les poussières.

La Figure 5.24 suivante présente la distribution des  $PM_{10}$  (en log) et de l'AOT dans chaque type de temps après enlèvement des observations ne correspondant pas aux cas de poussières. Comparé à la Figure 5.21, nous remarquons qu'en enlevant les données correspondant au coefficient d'Angström > 0.4, les valeurs des médianes et des quartiles des types de temps de l'ensemble des trois stations augmentent. Cependant les types de temps WR1 et WR2 caractéristiques du flux de mousson ont toujours des médianes et des quartiles moins forts par rapport aux autres. Nous pouvons dire que sur l'ensemble des trois stations, les poussières dominent car en ne conservant sur les classes de chaque station que les données correspondant au coefficient d'Angström < 0.4, on observe une augmentation de la moyenne de la variable  $PM_{10}$  et de l'AOT dans ces classes.



Figure 5.24 – Boxplot présentant la Médiane (lingne rouge), les quartiles 25% et 75% des concentrations en  $PM_{10}$  et de l'AOT pour chaque régime de temps et pour chaque station après le tri sur le coefficient d'Angstrom.

Bani	WR 1	WR 2	WR 3	WR4	WR5	WR6
NParclasse	274	572	501	651	475	487
N pour Alpha < 0.4	188	526	348	383	237	392
Pourcentage	68.61%	92%	69.46%	59%	50%	80.49%
Cinzana	WR 1	WR 2	WR 3	WR4	WR5	WR6
NParclasse	328	185	443	458	472	405
N pour Alpha < 0.4	187	156	283	261	213	342
Pourcentage	57%	$84{,}32\%$	64%	57%	45.12%	84.44%
Mbour	WR 1	WR 2	WR 3	WR4	WR5	WR6
NParclasse	133	654	544	439	313	466
N pour Alpha < 0.4	82	529	379	290	137	233
Pourcentage	62%	81%	70%	66.05%	44%	50%

Tableau 5.10 – Nombre d'observations avant et après le tri des cas de poussière

### Estimation de PM<sub>10</sub> à partir de l'AOT et des paramètres de mesure in-situ

Pour mieux affiner la relation  $AOT-PM_{10}$ , nous avons utilisé les données in-situ de chaque station captées par chaque classe, pour estimer les  $PM_{10}$  à partir de l'AOT et des variables météorologiques. L'algorithme des cartes auto-organisatrices est utilisé ici, mais dans un but de prédiction. Nous allons donc utiliser les cartes auto-organisatrices pour prévoir la concentration massique en  $PM_{10}$  à partir des AOT et cela pour chaque régime de temps, ceci afin de modéliser la relation  $AOT-PM_{10}$  pour un régime de temps spécifique et une localisation précise. Ainsi nous utiliserons comme variables d'entrée de la carte auto organisatrice, les variables d'AOT et les variables météorologiques locales : vitesse du vent, direction du vent, température, humidité relative et les variables de coefficient d'Angström de chaque station de la base définie plus haut. Pour les régimes de temps, on a donc six cartes auto organisées. Chaque neurone de chaque carte peut alors être associé à une moyenne de  $PM_{10}$  des données captées par ce dernier. Pour chacune des six cartes apprises par station, la Figure 5.25 et le Tableau 5.11 ci dessous, présentent les nuages des points et les corrélations entre  $PM_{10}$  estimée et  $PM_{10}$  mesurée de la base de test par régime de temps. Nous remarquons que sur les stations de Cinzana et Banizoumbou, les types de temps correspondant au flux de mousson ont toujours des corrélations moins élevées par rapport aux autres.



Cinzana



Figure 5.25 – Nuage des points entre  $PM_{10}$  mesuré et  $PM_{10}$  estimé pour chaque classe sur chaque station (Banizoumbou, Cinzana Mbour).

Tableau 5.11 – Corrélations entre  $PM_{10}$  estimée et  $PM_{10}$  mesurée de la base test pour chaque classe.

	WR 1	WR 2	WR 3	WR4	WR5	WR6	WR T
Bani	0,48	$0,\!55$	0,79	$0,\!86$	$0,\!82$	0,89	$0,\!83$
Cinzana	0,83	0,70	0,92	$0,\!87$	0,94	$0,\!87$	0,862
Mbour	$0,\!50$	$0,\!78$	0.81	0,76	$0,\!69$	$0,\!33$	0,70

Cependant, en récupérant et en fusionnant les données estimées et mesurées de chaque type de temps sur chaque station, on a tracé le Tableau 5.12 suivante qui représente les corrélations entre  $PM_{10}$  estimé et  $PM_{10}$  mesuré sur toute la base d'apprentissage et de test. Nous présentons ainsi sur toutes les stations, la relation à travers la corrélation aussi bien du test que de l'apprentissage sur données estimées et mesurées par rapport à la relation en types de temps. La corrélation obtenue après reconstitution des données estimées et mesurées est de 0.90 pour Cinzana, 0.80 pour la station de Mbour et 0.88 à Banizoumbou pour l'apprentissage et 0.86 à Cinzana , 0.70 à Mbour et 0.83 à Banizoumbou pour la base de test. Soit de très bonnes performances pour la modélisation sur une localisation spécifique de  $PM_{10}$  intégrant aussi le régime de temps présent.

	Apprentissage	Test
Bani	$0,\!88$	$0,\!83$
Cinzana	$0,\!90$	$0,\!86$
Mbour	$0,\!80$	0,70

Tableau 5.12 – Corrélations linéaire entre  $PM_{10}$  estimé  $PM_{10}$  mesuré du test et de l'apprentissage pour les trois stations.

### 5.2.4 Conclusion

Ce travail nous a permis de caractériser les situations météorologiques similaires en Afrique de l'Ouest pendant la saison sèche de 2006 à 2010 et leur influence sur la relation AOT- $PM_{10}$ . Pour ce faire nous avons appliqué les méthodes de classification et principalement une méthode de subspace clustering basée sur les cartes auto organisatrices de Kohonen combinée avec la Classification Ascendante Hiérarchique et les tests statistiques sur les données de profils verticaux composées des paramètres tels que (le vent zonal, le vent méridien, la température, l'humidité spécifique, la vitesse verticale et la hauteur de la couche limite). Ainsi 6 régimes de temps décrivant la saisonnalité climatique sur trois stations de l'Afrique de l'ouest ont été obtenus. Nous avons pu voir que les régimes de temps ne sont pas sensibles au cycle diurne. Par contre ils caractérisent fortement le cycle saisonnier. Dans les basses couches, nous avons identifié une dynamique régionale des régimes de temps, avec ceux caractéristiques du flux de mousson sur l'ensemble des trois stations (WR1 et WR2) et des types de temps caractéristiques du flux d'harmattan (WR3, WR4, WR5 et WR6). Le pic de vent zonal observé à 700 hPa sur les régimes de temps 1 et 2 signale la présence du JEA. La présence du NLLJ a aussi été mise en évidence avec le pic de vent à  $925 \ hPa$ .

La relation AOT- $PM_{10}$  examinée sur chaque régime de temps nous a permis de voir que les régimes de temps caractéristiques du flux d'harmattan présentent une relation meilleure en terme de corrélation par rapport a la relation avant la séparation en type de temps avec de forte valeurs d'AOT et de  $PM_{10}$  pour les stations continentales de Cinzana et Banizoumbou. Le fait que la relation  $AOT-PM_{10}$  soit meilleure en saison d'harmattan qu'en saison de mousson peut être du au changement de direction des vents, et par le fait que le maximum des concentrations de poussières au niveau du sol a lieu pendant la saison sèche alors que pendant la saison de mousson, les poussières sont généralement transportées en altitude. Ceci nous permet de dire que la météorologie joue un rôle important dans la relation AOT- $PM_{10}$ . L'estimation des concentrations en fonction des situations météorologiques locales de chaque régime de temps a permis d'obtenir une amélioration de la relation AOT- $PM_{10}$  dans la plupart des types de temps. Nous avons aussi obtenu une meilleure estimation globale des  $PM_{10}$  après la discrimination par les régimes de temps comparée à l'estimation sans types de temps. Par ailleurs, nous avons vu dans cette partie que ces résultats sont largement dominés par le cycle saisonnier. Ainsi dans le chapitre suivant, on filtrera les données en enlevant le cycle saisonnier moyen.

# 5.3 Régimes de temps obtenus sur les données en enlevant le cycle saisonnier moyen

Après avoir développé la même approche que Yahi et al. sur les données brutes qui sont dominées par le signal saisonnier, nous considérons maintenant les données désaisonnalisées par l'élimination du cycle saisonnier moyen calculé jour à jour afin de faire émerger la variabilité résiduelle, particulièrement d'échelle interannuelle. En effet Yahi et al. ont noté que l'utilisation des données brutes ne permet pas d'appréhender les origines d'évènements de soulèvement/transport de poussières importants mais sporadiques et d'échelle synoptique, qui sont observables sur le Sahel [Klose et al., 2010]. Au delà de ce travail, nous avons aussi tenté de combiner les résultats de ces deux approches pour évaluer la potentialité de mettre en place un système d'alerte statistique opérationnel aux évènements forts de poussière.

Cette partie se propose donc de définir de nouveaux types de temps en éliminant la variabilité saisonnière. Ainsi on a filtré les données de sorte que pour chaque variable de la base de données, on enlève, jour à jour, le cycle saisonnier moyen sur les cinq ans. Ce cycle saisonnier étant éliminé, il va rester la variabilité interannuelle et la variabilité intra-saisonnière.

Le procédé pour définir les régimes de temps sera le même que celui utilisé pour la définition des types de temps précédents. Nous comparerons ces deux ensembles de types de temps afin de pouvoir mieux interpréter ces nouveaux types de temps. Leurs liens et leurs apports pour une modélisation de la variabilité interannuelle des poussières seront examinés.

# 5.3.1 Caractérisation des régimes de temps

Apres avoir appliqué 2S - SOM sur la base de données désaisonnalisée, on a défini 6 classes qui correspondent aux 6 régimes de temps. Le nombre de classes a été défini en choisissant la partition la plus discriminante en tenant compte de la règle du « coude » appliquée sur le dendrogramme de la CAH puis en tenant compte aussi des indices de Davies Boudin.

Le Tableau 5.13 suivant présente le pourcentage d'occurrences de chaque régime de temps sur les cinq saisons sèches pour les trois stations. Nous remarquons que les pourcentages d'occurrences assez bien équilibrés entre les 6 types de temps de chaque station. Cependant pour les stations de Cinzana et Mbour, le régime de temps le plus fréquent est DWR4 (24,71% à Cinzana; 21,60% à Mbour) et le moins fréquent est DWR2 (10,43) à Cinzana DWR6 à Mbour alors qu'a Banizoumbou c'est DWR1 (20,14%) qui est le plus fréquent et DWR6 (11,15) le moins fréquent pour Banizoumbou et Mbour.

Tableau 5.13 – Pourcentage d'occurrence d	es données de chaq	que classe et de ch	aque station.
---	--------------------	---------------------	---------------

	DWR 1	DWR 2	DWR 3	DWR 4	DWR5	DWR6
Bani	14,40	$10,\!43$	14,40	24,71	$15,\!34$	20,71
Cinzana	20,81	$14,\!49$	$19,\!69$	$21,\!60$	14,72	$^{8,63}$
Mbour	$20,\!14$	$16,\!27$	$19,\!98$	$18,\!59$	$13,\!80$	$11,\!15$

### 5.3.1.1 Matrices de confusion

Comme pour l'approche précédente, nous avons recherché l'existence d'une dynamique régionale des régimes de temps en commençant par les tableaux suivants. Ainsi les résultats de la matrice de confusion (Tableau 5.14) nous a permis de renommer les classes obtenues dans cette partie par  $DWR1 \dots DWR6$  ou DWR constituent des régimes de temps de Banizoumbou, Cinzana et Mbour.

Le Tableau 5.14 montre ainsi les pourcentages de confusion entre les régimes de temps de Cinzana et Banizoumbou d'une part, puis Cinzana et Mbour d'autre part. Globalement, comme pour les données brutes (non désaisonnalisées), les pourcentages sur la diagonale du tableau sont plus élevés entre les pourcentages de Cinzana et Banizoumbou qu'entre les pourcentages de Cinzana et Mbour, signifiant de même mais à cette nouvelle échelle de temps, la plus proche similarité des régimes de temps entre les deux stations continentales qu'avec M'Bour sous l'influence océanique non seulement en terme de cycle saisonnier mais aussi en terme de variabilité interannuelle. Par contre, comparativement aux données brutes, les pourcentages sur la diagonale sont globalement plus faibles, ce qui est logique a priori car on peut penser que les structures propres au cycle saisonnier ont une cohérence spatiale plus élevée. Ainsi les régimes de temps 1, 2 et 4, à Cinzana et Banizoumbou montrent une cohérence mutuelle forte (une majorité d'observations appartenant par exemple au régime DWRB1 appartiennent aussi au régime DWRC1), mais le régime 3 n'a que 40% d'observations communes, et les deux derniers encore moins. On observe la même chose avec M'Bour.

Tableau 5.14 – Pourcentage de confusion des régimes de temps entre la station de Cinzana et Banizoumbou d'une part et de Cinzana et Mbour d'autre part.

a°)	WRB1	WRB2	WRB3	WRB4	WRB5	WRB6
WRC1	$54,\!29$	7,93	6,07	1,64	24,71	5,36
WRC2	8,97	$67,\!06$	6,41	0,00	0,59	16,96
WRC3	$16,\!14$	$5,\!50$	$41,\!71$	$15,\!50$	$15,\!57$	$5,\!57$
WRC4	7,12	0,00	$^{8,58}$	$53,\!66$	$24,\!65$	6,00
WRC5	28,10	$39,\!97$	$22,\!20$	$0,\!60$	$3,\!42$	5,70
WRC6	14,51	$5,\!91$	$33,\!38$	$13,\!41$	6,41	$26,\!38$
b°)	WRM1	WRM2	WRM3	WRM4	WRM5	WRM6
	WRM1 40,36	WRM2 20,50	WRM3 11,29	WRM4 6,93	$\frac{WRM5}{15,00}$	WRM6 5,93
$b^{\circ})$ $WRC1$ $WRC2$	WRM1 <b>40,36</b> 9,66	WRM2 20,50 44,58	$           WRM3 \\         11,29 \\         17,16     $	WRM4 = 6,93 = 0,30	$\frac{WRM5}{15,00} \\ 4,73$	$\frac{WRM6}{5,93}\\23,57$
$b^{\circ})$ $WRC1$ $WRC2$ $WRC3$	WRM1 <b>40,36</b> 9,66 16,93	WRM2 20,50 <b>44,58</b> 2,64	WRM3 11,29 17,16 <b>26,21</b>	$\frac{WRM4}{6,93} \\ 0,30 \\ 38,79$	$\begin{array}{c} WRM5 \\ 15,00 \\ 4,73 \\ 11,93 \end{array}$	$\begin{array}{c} WRM6 \\ 5,93 \\ 23,57 \\ 3,50 \end{array}$
$b^{\circ})$ $WRC1$ $WRC2$ $WRC3$ $WRC4$	WRM1 40,36 9,66 16,93 9,66	$\begin{tabular}{c} \hline WRM2 \\ \hline 20,50 \\ \bf 44,58 \\ 2,64 \\ 2,25 \end{tabular}$	WRM3 11,29 17,16 <b>26,21</b> 15,15	WRM4 6,93 0,30 38,79 <b>51,29</b>	$\begin{array}{r} WRM5 \\ 15,00 \\ 4,73 \\ 11,93 \\ 20,48 \end{array}$	$\begin{array}{c} WRM6 \\ 5,93 \\ 23,57 \\ 3,50 \\ 1,17 \end{array}$
	WRM1 40,36 9,66 16,93 9,66 34,94	WRM2           20,50           44,58           2,64           2,25           22,54	WRM3           11,29           17,16 <b>26,21</b> 15,15           11,54	WRM4 6,93 0,30 38,79 <b>51,29</b> 3,89	WRM5           15,00           4,73           11,93           20,48           14,82	$\begin{tabular}{c} \hline WRM6 \\ \hline 5,93 \\ 23,57 \\ 3,50 \\ 1,17 \\ 12,27 \end{tabular}$

### 5.3.1.2 Distributions diurnes à inter-annuelles

Dans cette partie, on vérifié aussi qu'un effet possible du cycle diurne sur l'attribution des régimes de temps est totalement éliminé et que comme pour les régimes de temps bruts on peut dire que les régimes de temps ne sont pas sensibles au cycle diurne d'occurrence. Ainsi en moyenne globale et en moyenne mensuelle, les types de temps ne traduisaient pas la variabilité diurne des conditions météorologiques dans les trois stations (cf figure A.3 en annexe).

La Figure 5.26 présente le cycle saisonnier d'occurrence des régimes de temps par stations. On voit sur cette figure, en comparant avec la Figure 5.6 des données brutes que le filtrage a fortement réduit l'effet de saisonnalité dans les occurrences de chacun des régimes de temps.

La Figure 5.27 montre l'évolution interannuelle d'occurrences des régimes de temps sur les trois stations sur chaque mois de l'année. La répartition de ces régimes de temps n'apparaît pas uniforme de 2006 à 2010, et globalement moins que pour les données brutes. Il semble aussi apparaître une corrélation entre l'évolution interannuelle de l'occurrence des régimes de temps de Banizoumbou et Cinzana à quelques différences près. Pour les deux stations continentales, l'occurrence du régime de temps DWR1 est forte pour les trois premières années de 2006 à 2007 d'octobre en mai et est faible entre 2009 et 2010 alors que le régime de temps 2 apparait faiblement les deux premières années 2006, 2007 sur l'ensemble des mois et fortement en 2009. Pour le type de temps DWR4, la variabilité interannuelle est très marquée sur l'ensemble des trois stations et semblent évoluer de la même façon. DWR4 apparaît être le type de temps le plus important de décembre 2009 à mai 2010 pour l'ensemble des trois stations et d'octobre 2006 à février 2007 uniquement à Cinzana. Pour la station de Banizoumbou, DWR5 est le type de temps le plus dominant entre avril et mais 2006 et 2008 à Mbour.



Figure 5.26 – Cycle saisonnier d'occurrence des types de temps en saison sèche pour les trois stations (Banizoumbou; Cinzan; Mbour).



Figure 5.27 – Evolution interannuelle d'occurrences des types de temps pour les trois stations (Banizoumbou; Cinzana; Mbour).

### 5.3.1.3 Persistances et transitions inter-classes

Le Tableau 5.15 suivant présente les probabilités de passer d'un régime de temps à un autre en un jour sur chacune des stations. Pour calculer les probabilités, on a d'abord converti notre base de données des régimes de temps horaire en base journalière en prenant pour chaque jour le régime de temps dominant. La probabilité de passer d'un type de temps
à un autre est alors obtenu par le rapport entre le nombre d'une telle transition sur le nombre total de transitions. La lecture du tableau se fait par ligne. Nous avons constaté que les régimes de temps sont nettement moins persistants en une journée que pour les données brutes, ce qui semble être lié à l'élimination du cycle saisonnier des données. Afin de mieux quantifier la persistance de chaque régime de temps des différentes stations, on a tracé la Figure 5.28 qui présente pour chaque station, la probabilité d'une classe de rester sur elle-même pendant un certain nombre de jours, c'est à dire la persistance des régimes de temps. Le zéro à l'abscisse correspond à la probabilité qu'un type de temps reste sur lui-même toutes les trois heures (base initiale), le 1 est la probabilité qu'un type de temps reste sur lui-même au moins une journée, etc, et le 5 la probabilité de rester sur lui-même durant 5 jours au moins. Nous remarquons globalement que plus on augmente le nombre de jours, plus la probabilité qu'une classe reste sur lui même diminue. Là aussi, la persistance est globalement plus faible que pour les données brutes où le cycle saisonnier est présent. Cependant pour un certain nombre de régimes de temps, après une décroissance les premiers jours, un plateau apparaît sur les jours suivants à des niveaux équivalents sur les 3 stations (valeurs entre 0.3 et 0.5). Ce plateau peut être du a la technique de transformation des données horaires aux données en un ou plusieurs jours (cf page 111).

Tableau 5.15 – Probabilité de passer d'une classe à une autre sur quatre jours pour les trois stations (a : Banizoumbou; b : Cinzana; c : Mbour).

a :Baniz	DWRB1	DWRB2	DWRB3	DWRB4	DWRB5	DWRB6
DWRB1	0,56	0,14	0,08	0,02	0,16	$0,\!05$
DWRB2	0,14	$0,\!31$	$0,\!25$	$0,\!08$	$0,\!04$	$0,\!18$
DWRB3	0,07	0,24	0,36	$0,\!14$	$0,\!09$	0,10
DWRB4	0,07	$0,\!12$	$0,\!12$	$0,\!56$	$0,\!09$	$0,\!04$
DWRB5	0,23	0,07	$0,\!07$	$0,\!14$	0,49	0,00
DWRB6	0,10	0,07	$0,\!30$	$0,\!20$	0,00	0,33
b :Cinzana	DWRC1	DWRC2	DWRC3	DWRC4	DWRC5	DWRC6
DWRC1	0,41	0,14	0,05	0,16	0,14	0,11
DWRC2	0,10	0,29	$0,\!13$	$0,\!23$	0,03	$0,\!23$
DWRC3	0,12	0,03	0,30	$0,\!12$	$0,\!27$	$0,\!15$
DWRC4	0,08	0,07	$0,\!08$	$0,\!52$	$0,\!12$	$0,\!13$
DWRC5	0,19	0,08	0,06	$0,\!13$	$0,\!29$	$0,\!25$
DWRC6	0,06	0,09	$0,\!13$	$0,\!20$	$0,\!13$	0,39
c :Mbour	DWRM1	DWRM2	DWRM3	DWRM4	DWRM5	DWRM6
DWRM1	0,46	0,16	0,09	0,16	0,06	0,06
DWRM2	0,28	0,23	$0,\!23$	0,03	$0,\!13$	$0,\!10$
DWRM3	0,12	0,08	0,32	$0,\!22$	$0,\!18$	0,08
DWRM4	0,19	0,06	$0,\!13$	$0,\!46$	$0,\!17$	0,00
DWRM5	0,09	0,19	$0,\!21$	$0,\!19$	$0,\!21$	$0,\!12$
DWRM6	0.04	0.13	0.35	0.22	0.09	0.17



Figure 5.28 – Persistance des régimes de temps sur 8 jours pour les trois stations (a : Banizoumbou; b : cinzana; c : Mbour).

Pour synthétiser les résultats présentés ci-dessus et pour mieux appréhender la signification de ces régimes de temps, nous avons appliqué comme dans la partie précédente une analyse factorielle des correspondances (AFC) associant les fréquences d'occurrence de chaque régime de temps pour chaque mois de chaque année, et permettant de projeter sur le premier plan factoriel les régimes de temps et les observations (mois et années). En effet, comme on l'a dit en introduction, le filtrage opéré élimine le cycle saisonnier moyen et conserve la variabilité interannuelle et intra-saisonnière sur les mois de Janvier à Mai et d'Octobre à Décembre de 2006 à 2010. On s'attend donc à ce que les 6 régimes de temps obtenus sur ces périodes représentent plutôt des anomalies d'échelles mensuelle et interannuelle. L'AFC devrait nous permettre de visualiser ces relations. Les pourcentages de variance concentrés sur les deux premiers axes factoriels (65.11% à Banizoumbou, 60.08% à Cinzana et 70% à Mbour) étant élevés (Tableau 5.16), cette visualisation est donc représentative de la proximité entre régimes de temps et anomalies d'échelles mensuelles à interannuelles. Par contre contrairement à la précédente analyse, l'axe 2 a un pourcentage de variance expliqué du même ordre de l'axe 1, lui donnant une importante équivalente. Les résultats obtenus sont aussi qualitativement cohérents avec ceux obtenus plus haut sur les transitions d'un régime à l'autre et sur la variabilité interannuelle des occurrences de ces régimes de temps.

Tableau 5.16 – Tableau des pourcentages de variance expliquée des deux premiers axes et leur somme pour chaque station.

	Axe factoriel 1	Axe factoriel 2	Somme
Banizoumbou	35.01%	30.1%	65.11%
Cinzana	34.72%	25.36%	60.08%
MBour	37.65%	31.95%	69.60%

Dans les figures qui suivent (Figure 5.29), nous présentons pour chaque station la projection des 6 régimes de temps de la station respective et la projection des mois de l'année, avec un plan de projection par année pour plus de clarté (on a réduit aussi pour cela la longueur des acronymes des régimes de temps). Ainsi pour une même station, les régimes de temps sont situés à la même place dans le plan factoriel pour chaque année.

Le premier élément à relever est l'organisation relative des différents régimes de temps. On observe à la fois pour Banizoumbou et Cinzana une proximité des régimes de temps 2, 3 et 6. Pour Banizoumbou, ces 3 régimes de temps s'opposent fortement aux régimes de temps 1 et 5 le long de l'axe 1 qui sont eux-mêmes opposés entre eux le long de l'axe 2. Le régime DB4 est proche de zéro sur le premier axe factoriel et relativement éloigné des autres régimes de temps le long du deuxième axe. A Cinzana, le contraste est moins marqué. DC2, DC3 et DC6 sont peu éloignés des autres régimes de temps suivant le premier axe factoriel et se distinguent des régimes DC5 et DC1 sur le deuxième axe, de moindre variance. DC4 est lui par contre relativement éloigné des autres régimes de temps le long du premier axe. On observe ainsi une cohérence globale dans la répartition des régimes de temps entre ces deux stations. Sur M'Bour, globalement la répartition des régimes de temps est différente des deux autres stations. Il n'y a pas de forte proximité sur un groupe de régimes de temps mais une distribution assez homogène sur le plan factoriel dont les axes totalisent environ 70% de variance expliqué. On peut noter cependant une proximité DM1-DM2, DM3-DM6, un isolement de DM4 et la position au centre du plan de DM5.

Si on considère maintenant les répartitions des différents mois de chaque année, on peut noter un certain nombre de regroupements autour d'un régime de temps individuel ou d'un groupe de régimes de temps, ces regroupements variant d'une année à l'autre.

Ainsi pour Banizoumbou, certaines années sont très structurées. Par exemple en 2010, la première partie de l'année est regroupée majoritairement autour de DB4 et la seconde partie de l'année autour du groupe DB2-DB3-DB6, les régimes de temps restants DB1et DB5 étant isolés, donc relativement absents en terme d'occurrences. Ainsi on pourra associer une forte contribution de DB4 aux anomalies interannuelles du début de l'année 2010, ainsi qu'une forte contribution des régimes DB2, DB3 et DB6 à celles de la fin de l'année. Les autres années sont structurées différemment. 2006 est centrée soit sur le groupe DB2-DB3-DB6, soit sur DB1, et dans une moindre mesure DB5; DB4 n'a alors pas de forte représentativité. 2008 est assez similaire à 2006. 2009 est assez fortement structurée autour des régimes DB2-DB3-DB6 de Janvier à Avril ainsi qu'Octobre. Enfin 2007 ne montre pas de forte structuration mais des mois répartis de manière relativement homogène dans le plan factoriel.

Concernant Cinzana, l'année 2010 a une structuration très similaire à Banizoumbou, avec un regroupement des mois équivalent autour du groupe DC2-DC3-DC6 et autour de DC4, et un isolement des régimes DC1 et DC5. 2009 a une certaine similarité avec 2010 par la concentration autour du groupe DC2-DC3-DC6. 2006 et 2008 montrent quelques aspects similaires à Banizoumbou (isolements de certains régimes de temps, regroupements partiellement similaires), et on retrouve pour 2007 la distribution relativement homogène des mois dans le plan factoriel.

Sur M'Bour, les observations (mois par année) sont globalement distribuées de manière plus homogène que pour les deux autres stations, montrant un contraste inter-régimes de temps et une prééminence de certains régimes de temps moins marquée sur plusieurs mois. Cependant 2010 montre un regroupement des mois de Janvier à Mars autour du régime DM4.On peut noter cependant suivant les années l'isolement de certains régimes de temps.







Figure 5.29 – Projection sur le premier plan factoriel année par année des régimes de temps de chaque station.

La Figure 5.30 suivante présente la projection de l'ensemble des types de temps des trois stations année par année, de manière à apporter une vision plus globale régionalement. Les pourcentages de variance expliquée sont de 30% pour le premier axe et de 25% pour le deuxième, donc assez similaires à ceux des projections par station. En accord avec les remarques précédentes sur les similitudes entre Banizoumbou et Cinzana, on retrouve le regroupement B2 - B3 - B6 - C2 - C3 - C6 auquel s'associent les régimes M3 et M6 de M'Bour. On observe un autre regroupement B4 - C4 - M4 à l'opposé du premier axe factoriel, ainsi qu'un regroupement B1 - C1 - M1 sur le deuxième axe. Enfin on a un regroupement C5 - M2 et un relatif isolement des régimes B5 et M5.

Année par année, comme vu précédemment la contribution des différents régimes est variable. On retrouve par exemple la particularité de 2010 avec une forte concentration des mois autour de deux groupes de régimes, B4 - C4 - M4 et B2 - B3 - B6 - C2 - C3 - C6 - M3 - M6, et en conséquence la faible contribution des autres régimes. On peut noter aussi l'éloignement relativement fréquent du groupe B4 - C4 - M4 (en 2006, 2007, 2008).

On a voulu faire une dernière analyse en intégrant le rôle possible de l'indice NAOdans l'interprétation de cette variabilité interannuelle. L'oscillation nord-atlantique (ou North Atlantic Oscillation en anglais, d'où le sigle NAO) est un phénomène atmosphérique et océanique, qui concerne principalement l'Atlantique Nord. Il décrit les variations du régime océan-atmosphère sur la région et se mesure généralement comme une différence de pression atmosphérique entre l'Anticyclone des Açores et la dépression d'Islande, en prenant l'anomalie de pression. *Chiapello and Moulin* [2002] ont étudiés la variabilité du transport de la poussière saharienne dessus de l'Atlantique Nord tropical en utilisant les données combinées de TOMS/Nimbus7 et Meteosat sur près de 20 années c'est à dire de 1979 à 1997. Ainsi, ils ont montré que l'occurrence interannuelle du transport des poussières pendant l'hiver est liée à la variation interannuelle l'Indice NAO.

Afin de voir le rôle possible de l'indice NAO dans ce travail, on a d'abord divisé les données de l'indice NAO de 2006 à 2010 à notre disposition en trois groupes, c'est à dire les données supérieures à 0.5, les données comprises entre -0.5 et 0.5 et les données inférieures à -0.5. Puis pour chaque groupe de données de NAO, on regarde le type de temps dominant pour chaque mois et on lui affecte la valeur du NAO. Ainsi à la fin on compte le nombre de données affectées à chaque type de temps selon le groupe de données, ce qui constitue les données projetées. Le Tableau 5.17 présente l'occurrence de chaque groupe de NAO dans chaque régime de temps pour chacun des trois stations.

Pour la station de Banizoumbou, nous remarquons que NAO1 et NAO2 sont beaucoup plus liées à DB1 que NAO3. En plus de DB1, NAO2 est liée à DB4. Par contre NAO3est beaucoup plus liée à DB5, DB3 et DB1. Pour la station de Cinzana, NAO1 est plus liée à DC1 et DC2, NAO2 plus liée à DC4, DC5 et DC6 et NAO3 est plus liée à DC4et DC6. Quand à la station de Mbour, nous remarquons que DM6 n'est liée à aucune NAO. Cependant, NAO1 et NAO2 sont plus liés à DM3 que NAO2 et que NAO2 met plus en avant DM1 et DM4.



Figure 5.30 – Projection de l'ensemble des types de temps des trois stations année par année.

a :Baniz	DWRB1	DWRB2	DWRB3	DWRB4	DWRB5	DWRB6
NAO1	5	1	2	1	1	0
NAO2	5	1	3	5	0	2
NAO3	4	1	4	0	4	1
b :Cinzana	DWRC1	DWRC2	DWRC3	DWRC4	DWRC5	DWRC6
NAO1	3	1	0	1	2	3
NAO2	1	1	2	5	4	3
NAO3	2	1	2	3	1	5
c :Mbour	DWRM1	DWRM2	DWRM3	DWRM4	DWRM5	DWRM6
NAO1	2	2	3	2	1	0
NAO2	4	1	1	9	1	0
NAO3	1	3	6	2	2	0

Tableau 5.17 – Relation entre les régimes de temps et la NAO dans chacune des trois station. (a : Banizoumbou; b : Cinzana; c : Mbour).

NAO1: données > à 0.5 ; NAO2: données < à -0.5 ; NAO3: données compris entre -0.5 et 0.5

Si on s'intéresse surtout aux valeurs NAO1 et NAO2, ce qu'on peut dire, c'est que NAO1 met en avant DB1, NAO2 DB1, DB4 et DB3 à la station de Banizoumbou. Pour la station de Cinzana, NAO1 met en avant DC1 et DC6 et NAO2, DC4 et DC5. Pour la station de Mbour, NAO1 met en avant DM3 et NAO2, DM4. Pour toutes les trois stations, NAO2 met en avant le régime de temps 4 qui se caractérise par une forte variation du cycle interannuel d'occurrence (Figure 5.27) avec la plus forte occurrence obtenue en 2010. Ce qu'on voit surtout sur les séries des valeurs NAO (non montré), c'est la phase négative de 2010, et on a vu que 2010 à une structuration particulière sur le plan factoriel, s'associant fortement aux régimes DB4 - DC4 - DM4 de Janvier à Mai 2010.

#### 5.3.1.4 Structures spatiales et interprétations météorologiques et climatiques

La Figure 5.31 suivante présente la climatologie de vents (vecteurs), géopotentiel (isolignes) et vitesse verticale (couleur) (500 hPa); vents (vecteurs) et humidité (700 hPa); vents (vecteurs) et humidité (925 hPa) des données brutes sur la zone d'étude. Nous remarquons qu'à 500 hPa, on a une zone de hauts géopotentiels sur toute la région, orientée zonalement, dont l'axe du maximum est situé vers 16°N organisant une circulation anticyclonique. Les verticales vitesses ascendantes sont localisées dans la zone Sud (golfe de Guinée) tandis que les valeurs positives (subsidence) sont localisées sur les latitudes les plus au nord. Dans les plus basses couches, à 700 hPa, la circulation des vents est similaire à celle obtenue à 500 hPa, et est associée à un gradient d'humidité vers le sud ; à 925 hPale gradient d'humidité est plus fort qu'à 700 hPa et associé à un flux d'harmattan au nord de 12°N et un flux de sud le long de la côte guinéenne.

Les champs de géopotentiel à 500 hPa pour les différents régimes de temps et les trois stations (Figure 5.32, le point sur la carte représente la station de mesure des données insitu) présentent une anomalie de circulation, soit cyclonique, soit anticyclonique, centrée sur le nord du domaine présenté. Ceci correspond à une modulation du champ climatologique moyen présenté au-dessus, se traduisant soit par une extension vers le sud des bas géopotentiels des latitudes moyennes, soit par un renforcement des hauts géopotentiels présents en moyenne sur le domaine. Ces structures atmosphériques sont associées à un forçage dynamique cohérent avec ce que l'on connaît des ondes atmosphériques de Rossby, à savoir un forçage ascendant dans la partie Est des thalwegs (vents de sud) et un forçage subsidant dans la partie Est des dorsales (vents de nord). Ces structures sont très cohérentes entre les régimes de Banizoumbou et de Cinzana. A M'Bour les structures sont un peu différentes des deux autres stations avec un centrage plus à l'ouest des structures d'anomalies.

Les Figure 5.33 et Figure 5.34 présentent la projection des différentes variables sur les types de temps pour chacune des trois stations : vent et humidité spécifique à 700 hPa, vent et humidité spécifique à 925 hPa. A 700 hPa, les structures de circulation sont très similaires à celles de 500 hPa, avec des zones plus humides dans les vents de sud et plus sèches dans les vents de nord. A 925 Pa, les structures de circulation sont là encore similaires à celles des niveaux supérieurs; par contre les anomalies d'humidité spécifique sont moins étendues et localisées sur la partie sud. A 925 hPa, DWR2 et DWR6 sont les régimes de temps les plus humides pour les station de Cinzana et Banizoumbou. Pour la station de Mbour, DWR4 et DWR6 sont les plus humides. Ici aussi, on retrouve la dynamique régionale observée sur les résultats de l'AFC avec une forte ressemblance des structures spécifiques des régimes de temps DWRB1 - DWRC1 - DWRM1, DWRB4 - DWRC4 - DWRM4, DWRB5 - DWRC5 - DWRM5 sur l'ensemble des trois stations.

Si on revient aux tableaux de probabilité de passage d'une classe à une autre (Tableau 5.14), les probabilités de passage d'une classe donnée à une classe différente sont relativement faibles, probablement car on décrit des structures liées à la variabilité interannuelle, donc avec une forte persistance intra-classe. On peut cependant relever une certaine cohérence entre les probabilités de passage les plus fortes d'une classe à une autre et une continuité des structures spatiales de ces classes, mais pas systématiquement. Par exemple pour Banizoumbou (Figure 5.32), les probabilités les plus fortes concernent le passage de DWRB6 à DWRB3 (p=0,30) avec un déplacement vers l'ouest de la circulation cyclonique localisée au nord du domaine. De même pour la transition DWRB5 vers DWRB1(p=0,19). Par contre le passage DWRB6 vers DWRB4 (p=0,20) se caractérise dans un renversement du sens de la circulation atmosphérique.

Comparativement à la classification sur les données brutes, on voit là que ce n'est pas la structuration du cycle saisonnier qui apparait, mais des régimes de temps caractéristiques de modulation d'ondes synoptiques qui peuvent être représentatifs des anomalies de circulation de certains mois de certaines années comme on l'a montré par les résultats de l'AFC.



Figure 5.31 – Cartes moyennes : (A : la Vitesse verticale (hPa/s) à 500 hPa (couleur), de Géopotentiel (mgb) à 500 hPa (isolignes) et de vents (m/s) à 500 hPa; B : de Vents (m/s) et Humidité (g/kg) à 700 hPa; C : de Vents (m/s) et Humidité (g/kg) à 925 hPa).







Figure 5.32 – Anomalie de la Vitesse verticale (hPa/s) à 500 hPa (couleur), de Géopotentiel (mgb) à 500 hPa (isolignes) et de vents (m/s) à 500 hPa (Vecteurs) pour les trois stations. (Couleur marron : valeurs négatives de vitesse verticale, ascendance; Couleur verte : valeurs positive de vitesse verticale).







Figure 5.33 – carte moyenne des anomalies de Vents (m/s) et Humidité (g/g) à 700 hPa pour les trois stations (A : Banizoumbou; B : Cinzana; C : Mbour).







Figure 5.34 – carte moyenne des anomalies de Vents (m/s) et Humidité (g/g) à 925 hPa pour les trois stations (A : Banizoumbou; B : Cinzana; C : Mbour).

Dans cette partie, comme pour la partie des régimes de temps bruts, nous avons voulu caractériser les régimes de temps par rapport à l'AI, mais ici en terme d'anomalie interannuelle. Ainsi on a projeté les données d'anomalies de l'AI sur les classes obtenues. La Figure 5.35 présente pour chaque régime de temps, les anomalies moyennes de géopotentiel à 925 hPa et de l'anomalie moyenne l'Aérosol Index » sur les trois stations. Les structures de circulation étant très différentes entre les régimes de temps bruts et désaisonnalisés, la répartition des anomalies de poussières le sont aussi. Les signaux d'anomalies de poussières ne sont pas toujours très nets suivants les régimes de temps mais on observe une certaine cohérence entre vents d'origine océanique (resp. continentale) et déficit (resp. excédent) de poussières. De plus on note une assez bonne cohérence entre les stations de Banizoumbou et Cinzana, les signaux étant plus nets pour Banizoumbou. Ainsi les régimes de temps comportant une composante de vent d'ouest et sud-ouest (WR2, WR3, WR4 et WR6) montrent une anomalie négative de contenu en poussière, l'inverse (vent d'est et anomalie positive de poussière) étant présent pour le régime WR5 sur Banizoumbou et dans une moindre mesure pour les régimes WR1 au-dessus de la côte ouest africaine et WR4. Sur M'Bour, les régimes WR3, WR5 et WR6 montrent une structuration et une distribution en poussières similaires à ceux de Banizoumbou. WR4 a une structure assez cohérente aussi avec un renforcement du contenu en poussières en lien avec une circulation d'est et nord-est renforcée; la structuration inverse est observée pour WR2 (et dans une moindre mesure pour WR1).







Figure 5.35 – Anomalie de l'indice d'aérosol (couleur), de géopotentiel (mgb) à 925 hPa (isolignes) et de vents pour les trois stations Banizoumbou, Cinzana et Mbour.

#### 5.3.2 Relations AOT- $PM_{10}$ avec et sans régimes de temps

Comme pour l'étude précédente, nous avons regroupé les données météorologiques en situations météorologiques présentant les mêmes caractéristiques appelées régimes de temps. Nous présentons ici aussi la relation AOT- $PM_{10}$  en anomalie dans chaque régime de temps afin de voir l'influence des variables atmosphériques sur la relation mais en anomalie. La Figure 5.36 suivante présente les nuages des points et les corrélations entre anomalies AOT et  $PM_{10}$  sur toute la base avant séparation en type de temps pour la base de données obtenue en filtrant les données par enlèvement du cycle saisonnier moyen sur chaque station. Cette dernière permet d'avoir une idée de la relation AOT- $PM_{10}$ dans chaque station avant l'introduction de la notion de régimes de temps. La corrélation est beaucoup plus faible à Mbour (0.59) qu'à Cinzana (0.74) et Banizoumbou (0.75). Il est intéressant de noter que ces corrélations sont du même ordre de grandeur que celles obtenues pour les données brutes (Figure 5.20; 0.51 à M'Bour, 0.72 à Cinzana, 0.69 à Banizoumbou).



Figure 5.36 – Nuages des points et corrélations entre anomalies AOT et  $PM_{10}$  sur toute la base avant séparation en régimes de temps.

Afin d'étudier la relation AOT- $PM_{10}$ , la base de données d'étude de cette dernière, c'est à dire la base de données météorologiques, pour laquelle pour chaque observation on connaît la valeur d'AOT et de  $PM_{10}$  a été projetée sur les cartes obtenues et réparties entre les six classes en associant à chaque valeur d'AOT et de  $PM_{10}$  le régime de temps associé à sa situation météorologique. Ainsi pour voir l'impact de la notion de « régime de temps », sur la relation AOT- $PM_{10}$ , nous avons étudié l'amélioration qui en résultait lorsque l'on fait des régressions pour chaque régime de temps.

La Figure 5.37 présente la relation AOT- $PM_{10}$  dans chaque régime de temps. Si l'on compare la Figure 5.37 et la Figure 5.36 obtenue sans faire intervenir la notion de type de temps, on observe une amélioration de la régression dans certains régimes de temps et une détérioration dans d'autres. Ainsi à Banizoumbou, seul DWR6 détériore la corrélation alors que les autres régimes de temps ont des performances équivalentes ou supérieures. Pour la station de Cinzana, DWR2 n'améliore pas la corrélation, celles ci étant de même ordre ou supérieures pour les autres régimes. Pour la station de Mbour les corrélations de DWR1 et DRW6 sont dégradées, et seul DRWM2 a une corrélation nettement supérieure.

Dans cette partie aussi comme dans la partie concernant les régimes de temps bruts,

nous avons voulu voir si les régimes de temps pouvaient être différenciés à travers les pentes des régressions en traçant la Figure 5.38. Globalement pour les trois stations, nous remarquons qu'il existe des classes caractéristiques de pentes faibles et des classes caractéristiques de pentes plus élevées. Pour les stations de Cinzana et Banizoumbou, c'est le régime de temps DWR6 qui a la pente la plus faible avec 229,8  $\mu g.m^{-3}$  pour DWRB6 et 298,48  $\mu g.m^{-3}$  pour DWRC6. La pente la plus forte est obtenue sur DWRB5 (626,6  $\mu g.m^{-3}$ ) et DWRC1 (543,9 $\mu g.m^{-3}$ ). Cependant pour la station de Mbour la pente la plus basse est obtenue sur DWRM6 (236,9  $\mu g.m^{-3}$ ) et la plus haute est obtenue sur DWRM2 (466,9  $\mu g.m^{-3}$ ). Il apparaît que, pour une valeur d'AOT donnée, la valeur de  $PM_{10}$  correspondante pour les classes ayant la pente la plus forte peut aller jusqu'à deux fois plus que la valeur de  $PM_{10}$  correspondant aux classes ayant la pente la plus faible. Par exemple pour une valeur d'AOT égale à 1, pour les stations de Cinzana et Banizoumbou nous avons respectivement DWRC6 a une valeur de (244  $\mu g.m^{-3}$ ) alors que DWRC1 a une valeur de (622  $\mu g.m^{-3}$ ) et DWRB6 (198  $\mu g.m^{-3}$ ) et DWRB5 (718  $\mu g.m^{-3}$ ).



DWRB 1	DWRB 2	DWRB 3	DWRB 4	DWRB 5	DWRB 6	Sans DWR
0,80	0,73	0,74	0,70	0,89	0,41	0,75

Banizoumbou





Figure 5.37 – Nuages des points et les corrélations entre AOT et  $PM_{10}$  dans chaque classe sur les trois stations.



Figure 5.38 – Droite de régression entre les anomalies de  $PM_{10}$  et AOT pour chaque régime de temps

#### 😰 Estimation PM à partir de l'AOT et les variables météorologiques

Nous avons constaté à partir de l'étude précédente que la relation entre les  $PM_{10}$  et l'AOT en terme de corrélation n'est pas toujours meilleure dans certaines classes par rapport à celle obtenue sans les régimes de temps. Ainsi nous avons décidé d'utiliser une méthode non linéaire pour la prévision de la concentration en  $PM_{10}$  à partir de l'AOT.

L'algorithme des cartes auto-organisatrices est alors utilisé ici aussi dans un but de prédiction. Nous avons donc utilisé les cartes auto-organisatrices pour prévoir la concentration massique en  $PM_{10}$  à partir des AOT et cela pour chaque régime de temps, afin d'affiner la relation AOT- $PM_{10}$  que nous voulons déterminer. Nous avons ainsi utilisé comme variables d'entrée de la carte auto organisatrice, la variable d'AOT et les variables météorologiques locales (température, vitesse du vent, direction du vent, humidité, coefficient d'angström) pour chaque classe, ceci afin de modéliser la relation AOT- $PM_{10}$  pour un régime de temps spécifique et une localisation précise. On a donc réalisé six cartes auto organisées et ainsi, chaque neurone est associé à une moyenne de  $PM_{10}$ . De cette manière, on obtient une fonction non paramétrique permettant de prédire la relation AOT- $PM_{10}$ .

Pour chaque classe, nous avons entrainé la carte auto organisatrice qui lui est dédiée à l'aide des données d'apprentissage qui lui sont dédiées et tester les performances sur les données de test dédiées. La Figure 5.39 suivante présente les nuages des points entre  $PM_{10}$ estimée et  $PM_{10}$  mesurée sur la base de test et les corrélations entre  $PM_{10}$  estimée et  $PM_{10}$ mesurée aussi bien sur la base de test que celle d'apprentissage. Les corrélations entre  $PM_{10}$ estimée et  $PM_{10}$  mesurée varient de 0.74 à 0.94 sur la base d'apprentissage et de 0.36 à 0.87 sur la base de test sur l'ensemble des trois stations. Sur la base d'apprentissage, les plus faibles corrélations sont obtenues au niveau du régime de temps DWR6 à Banizoumbou (0.74) et à Cinzana (0.77) et du régime de temps DWR4 à Mbour (0,74). De même sur la base de test la plus faible corrélation est obtenue sur DWR6 à Banizoumbou avec une corrélation de 0.52 et DWR5 à Cinzana avec une corrélation de 0.61. Pour la station de Mbour, la plus faible corrélation est obtenue sur le régime de Temps DWR6 (0,36). Afin de voir la relation  $PM_{10}$  estimée- $PM_{10}$  mesurée obtenue sur le processus des régimes de temps sur l'ensemble de la base, on a regroupé les résultats  $PM_{10}$  estimée- $PM_{10}$  mesurée dans chaque régime de temps puis on a calculé la corrélation sur la base d'apprentissage et de test pour l'ensemble des trois stations. La relation entre AOT- $PM_{10}$  estimée obtenue dans chaque régime de temps est présenté ci dessous (Tableau ??) pour chaque station. On obtient ainsi d'excellentes performances pour la modélisation sur une localisation spécifique de  $PM_{10}$  intégrant aussi le régime de temps présent, excepté pour M'Bour sur la base test.







Figure 5.39 – Les Nuages de points entre PM estimée et PM mesurée dans chaque classe sur chacune des trois stations (Banizoumbou; Cinzana; Mbour).

Tableau 5.18 – Corrélations linéaire entre  $PM_{10}$  estimé  $PM_{10}$  mesuré du test et de l'apprentissage pour les trois stations.

	Apprentissage	Test
Bani	$0,\!87$	0,75
Cinzana	$0,\!87$	0,72
Mbour	0,81	$0,\!53$

### 5.3.3 Travail exploratoire sur un système d'alerte opérationnel

Dans la partie précédente, de manière similaire à l'approche de Yahi et al. [2013], on a montré les liens entre les régimes de temps et les contenus en poussières à une échelle de temps fortement contrôlée par la saisonnalité. Ici, on a fait de même après avoir éliminé le cycle saisonnier moyen, mettant en exergue la variabilité résiduelle comprenant principalement la variabilité interannuelle mais aussi la variabilité intra-saisonnière. On a montré que les régimes de temps avaient une structure atmosphérique faisant penser à des perturbations d'échelle synoptique mais représentatives d'anomalies de certains mois de certaines années (résultats de l'AFC), et qu'ils montrent une bonne cohérence avec les champs d'anomalies de contenus en poussières.

On est ainsi en présence de deux types d'informations (saisonnière et désaisonnalisée) montrant des liens entre régimes de temps et contenus en poussières. Nous avons exploré la possibilité de les combiner pour évaluer l'information statistique que l'on pourrait en extraire dans le cadre d'un système d'alerte opérationnel appliqué aux évènements de contenus les plus forts de poussière pour chacune des trois stations. La première étape est simplement la connaissance du cycle saisonnier qui fait que certains mois (représentés par certains des régimes de temps bruts) ont un contenu en poussière plus élevé que d'autres mois. Ce résultat est peu informatif. On peut alors s'intéresser à l'occurrence des régimes de temps interannuels pour chacun des régimes de temps bruts et voir si certaines combinaisons peuvent être particulièrement favorables à l'occurrence d'épisodes à forts contenus en poussière. De plus comme les données interannuelles ne nécessitent que d'enlever la valeur du cycle saisonnier moyen jour à jour, on peut utiliser cette approche en temps réel et donc l'insérer dans un système d'alerte opérationnel.

Pour tester cette approche on s'est fixé le seuil de 15% des valeurs de  $PM_{10}$  les plus élevées sur l'échantillon total pour chacune des stations (Tableau 5.19), ceci à partir des données de  $PM_{10}$  brutes (dans ce cas on sélectionne les contenus les plus forts) ou bien à partir des données désaisonnalisées (dans ce cas on sélectionne les anomalies les plus fortes par rapport au cycle saisonnier moyen).

$PM_{10}$ seuils 15% max $(\mu g/m^3)$	Brut	Désaisonnalisé
Bani	4134	2877
Cinzana	3019	1885
Mbour	2637	1853

Tableau 5.19 – Les seuils de  $PM_{10}$  pour chaque filtrage et pour chacune des trois stations.

Cette approche est présentée maintenant en détail pour les données  $PM_{10}$  brutes des trois stations. Sur l'ensemble de l'échantillon disponible pour chaque station, (2960 cas) de la station de Banizoumbou on identifie 2288 cas pour la station de Cinzana et 2545 cas pour la station de Mbour. Le nombre de cas correspondant (à titre de test) aux 15% des valeurs les plus élevés de  $PM_{10}$  est 444 cas pour la station de Banizoumbou, 341 cas pour la station de Cinzana et 378 cas pour la station de Mbour. Pour chacun de ces cas on identifie le régime de temps brut (WR) et le régime de temps interannuel (DWR)correspondant. On construit ensuite le tableau de contingence correspondant pour tout l'échantillon (Tableau 5.20) et cela pour chaque station. Dans ce tableau (par exemple pour la station de Banizoumbou), la première case de la première ligne, « 2/87 », signifie qu'on a 2 cas où on a observé la concentration en  $PM_{10}$  dans les 15% les plus forts, sur 87 cas où le régime WR1 est présent en même temps que le régime DWR1. On fait de même pour les autres régimes DWR et pour les autres stations. Cependant la dernière case de chaque ligne représente les totaux de la ligne. Le tableau suivant (Tableau 5.21) montre les mêmes résultats exprimés en pourcentages d'occurrence (on a effectué les divisions dans chaque case).

Tableau 5.20 – Tableau de confusion entre les classes de la base de données brutes correspondant aux maximums de  $PM_{10}$  à 15% et celles des la bases de données désaisonnalisées correspondante par rapport au totale de confusion sur l'ensemble de la base de données pour les trois stations.

Baniz	DWRB1	DWRB2	DWRB3	DWRB4	DWRB5	DWRB6	Total
WRB1	2/87	1/52	0/45	0/14	0/8	2/68	5/274
WRB2	10/98	3/42	7/112	6/142	0/25	5/150	31 / 569
WRB3	22/95	34/130	20/160	4/54	9/23	3/39	92 / 501
WRB4	24/112	5/36	13/111	15/172	76/212	1/8	134 / 651
WRB5	39/170	14/148	9/74	0/9	15/55	0/19	77 / 475
WRB6	32/126	1/20	13/92	26/150	33/81	0/18	105 / 487
Total	129/688	58/428	62/594	51/541	133/404	11/302	444 / 2957=15%
Cinzana	DWRC1	DWRC2	DWRC3	DWRC4	DWRC5	DWRC6	Total
WRC1	0/41	0/12	0/27	0/60	1/62	1/125	2/327
WRC2	10/7	0/4	3/18	5/61	0/5	7/88	15/183
WRC3	3/51	4/43	12/67	10/72	23/90	19/120	71/443
WRC4	37/108	2/5	22/91	33/183	1/24	5/47	100/458
WRC5	28/133	9/126	3/29	4/34	7/66	5/84	56/472
WRC6	2/19	0/4	10/62	55/137	13/94	17/89	97/405
Total	70/359	15/194	50/294	107/547	45/341	54/553	341/2288 = 15%
Mbour	DWRM1	DWRM2	DWRM3	DWRM4	DWRM5	DWRM6	Total
WRM1	2/57	1/18	1/20	1/15	0/19	0/4	5/133
WRM2	18/202	1/33	4/177	29/142	13/78	0/18	65/650
WRM3	42/197	9/37	6/23	39/158	2/15	0/9	98/439
WRM4	6/48	5/39	15/60	18/75	20/64	4/27	68/313
WRM5	5/50	39/154	8/68	0/24	11/64	11/104	74/466
WRM6	7/80	7/60	11/154	12/82	23/97	8/71	68/544
Total	80/634	62/343	45/502	99/496	69/337	23/233	378/2545=15%

Tableau 5.21 – Pourcentage de données des classes de la base de données brutes correspondant aux maximums de  $PM_{10}$  à 15% contenu dans les classes de la base de données désaisonnalisées correspondantes.

Baniz	DWRB1	DWRB2	DWRB3	DWRB4	DWRB5	DWRB6	Total
WRB1	2,3	1,9	0	0	0	2,9	1,8
WRB2	10,2	$^{7,1}$	$6,\!3$	$^{4,2}$	0	$^{3,3}$	5,4
WRB3	23,2	26,2	12,5	$^{7,4}$	39,1	$^{7,7}$	18,4
WRB4	21,4	$13,\!9$	11,7	$^{8,7}$	$35,\!8$	12,5	20,6
WRB5	22,9	$^{9,5}$	12,2	0	$27,\!3$	0	16,2
WRB6	25,4	5	14,1	$13,\!3$	40,7	0	21,6
Cinzana	DWRC1	DWRC2	DWRC3	DWRC4	DWRC5	DWRC6	Total
WRC1	0	0	0	0	$1,\!6$	0,8	0,6
WRC2	0	0	16,7	$^{8,2}$	0	$7,\!9$	8,2
WRC3	$5,\!9$	9,3	$17,\!9$	$13,\!9$	$25,\!6$	$15,\!8$	16,0
WRC4	34,2	40,0	24,2	18,1	$^{4,2}$	$10,\!6$	21,8
WRC5	21,0	$^{7,1}$	10,3	$11,\!8$	$10,\!6$	$5,\!9$	11,9
WRC6	10,5	0	16,1	40,1	$13,\!8$	19,1	23,9
Mbour	DWRM1	DWRM2	DWRM3	DWRM4	DWRM5	DWRM6	Total
WRM1	3,5	$5,\!6$	$5,\!0$	6,7	0,0	0,0	3,8
WRM2	8,9	$^{3,0}$	$^{2,3}$	20,4	16,7	$0,\!0$	10,0
WRM3	21,3	24,3	26,1	24,7	$13,\!3$	$0,\!0$	22,3
WRM4	12,5	$12,\!8$	25,0	$24,\!0$	$31,\!3$	$14,\!8$	21,7
WRM5	10,0	25,0	$11,\!8$	$0,\!0$	17,2	$10,\!6$	15,9
WRM6	8,8	11,7	$^{7,1}$	$14,\! 6$	23,7	$11,\!3$	12,5

On peut interpréter ces résultats de la manière suivante. Globalement on a, par définition 15% de chance de se trouver en présence d'un contenu  $PM_{10}$  supérieur au seuil de 4134  $\mu g/m^3$  pour la station de Banizoumbou, 3019  $\mu g/m^3$  pour Cinzana et 2637  $\mu g/m^3$  pour la station de Mbour. Si on considère l'information donnée par les régimes WR seuls (Tableau 5.21, colonne de droite) pour la station de Banizoumbou, on voit que la probabilité de dépasser ce seuil est globalement minime pour les régimes WR1 et WR2 (1,8% et 5,4%) et nettement plus élevé pour les autres régimes (compris entre 16% et 22%). Pour la station de Mbour, on voit également que la probabilité de dépasser le seuil de 2637  $\mu g/m^3$ 3 est globalement faible pour les régimes WR1 et WR2 (3,8% et 10%) par rapport aux autres WR, mais un peut plus élevée que WR1 et WR2 (0,6% et 8,2%) de la station de Cinzana. Pour les WR3 - 6, les probabilités de dépasser le seuil de 3019  $\mu g/m^3$  pour la station de Cinzana varient entre (de 11% à 23%) et celles de dépasser le seuil de 2637  $\mu g/m^3$  pour la station de Mbour varient entre (12,5% et 22,3%).

Ceci est cohérent avec l'analyse faite sur ces régimes de temps caractérisant soit les mois de début/fin de saison de mousson, soit les mois de saison sèche pour l'ensemble des trois stations. Si on considère ensuite la modulation de ces probabilités par les régimes DWR (Tableau 5.21, autres cases), l'information devient alors beaucoup plus discriminante. Pour la station de Banizoumbou, on voit que pour les régimes WR5 et WR6, il existe des configurations de régimes DWR où la probabilité est minime (<5%) bien que l'on soit dans des régimes WR propices à de forts contenus en poussières. Cette configuration de faible probabilité peut être observée sur les régimes de temps (WR6 - DWR2) de la station de Cinzana et (WR5 - DWR4) de la station de Mbour propices à de forts contenus en poussières. Inversement, pour chacune des stations, d'autres configurations font monter les probabilités jusqu'à 40%, largement supérieurs à la probabilité moyenne de 15%. Globalement, les contrastes (contenus faibles et contenus élevés) apparaissent plus

nets pour les deux stations continentales que pour Mbour.

La même approche est réalisée sur les trois stations (Banizoumbou, Cinzana et Mbour) mais cette fois ci sur les anomalies désaisonnalisées de  $PM_{10}$ . Le seuil d'anomalie de  $PM_{10}$  à 15% est maintenant de 2877  $\mu g/m^3$  à Banizoumbou, 1885  $\mu g/m^3$  à Cinzana et 1853  $\mu g/m^3$ à Mbour. On obtient les tableaux suivants (Tableau 5.22 et Tableau 5.23). On voit dans ce cas que les régimes désaisonnalisés ayant la plus grande probabilité d'avoir des épisodes de  $PM_{10}$  élevés sont (Tableau 5.23, colonne de droite) DWR1(18,6%) et DWR5 (12,9%) à Banizoumbou, DWR1(23,7%) et DWR3 - DWR4 (19,7%) à Cinzana, et DWR4 (22,0%) et DWR5 (21,4%) à Mbour, et la plus faible probabilité pour DWR6 (6%) à Banizoumbou, DWR2 (6,7%) à Cinzana et DWR3 (9,6%) à Mbour, ce qui n'est relativement pas très discriminant pour chaque station. Si on tient compte de la combinaison avec les régimes WR, on obtient une meilleure discrimination avec un certain nombre de configurations de probabilités minimes (<5%) et de probabilités allant jusqu'à 43,5%.

Tableau 5.22 – Tableau de confusion entre les classes de la base de données désaisonnalisées correspondant aux maximums de  $PM_{10}$  à 15% et celles des la bases de données brutes correspondante par rapport au totale de confusion sur l'ensemble de la base de données pour les trois stations .

Baniz	WRB1	WRB2	WRB3	WRB4	WRB5	WRB6	Total
DWRB1	5/87	14/98	17/95	24/112	40/170	28/126	128/688
DWRB2	1/52	3/42	21/130	2/36	17/148	1/20	45/428
DWRB3	0/45	6/112	18/160	13/111	7/74	5/92	49/594
DWRB4	0/14	6/142	5/54	18/172	0/9	23/150	52/541
DWRB5	0/8	0/25	10/23	89/212	21/55	32/81	152/404
DWRB6	1/68	8/150	4/39	2/8	3/19	0/18	18 / 302
Total	7	37	75	148	88	89	444/2957 = 15%
Cinzana	WRC1	WRC2	WRC3	WRC4	WRC5	WRC6	Total
DWRC1	1/41	1/7	6/51	42/108	33/133	2/19	85/359
DWRC2	0/12	1/4	3/43	2/5	7/126	0/4	13/194
DWRC3	0/27	5/18	11/67	24/91	4/29	14/62	58/294
DWRC4	4/60	7/61	10/72	37/183	7/34	43/137	108/547
DWRC5	3/62	0/5	15/90	4/24	6/66	10/94	38/341
DWRC6	3/125	3/88	15/120	4/47	5/84	11/89	41/533
Total	11/327	17/183	60/443	113/458	62/472	80/405	343/2288 = 15%
Mbour	WRM1	WRM2	WRM3	WRM4	WRM5	WRM6	Total
DWRM1	0/57	25/202	35/197	2/48	3/50	10/80	75/631
DWRM2	3/18	1/33	8/37	4/39	30/156	8/60	54/343
DWRM3	1/20	6/177	5/23	12/60	9/68	15/154	48/502
DWRM4	2/15	35/142	36/158	18/75	1/24	17/82	109/496
DWRM5	1/19	16/78	2/15	17/64	10/64	26/97	72/337
DWRM6	0/4	0/18	1/9	4/27	10/104	9/71	24/233
Total	7/133	83/650	87/439	57/313	63/466	85/544	378/2545 = 15%

Tableau 5.23 –	Pou	rcentage de	e doi	nnées of	$\operatorname{des}$	classe	es de la	base de	e do	nnées (	désai	sor	nalis	ées
correspondant	aux	maximum	s de	$PM_{10}$	oà	15%	conten	u dans	les	classes	s de	la	base	de
données brutes														

Baniz	WRB1	WRB2	WRB3	WRB4	WRB5	WRB6	Total
DWRB1	5,7	14,3	17,9	21,4	23,5	22,2	18,6
DWRB2	1	$^{7,1}$	$16,\! 0$	$^{5,6}$	$11,\!5$	5	10,5
DWRB3	$1,\!9$	$^{5,4}$	$11,\!3$	11,7	$^{9,5}$	$^{5,4}$	8,3
DWRB4	0	$^{4,2}$	$_{9,3}$	10,5	0	$15,\!3$	9,6
DWRB5	0	0	$43,\!5$	42,0	38,2	$39,\!5$	12,9
DWRB6	$^{1,5}$	$^{5,3}$	10,3	25	$15,\!8$	0	6,0
Cinzana	WRC1	WRC2	WRC3	WRC4	WRC5	WRC6	Total
DWRC1	2,4	14,3	11,8	38,9	24,8	10,5	23,7
DWRC2	0	25	$^{6,9}$	40	$^{5,6}$	0	6,7
DWRC3	0	$27,\!8$	16,4	26,4	$13,\!8$	$22,\!6$	19,7
DWRC4	$^{6,7}$	$11,\!5$	$13,\!9$	20,2	$20,\!6$	$31,\!4$	19,7
DWRC5	4,8	0	16,7	16,7	$^{9,1}$	$10,\! 6$	11,1
DWRC6	$^{2,4}$	$^{3,4}$	12,5	$^{8,5}$	$^{5,9}$	$12,\!4$	7,7
Mbour	WRM1	WRM2	WRM3	WRM4	WRM5	WRM6	Total
DWRM1	0,0	12,4	17,8	4,2	6,0	12,5	11,8
DWRM2	16,7	$^{3,0}$	$21,\!6$	10,3	19,2	$13,\!3$	15,7
DWRM3	$^{5,0}$	$^{3,4}$	21,7	20,0	$13,\!2$	$^{9,7}$	9,6
DWRM4	$13,\!3$	$24,\! 6$	$22,\!8$	24,0	$^{4,2}$	20,7	22,0
DWRM5	$^{5,3}$	20,5	$13,\!3$	$26,\!6$	$15,\! 6$	$26,\!8$	21,4
DWRM6	$^{0,0}$	$0,\!0$	$11,\!1$	$14,\!8$	$^{9,6}$	12,7	10,3

Ces résultats mettent l'accent sur le potentiel d'une utilisation croisée des régimes de temps bruts et interannuels pour caractériser certaines combinaisons propices à des épisodes de forts contenus en poussière. Concernant les régimes de temps bruts, nous avons constaté sur l'ensemble des trois stations que la probabilité de dépasser le seuil de contenu élevé en poussière (les 15% les plus élevés) est faible pour les régimes de temps caractéristiques du flux de mousson, et ce quelque soit les régimes de temps interannuels. Par contre, pour les régimes de temps caractéristiques du flux d'harmattan, alors que la probabilité de passer ce seuil est en moyenne importante, elle peut être très faible en occurrence de certains régimes de temps interannuels.

## 5.3.4 Conclusion

Le but de cette partie a été de définir les types de temps qui pourraient avoir un impact sur la variabilité intra-annuelle des aérosols pendant la saison sèche sur trois stations de l'Afrique de l'Ouest. Ainsi nous avons déterminé les régimes de temps en saison sèche sur les trois stations en éliminant cette fois ci l'influence de la composante saisonnière sur les données. Comme précédemment, après avoir appliqué la méthode sur les données, six régimes de temps ont été identifiés. Pour avoir une dynamique régionale on a renommé les régimes de temps par DWR1...DWR6 de sorte à avoir pour les régimes de temps qui se ressemblent de Banizoumbou à Mbour le même numéro. Nous avons remarqué dans cette partie comme pour la partie précédente sur les données brutes, que sur l'ensemble des trois stations le cycle diurne n'influence pas les régimes de temps. Cependant ces régimes de temps sont caractérisés par une forte variabilité interannuelle. Nous avons remarqué aussi que les régimes de temps sont moins persistants que les régimes de temps bruts. Les champs de géopotentiel à 500 hPa pour les différentes régimes de temps des trois stations sont évidemment très différents de ceux des données brutes qui caractérisaient le signal saisonnier, et représentent des anomalies de circulation, soit cyclonique, soit anticyclonique, d'échelle synoptique et centrée sur le nord du domaine. La mise en relation entre des régimes de temps et la NAO permet de voir qu'il y a des régimes de temps qui lui sont plus liés et d'autres moins.

L'étude de la relation AOT- $PM_{10}$  obtenue dans cette échelle interannuelle a permis de voir que les types de temps peuvent être différenciés selon les pentes de la régression. Ainsi selon les pentes, une même valeur d'AOT ne donne pas la même concentration de surface. Cependant en terme de corrélation, il n'existe pas un lien particulier entre les régimes de temps et la relation entre les deux variables. Ce qui a été constaté est qu'il y a des régimes de temps qui améliorent la relation en terme de corrélation et d'autres la détériorent, mais globalement les corrélations sont améliorées. En intégrant le processus des régimes de temps avec les informations météorologiques locales spécifiques à chaque station, puis en regroupant les résultats sur l'ensemble des régimes de temps, on obtient d'excellents résultats dans la prévision des valeurs de  $PM_{10}$ .

Les résultats présentés sur le système d'alerte opérationnel ont permis de montrer le potentiel d'une utilisation croisée des régimes de temps bruts (saisonnier) et interannuels (désaisonnalisés) pour caractériser certaines combinaisons propices à des épisodes de forts contenus en poussière. Il faudra pour concrétiser cela évaluer le niveau de prévisibilité de ces combinaisons de types de temps brut-désaisonnalisé, avec par exemple les prévisions à 15 jours fournies opérationnellement par ECMWF.

# 5.4 Conclusion

Dans ce chapitre nous avons prolongé le travail développé dans Yahi et al. [2013] sur l'analyse des relations entre AOT et  $PM_{10}$  sur les stations de Banizoumbou et Cinzana durant la saison sèche (d'Octobre à Mai). Un nombre plus important de variables ont été inclus de façon à mieux caractériser les régimes de temps par leur composante dynamique et thermodynamique. En conséquence une nouvelle méthode de sélection de variables a été développée, par bloc, en posant dans un bloc les variables dynamiques et dans un autre bloc les variables thermodynamiques. On a sélectionné 6 régimes de temps représentant comme dans Yahi et al., des structures de grande échelle caractéristiques de différentes phases du cycle saisonnier durant la saison sèche. Par cette approche on a obtenu une discrimination un peu meilleure, caractérisée en particulier par une détection de l'AEJdans les profils verticaux de vents pour les régimes de transition entre saison sèche et saison de mousson en plus du NLLJ identifié sur les régimes de temps caractéristiques de la saison sèche.

Yahi et al. utilisant l'information sur les régimes de temps ont montré ensuite qu'il existe un lien entre la direction du vent et la concentration de poussières ou l'épaisseur optique en aérosols. L'analyse en type de temps a montré que les fortes concentrations de poussières sont enregistrées pour certaines configurations météorologiques qui concernent tout le Sahel. Ainsi les types de temps caractéristiques des vents d'harmattan sont responsables de la forte présence des concentrations en  $PM_{10}$  et l'AOT avec une corrélation significative entre les deux paramètres alors que les types de temps caractéristiques des vents de mousson présentent le contraire. La pente de régressions obtenues sur la relation entre les concentrations en  $PM_{10}$  et l'AOT diffèrent suivant les régimes de temps. Dans une seconde étape, pour chaque régime de temps de grande échelle et à chaque station, l'information météorologique locale de la station permet d'améliorer la relation AOT- $PM_{10}$ . A ce niveau nous avons aussi considéré dans les variables le coefficient d'Angstrom souvent utilisé pour caractériser les différents types d'aérosols, en particulier les poussières désertiques des aérosols de pollution. Ce qui a été identifié en plus des travaux de Yahi et al. [2013], est que les régimes de temps caractéristiques des mois de transition entre la saison sèche et la saison des plus sont dominés par les poussières désertiques alors que
certains régimes de temps caractéristiques des mois de la saison sèche sont dominés pas des cas d'aérosols de pollution .

Dans cette approche générale, la station de M'Bour, située le long de la petite côte, a été ajoutée à cette analyse sur les stations continentales de Cinzana et Banizoumbou développée par Yahi et al.. Ceci a permis de couvrir plus largement le transect sahélien et d'observer les contrastes possibles en longitude. Cependant bien que des ressemblances peuvent être identifiées entre les trois stations, des différences peuvent aussi être notées entre la station côtière de Mbour et les stations continentales (Banizoumbou et Cinzana). Par exemple, on a trouvé que la relation AOT- $PM_{10}$  est meilleure sur les régimes de temps de la saison sèche que ceux des mois de transition sur les stations de Banizoumbou et Cinzana, Par contre sur la station côtière de Mbour, il existe un régime de temps des mois de transition entre la saison sèche et la saison des pluies dont la relation AOT- $PM_{10}$ est meilleure que certains régimes de temps de la saison sèche, ce qui semble être lié à sa position géographique, par le fait que les particules marines affectent les concentrations de poussières.

Nous avons ensuite développé la même approche que Yahi et al. mais sur une autre échelles de temps. En effet Yahi et al. avaient noté que l'utilisation des données brutes (sans filtrage préalable) ne permet pas d'appréhender les origines d'évènements de soulè-vement/transport de poussières importants mais sporadiques et d'échelle synoptique, qui sont observables sur le Sahel. On a donc appliqué cette approche aux données désaisonnalisées en éliminant le cycle saisonnier moyen calculé jour à jour, de manière à supprimer l'effet de la saisonnalité qui domine largement dans les résultats de Yahi et al., et faire apparaître la variabilité résiduelle, particulièrement d'échelle interannuelle. Ainsi nous avons constaté que les régimes de temps obtenus présentent des anomalies de circulation, soit cyclonique, soit anticyclonique, d'échelle synoptique et centrée sur le nord de notre domaine d'étude. En comparant les résultats obtenus ici aux résultats des régimes de temps bruts, on remarque que ce n'est pas la structuration du cycle saisonnier qui apparaît, mais des régimes de temps caractéristiques de modulation d'ondes synoptiques qui peuvent être représentatifs des anomalies de circulation de certains mois de certaines années comme on l'a montré par les résultats de l'AFC.

La mise en relation entre les régimes de temps et la relation AOT- $PM_{10}$  a permis de voir que pour chaque station, il y a des régimes de temps qui améliorent la relation er d'autre qui la détériorent. Cependant globalement les corrélations sont améliorées. Comme pour les régimes de temps bruts, nous avons pu voir que les régimes de temps peuvent être différenciés par la pente de la régression AOT- $PM_{10}$ . En intégrant tout le processus des régimes de temps avec les informations météorologiques locales spécifiques à chaque station, puis en regroupant les résultats sur l'ensemble des régimes de temps, on arrive à bien prévoir des valeurs de  $PM_{10}$ .

Dans cette partie, nous avons aussi exploré la potentialité de mettre en place un système d'alerte statistique opérationnel aux évènements forts de poussière en combinant les résultats des classifications sur les données brutes et sur les données filtrées du cycle saisonnier moyen. Les résultats présentés sur le système d'alerte opérationnel ont permis de montrer le potentiel d'une utilisation croisée des régimes de temps bruts (saisonnier) et interannuels (désaisonnalisés) pour caractériser certaines combinaisons propices à des épisodes de forts contenus en poussière. Il faudra pour concrétiser cela évaluer le niveau de prévisibilité de ces combinaisons de types de temps brut-désaisonnalisé, avec par exemple les prévisions à 15 jours fournies opérationnellement par ECMWF.

### Chapitre 6

# Conclusion générale et perspectives

L'objectif principal de cette thèse était d'étudier la variabilité des aérosols désertiques en Afrique de l'Ouest et de comprendre les mécanismes météorologiques qui la contrôlent. Ainsi deux grandes parties ont été retenues.

La première étape a consisté à étudier la variabilité des concentrations des aérosols en analysant le cycle saisonnier et le cycle diurne des concentrations en  $PM_{10}$  sur la période de 2006 à 2010, des données disponibles pour les trois stations sahéliennes du « Sahelian Dust Transect » (SDT) afin de comprendre les phénomènes météorologiques locaux qui contrôlent cette variabilité. L'analyse des concentrations de poussières mesurées entre 2006 et 2010 a permis de confirmer un cycle saisonnier régional présenté par Marticorena et al. [2010], caractérisé par un maximum de concentration en saison sèche et un minimum pendant la saison des pluies sur l'ensemble de trois stations. La mise en relation entre les cycles saisonniers de la concentration en  $PM_{10}$  et la vitesse du vent montre un décalage dans le temps des fortes valeurs des deux paramètres sur les stations de Cinzana et Banizoumbou. Les  $PM_{10}$  les plus forts sont enregistrés en Mars alors que les vents les plus forts sont enregistrés en début de saison des pluies, au mois de Juin. Ce décalage montre que le vent local n'est pas responsable des maximums de  $PM_{10}$  de chaque station, ce qui conforte l'idée selon laquelle ce sont les transports d'aérosols sahariens qui sont responsables des maxima de concentration observés en saison sèche au Sahel (Marticorena et al., 2010). Par ailleurs le gradient d'Est en Ouest, c'est à dire de Banizoumbou à Mbour, des concentrations médianes mensuelles apparaît être lié à la position géographique de chaque station. Banizoumbou où les plus fortes concentrations ont été enregistrées est beaucoup plus proche des zones reconnues comme étant parmi les plus grandes productrices en aérosols désertiques en Afrique telles que le sud de Algérie, en Libye et au Tchad (dépression de Bodélé) [Prospero et al., 2002]. Les données recueillies sur les cinq années ont permis de quantifier la variabilité des concentrations moyennes mensuelles. Il existe une très forte variabilité des concentrations mensuelles sur les cinq années de mesure qui peut être en partie liée à la variabilité interannuelle. Les concentrations médianes mensuelles très élevées ou faibles sont enregistrées à une période spécifique de l'année. C'est le cas du mois de Janvier 2007, où la concentration médiane est significativement plus élevée que les autres années pour l'ensemble des trois stations. L'année 2009 apparaît comme l'année la moins poussiéreuse sur la période des cinq années dans les trois stations. Elle est également parmi les années les plus sèches à Cinzana et M'Bour sur la période des cinq années. Les plus faibles concentrations sont enregistrées pendant la saison des pluies coïncidant avec la période de maximum de précipitations obtenu en Aout à Cinzana et Banizoumbou et en Septembre à Mbour. Les concentrations médianes mensuelles en saison des pluies sont trois fois plus élevées à M'Bour (~  $30 \ \mu g/m^3$ ) qu'à Banizoumbou et Cinzana (  $10 \ \mu g.m^{-3}$ ) malgré le fait que les précipitations annuelles sont dans le même ordre de grandeur qu'à Banizoumbou.

Les résultats présentés ici nous ont permis aussi de voir qu'il existe un cycle diurne clair et distinct entre la saison des pluies et la saison sèche sur les trois stations. Pendant la saison sèche, le cycle diurne de la concentration en  $PM_{10}$  est maximal le matin entre 7h et 14h avec un pic à 10h et minimal pendant la nuit, coïncidant avec le cycle diurne de la vitesse du vent. Au Sahel, le cycle diurne des vents de surface est fortement lié à la dynamique du jet nocturne de basses couches. Le pic de vent du matin (10 h) coïncidant avec celui de la concentration en  $PM_{10}$  nous a permis de penser que le jet nocturne saharien module aussi le cycle diurne des concentrations en particules mesurées au Sahel. Pendant la saison des pluies, le cycle diurne de la concentration en  $PM_{10}$  apparaît diffèrent de celui de la saison sèche. La concentration en  $PM_{10}$  maximale en fin d'après midi entre 16h et 23h sur les stations de Banizoumbou et Cinzana coïncidant avec un renforcement des directions de vent du flux de mousson (180-270°). Pour la station de Mbour, le cycle diurne est différent des autres stations avec un pic aussi bien en fin d'après midi (19h) qu'en début de matinée (8h). Cependant le pic de fin d'après midi (19h) coïncide avec un renforcement des vents de secteur ouest nord-ouest. Par ailleurs le cycle diurne de la concentration en  $PM_{10}$  est en déphasage avec celui de la vitesse de vent ayant un pic le matin aux environs de 10 h à Banizoumbou et Cinzana et un pic entre (15-17 h) pour la station de Mbour. Ainsi nous pensons que la variabilité des concentrations de poussière en saison des pluies apparaît être modulée par l'évolution des systèmes convectifs a l'échelle régionale. Cependant, les émissions locales par les systèmes convectifs sont des événements de courte durée qui ne sont pas entièrement recherchés dans l'ensemble de données horaires utilisé dans cette étude.

La deuxième étape a consisté à réaliser une analyse climatologique en types de temps en saison sèche (Octobre-Mai) sur les trois stations afin de construire à l'échelle de l'Afrique de l'Ouest des relations entre la mesure de la concentration des particules au sol,  $PM_{10}$  et la mesure de l'épaisseur optique des aérosols en identifiant les situations météorologiques à grande échelle qui influencent ces relations. Ceci a permis de mieux prévoir les concentrations massiques en  $PM_{10}$  à partir des mesures d'épaisseurs optiques pour chaque régime de temps. Cette approche s'inspire du travail réalisé par Yahi et al. [2013] sur les stations de Banizoumbou et Cinzana que nous avons étendu selon plusieurs axes. Tout d'abord pour mener à bien cette étude, nous avons voulu tout d'abord développer une méthode capable de prendre en compte les difficultés liées à l'origine multi-sources et multi-blocs des données. En effet nous avons voulu compléter le jeu de variables, définies dans Yahi et al. sur un carré de 4° de latitude/longitude autour de chacune des stations, de façon à mieux caractériser les régimes de temps, en combinant des variables décrivant la circulation de grande échelle à des variables décrivant la composante thermodynamique de l'atmosphère et son instabilité verticale. Nous avons aussi considéré dans les variables le coefficient d'Angstrom souvent utilisé pour caractériser les différents types d'aérosols, en particulier les poussières désertiques des aérosols de pollution. Enfin la station de M'Bour, située le long de la petite côte, a été ajoutée à cette analyse sur les stations continentales de Cinzana et Banizoumbou, afin de couvrir plus largement le transect sahélien et d'observer les contrastes possibles en longitude. Afin de mieux prendre en compte les difficultés liées à l'étude, nous nous sommes intéressé aux méthodes de classifications, précisément les méthodes de classifications multi bloc. Ce qui nous a permis de mettre en place une approche de sélection de variables en classification. A partir de cela, nous avons développé une approche statistique permettant de définir les types de temps sur deux échelles de temps différentes. En effet Yahi et al. ont noté que l'utilisation des données brutes (sans filtrage préalable) ne permet pas d'appréhender les origines de certains évènements de soulèvement/transport de poussières importants mais sporadiques et d'échelle synoptique, qui sont observables sur le Sahel. On a donc appliqué leur approche, (i) aux données brutes, (ii) aux données désaisonnalisées en éliminant le cycle saisonnier moyen calculé jour à jour, de manière à supprimer l'effet de saisonnalité qui domine largement dans leurs résultats et faire apparaître la variabilité d'échelle interannuelle. Dans un premier temps, on a défini 6 régimes de temps « bruts » qui décrivent la saisonnalité climatique sur les trois stations en utilisant les champs météorologiques sur neuf niveaux (1000 hPa, 925 hPa, 850 hPa, 700 hPa, 500 hPa, 400 hPa, 300 hPa, 250 hPa, 200 hPa) que sont la température (T), le vent zonal (U), le vent méridien (V), le géopotentiel (Z), l'humidité spécifique (Q), la vitesse verticale (W), ainsi que la variable hauteur de la couche limite (BLH). Nous avons pu voir que les régimes de temps ne sont pas sensibles au cycle diurne. Cependant ils sont sensibles au cycle saisonnier. Les régimes de temps obtenus ont été classés en deux grandes familles météorologiques. Dans les basses couches, nous nous avons identifié une dynamique régionale des régimes de temps, avec ceux caractéristiques du flux de mousson en début et fin de saison de mousson sur l'ensemble des trois stations (WR1 et WR2) et les quatre autres types de temps caractéristiques du flux d'harmattan (WR3, WR4, WR5 et WR6). Les régimes de temps caractéristiques du flux d'harmattan sont associés à un pic de vent zonal et de vent méridien à 925 hPa signalant la présence du jet nocturne de basses couches alors que les régimes de temps 1 et 2 caractéristiques du flux de mousson sont associées à un pic de vent zonal observé à 700 hPa signalant ainsi la présence du JEA (qui n'était pas clairement discriminé dans l'analyse en 5 classes de Yahi et al.). La quantification du degré de confusion des types de temps des trois stations a permis de voir les régimes de temps des stations continentales (Banizoumbou et Cinzana) se ressemblent le plus, c'est à dire qu'il y a donc plus de concordance entre les régimes de temps de Banizoumbou /Cinzana que les régimes de temps des stations Cinzana/Mbour, ce qui semble lié a la position côtière de la station de Mbour. L'utilisation de l'information sur les régimes de temps a permis de construire de meilleures régressions linéaires AOT- $PM_{10}$  pour une majorité de ces régimes. La relation AOT- $PM_{10}$  examinée sur chaque régime de temps nous a permis de voir que les régimes de temps caractéristiques du flux d'harmattan présentent une relation meilleure en terme de corrélation par rapport a la relation avant la séparation en type de temps avec de forte valeurs d'AOT et de  $PM_{10}$  pour les stations continentales de Cinzana et Banizoumbou. Ce qui nous a permis de dire que la météorologie joue un rôle important dans la relation AOT- $PM_{10}$ . Nous avons montré aussi que la pente de ces régressions entre AOT et  $PM_{10}$  diffère suivant les régimes de temps, pouvant donner des correspondances de  $PM_{10}$  en fonction d'AOT du simple au double. Dans une seconde étape, pour chaque régime de temps de grande échelle et à chaque station, l'information météorologique locale de la station a permis d'améliorer la relation  $AOT-PM_{10}$ . Nous avons ainsi obtenu une meilleure estimation globale des  $PM_{10}$  après les régimes de temps comparée à l'estimation sans prise en compte de ces régimes de temps. A ce niveau nous avons aussi considéré dans les variables le coefficient d'Angstrom souvent utilisé pour caractériser les différents types d'aérosols, en particulier les poussières désertiques des aérosols de pollution. Nous avons pu voir que la relation AOT- $PM_{10}$  est meilleure sur les régimes de temps des stations continentales que ceux de la station côtière. Nous avons ensuite développé une approche similaire mais en filtrant les données par élimination du cycle saisonnier moyen jour à jour. Après avoir appliqué la méthode sur les données, six régimes de temps ont aussi été identifiés. Ces régimes de temps sont caractérisés par une forte variabilité interannuelle, et comme pour la partie précédente sur les données brutes, sur l'ensemble des trois stations le cycle diurne n'influence pas l'occurrence de ces régimes de temps. Nous avons remarqué aussi que les régimes de temps sont moins persistants que les régimes de temps bruts. Les champs atmosphériques (géopotentiel à 500 hPa par exemple) pour les différents régimes de temps des trois stations présentent principalement des anomalies de circulation d'échelle synoptique, soit cyclonique, soit anticyclonique, centrée sur le nord du domaine. Ils présentent aussi une forte cohérence inter-stations. Ces régimes de temps, définis sur un carré de 4° de latitude/longitude autour de chacune des stations pour cibler les relations AOT- $PM_{10}$ , possèdent ainsi des structures spatiales de circulation atmosphérique très cohérentes sur le domaine ouest-africain dans sa totalité. Enfin, la mise en relation entre des régimes de temps et la NAO permet de voir qu'il y a des régimes de temps qui lui sont plus liés et d'autres moins. L'étude de la relation AOT- $PM_{10}$  obtenue dans cette échelle interannuelle a montré qu'il y a des régimes de temps qui améliorent la relation et d'autres qui la détériorent, même si globalement le résultat intégré est meilleur avec la prise en compte de ces régimes. Dans cette analyse à l'échelle interannuelle, nous avons aussi exploré la potentialité de mettre en place un système d'alerte statistique opérationnel aux évènements forts de poussière en combinant les résultats des classifications sur les données brutes et sur les données filtrées du cycle saisonnier moyen. Nous avons montré qu'il existe une réelle perspective de pouvoir utiliser l'information apportée par les régimes de temps car certaines configurations atmosphériques montrent une probabilité élevée d'occurrence de contenu en  $PM_{10}$  et d'autres configurations une probabilité quasi-nulle.

Les perspectives inspirées par cette étude sont nombreuses :

- L'un des problèmes rencontrés dans cette étude est celui des différentes échelles sur lesquelles sont mesurées les différentes variables de l'étude. Par exemple les variables météorologiques que nous avons utilisées pour déterminer les régimes de temps sont à une résolution de 55km, alors que les concentrations en particules sont mesurées sur un point. L'utilisation d'une météorologie plus locale (c'est à dire à une échelle plus petite) pour la détermination des régimes de temps devrait permettre d'affiner plus les résultats.
- L'application de la méthodologie à de plus grandes séries chronologiques de données doit permettre d'affiner les fonctions de prédiction de la relation AOT- $PM_{10}$
- Il serait intéressant d'établir une méthode statistique permettant d'identifier les groupes de variables pouvant former les blocs de variables à faire entrer dans l'algorithme. En effet les deux blocs de variables constitués dans cette étude ont été formés sur l'avis d'expert.
- L'apport d'information sur les profils verticaux d'aérosols par mesures lidar disponibles à la station de M'Bour pourront aider à affiner les relations  $AOT-PM_{10}$  sur cette station, en particulier pour les régimes de temps où cette relation est déficiente.
- On pourra aussi voir comment généraliser l'extension à l'échelle de l'Afrique de l'Ouest des relations  $AOT-PM_{10}$  déterminées localement, afin d'obtenir un suivi sur une couverture spatiale sur l'évolution des contenus en  $PM_{10}$ . Puis il serait intéressant d'appliquer la méthode développée ici à d'autres zones géographiques afin de mesurer la stabilité de la méthode.
- Il existe une réelle perspective de pouvoir développer un système d'alerte opérationnel de certains forts épisodes de poussières, d'une part en combinant les résultats des classifications sur les données brutes et sur les données filtrées du cycle saisonnier moyen, puis en intégrant les prévisions météorologiques opérationnelles de ECMWFde l'information « locale » autour des stations aux échéances de 1 à 15 jours (on a vu que les régimes de temps sont relativement persistants, donc potentiellement prévisibles à ces échéances) ; ou inversement à partir de la projection des prévisions ECMWF sur les structures spatiales des régimes de temps à l'échelle de l'Afrique de l'ouest.
- Pour obtenir des scores de prévision plus discriminants, il sera vraisemblablement nécessaire d'étudier plus précisément la dynamique temporelle de ces épisodes de poussières et affiner la définition des régimes de temps en fonction de cette dynamique, sur des échelles de temps plus fines. Par exemple une étude complémentaire pourrait être réalisée à une échelle temporelle plus fine en faisant un lissage des données par moyenne mobile afin de déterminer les régimes de temps qui pourraient avoir un impact sur la variabilité intra-saisonnière des aérosols en saison sèche (cette

approche a été démarrée mais n'a finalement pas été incluse dans ce document de thèse).

- Il serait intéressant aussi dans la suite de travailler sur la persistance des combinaisons « régimes de temps bruts-désaisonnalisés » à forts contenus de  $PM_{10}$  en lien avec les cumuls  $PM_{10}$  sur plusieurs jours.
- Ce type de système de prévision du contenu en aérosols désertiques en lien avec les régimes de temps pourrait être intégré dans un système de prévision de risque de méningite en s'appuyant sur l'ensemble des études liant épidémie de méningite et conditions environnementales.
- Ces résultats de thèse seraient utiles pour aborder le volet de modélisation, en particulier l'évaluation de modèle déterministe du cycle d'aérosols désertiques de type « chimie-transport » comme CHIMERE-Dust à reproduire et prévoir correctement la variabilité observée des teneurs en aérosols désertiques, (concentrations de surface, répartition verticale, phases de maximum, ...), et ses performances en fonction des différents régimes de temps.

## Bibliographie

- Agrawal, R., J. Gehrke, D. Gunopulos, and P. Raghavan, Automatic subspace clustering of high dimensional data for data mining applications, vol. 27, ACM, 1998.
- Allab, K., K. Benabdeslem, and A. Aussem, Une approche de co-classification automatique à base des cartes topologiques, *Revue des Nouvelles Technologies de l'information* (*RNTI*), pp. 1–24, 2011.
- Andreae, M. O., Climatic effects of changing atmospheric aerosol levels, World survey of climatology, 16, 347–398, 1995.
- Bach, W., Global air pollution and climatic change, *Reviews of Geophysics*, 14(3), 429–474, 1976.
- Balme, M., and R. Greeley, Dust devils on earth and mars, *Reviews of Geophysics*, 44(3), 2006.
- Banta, R. M., Y. L. Pichugina, and R. K. Newsom, Relationship between low-level jet properties and turbulence kinetic energy in the nocturnal stable boundary layer, *Journal of the atmospheric sciences*, 60(20), 2549–2555, 2003.
- Banta, R. M., Y. L. Pichugina, and W. A. Brewer, Turbulent velocity-variance profiles in the stable boundary layer generated by a nocturnal low-level jet, *Journal of the atmospheric sciences*, 63(11), 2700–2719, 2006.
- Bergametti, G., L. Gomes, G. Coudé-Gaussen, P. Rognon, and M.-N. Le Coustumer, African dust observed over canary islands : Source-regions identification and transport pattern for some summer situations, *Journal of Geophysical Research : Atmospheres* (1984–2012), 94 (D12), 14,855–14,864, 1989.
- Blackadar, A. K., Boundary layer wind maxima and their significance for the growth of nocturnal inversions, Bull. Amer. Meteor. Soc, 38(5), 283–290, 1957.
- Bou Karam, D., C. Flamant, P. Knippertz, O. Reitebuch, J. Pelon, M. Chong, and A. Dabas, Dust emissions over the sahel associated with the west african monsoon intertropical discontinuity region : A representative case-study, *Quarterly Journal of the Royal Meteorological Society*, 134(632), 621–634, 2008.
- Bou Karam, D., C. Flamant, P. Tulet, M. C. Todd, J. Pelon, and E. Williams, Dry cyclogenesis and dust mobilization in the intertropical discontinuity of the west african monsoon : A case study, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 114 (D5), 2009.
- Brooks, N., and M. Legrand, Dust variability over northern africa and rainfall in the sahel, in *Linking climate change to land surface change*, pp. 1–25, Springer, 2000.
- Burpee, R. W., The origin and structure of easterly waves in the lower troposphere of north africa, *Journal of the Atmospheric Sciences*, 29(1), 77–90, 1972.

- Burpee, R. W., Characteristics of north african easterly waves during the summers of 1968 and 1969, *Journal of the Atmospheric Sciences*, 31(6), 1556–1570, 1974.
- Charba, J., Application of gravity current model to analysis of squall-line gust front, Monthly Weather Review, 102(2), 140–156, 1974.
- Charney, J., and M. Stern, On the stability of internal baroclinic jets in a rotating atmosphere, *Journal of the Atmospheric Sciences*, 19(2), 159–172, 1962.
- Charrad, M., and M. B. Ahmed, Simultaneous clustering : A survey, in *Pattern Recognition* and *Machine Intelligence*, pp. 370–375, Springer, 2011.
- Chavent, M., C. Lacomblez, and B. Patouille, Critère de rand asymétrique, *Proceedings* SFC, 8(82-88), 69, 2001.
- Chen, X., Y. Ye, X. Xu, and J. Z. Huang, A feature group weighting method for subspace clustering of high-dimensional data, *Pattern Recognition*, 45(1), 434–446, 2012.
- Chen, Y., S. M. Kreidenweis, L. M. McInnes, D. C. Rogers, and P. J. DeMott, Single particle analyses of ice nucleating aerosols in the upper troposphere and lower stratosphere, *Geophysical research letters*, 25(9), 1391–1394, 1998.
- Chiapello, I., and C. Moulin, Toms and meteosat satellite records of the variability of saharan dust transport over the atlantic during the last two decades (1979–1997), *Geophysical Research Letters*, 29(8), 17–1, 2002.
- Chiapello, I., J. Prospero, J. Herman, and N. Hsu, Detection of mineral dust over the north atlantic ocean and africa with the nimbus 7 toms, *Journal of Geophysical Research :* Atmospheres (1984–2012), 104 (D8), 9277–9291, 1999.
- Chiapello, I., C. Moulin, and J. M. Prospero, Understanding the long-term variability of african dust transport across the atlantic as recorded in both barbados surface concentrations and large-scale total ozone mapping spectrometer (toms) optical thickness, *Journal of Geophysical Research : Atmospheres (1984–2012), 110*(D18), 2005.
- d'Almeida, G. A., On the variability of desert aerosol radiative characteristics, Journal of Geophysical Research : Atmospheres (1984–2012), 92(D3), 3017–3026, 1987.
- Davies, D. L., and D. W. Bouldin, A cluster separation measure, Pattern Analysis and Machine Intelligence, IEEE Transactions on, (2), 224–227, 1979.
- Deroubaix, A., N. Martiny, I. Chiapello, and B. Marticoréna, Suitability of omi aerosol index to reflect mineral dust surface conditions : Preliminary application for studying the link with meningitis epidemics in the sahel, *Remote Sensing of Environment*, 133, 116–127, 2013.
- Droegemeier, K. K., and R. B. Wilhelmson, Numerical simulation of thunderstorm outflow dynamics. part i : Outflow sensitivity experiments and turbulence dynamics, *Journal of* the Atmospheric Sciences, 44(8), 1180–1210, 1987.
- Dubovik, O., B. Holben, T. F. Eck, A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanré, and I. Slutsker, Variability of absorption and optical properties of key aerosol types observed in worldwide locations, *Journal of the atmospheric sciences*, 59(3), 590–608, 2002.
- Duce, R. A., and N. W. Tindale, Atmospheric transport of iron and its deposition in the ocean, *Limnology and Oceanography*, 36(8), 1715–1726, 1991.

- Engelstaedter, S., and R. Washington, Atmospheric controls on the annual cycle of north african dust, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 112(D3), 2007.
- Engelstaedter, S., I. Tegen, and R. Washington, North african dust emissions and transport. earth sci. rev. 79, 73–100, 2006.
- Evan, A. T., A. K. Heidinger, and P. Knippertz, Analysis of winter dust activity off the coast of west africa using a new 24-year over-water advanced very high resolution radiometer satellite dust climatology, *Journal of Geophysical Research : Atmospheres* (1984–2012), 111(D12), 2006 b.
- Fiedler, S., K. Schepanski, B. Heinold, P. Knippertz, and I. Tegen, Climatology of nocturnal low-level jets over north africa and implications for modeling mineral dust emission, *Journal of Geophysical Research : Atmospheres*, 118(12), 6100–6121, 2013.
- Flamant, C., J.-P. Chaboureau, D. Parker, C. Taylor, J.-P. Cammas, O. Bock, F. Timouk, and J. Pelon, Airborne observations of the impact of a convective system on the planetary boundary layer thermodynamics and aerosol distribution in the inter-tropical discontinuity region of the west african monsoon, *Quarterly Journal of the Royal Me*teorological Society, 133(626), 1175–1189, 2007.
- Flamant, C., P. Knippertz, D. Parker, J.-P. Chaboureau, C. Lavaysse, A. Agusti-Panareda, and L. Kergoat, The impact of a mesoscale convective system cold pool on the northward propagation of the intertropical discontinuity over west africa, *Quarterly Journal of the Royal Meteorological Society*, 135(638), 139–159, 2009.
- Garratt, J., et al., The atmospheric boundary layer, cambridge atmospheric and space science series, *Cambridge University Press, Cambridge*, 416, 444, 1992.
- Ginoux, P., and O. Torres, Empirical toms index for dust aerosol : Applications to model validation and source characterization, *Journal of Geophysical Research : Atmospheres* (1984–2012), 108(D17), 2003.
- Gordon, A., A survey of constrained classification, Computational Statistics & Data Analysis, 21(1), 17–29, 1996.
- Govaert, G., Classification croisée, These d'état, Université Paris, 6, 1983.
- Govaert, G., Classification simultanée de tableaux binaires, *Data analysis and informatics*, 3, 233–236, 1984.
- Griffin, D., C. Kellog, V. Garrison, and E. Shinn, La mondialisation des poussières, Pour la science, 309, 80–5, 2003.
- Gueye, A., S. Janicot, A. Niang, S. Sawadogo, B. Sultan, A. Diongue-Niang, and S. Thiria, Weather regimes over senegal during the summer monsoon season using self-organizing maps and hierarchical ascendant classification. part i : synoptic time scale, *Climate* dynamics, 36(1-2), 1–18, 2010.
- Gueye, A., S. Janicot, A. Niang, S. Sawadogo, B. Sultan, A. Diongue-Niang, and S. Thiria, Weather regimes over senegal during the summer monsoon season using self-organizing maps and hierarchical ascendant classification. part ii : interannual time scale, *Climate dynamics*, 36(1-2), 1–18, 2012.
- Guichard, F., L. Kergoat, E. Mougin, F. Timouk, F. Baup, P. Hiernaux, and F. Lavenu, Surface thermodynamics and radiative budget in the sahelian gourma : seasonal and diurnal cycles, *Journal of Hydrology*, 375(1), 161–177, 2009.

- Halkidi, M., Y. Batistakis, and M. Vazirgiannis, Cluster validity methods : part i, ACM Sigmod Record, 31(2), 40–45, 2002.
- Hamilton, R., J. Archbold, and C. Douglas, Meteorology of nigeria and adjacent territory, Quarterly Journal of the Royal Meteorological Society, 71 (309-310), 231–264, 1945.
- Hartigan, J. A., Clustering algorithms, John Wiley & Sons, Inc., 1975.
- Hastenrath, S., Climate and circulation of the tropics, D. Reidel Publishing Company, 112(6), 1988.
- Holben, B., et al., Aeronet-a federated instrument network and data archive for aerosol characterization, *Remote sensing of environment*, 66(1), 1–16, 1998.
- Holben, B., et al., An emerging ground-based aerosol climatology-aerosol optical depth from aeronet, Journal of Geophysical Research, 106(D11), 12,067–12,097, 2001.
- Holton, J. R., The diurnal boundary layer wind oscillation above sloping terrain1, *Tellus*, 19(2), 199–205, 1967.
- Hoose, C., U. Lohmann, R. Bennartz, B. Croft, and G. Lesins, Global simulations of aerosol processing in clouds, Atmos. Chem. Phys, 8(23), 6939–6963, 2008.
- Hoxit, L. R., Diurnal variations in planetary boundary-layer winds over land, Boundary-Layer Meteorology, 8(1), 21–38, 1975.
- Huang, J. Z., M. K. Ng, H. Rong, and Z. Li, Automated variable weighting in k-means type clustering, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 27(5), 657–668, 2005.
- Hubert, L., and P. Arabie, Comparing partitions, *Journal of classification*, 2(1), 193–218, 1985.
- Huneeus, N., et al., Global dust model intercomparison in aerocom phase i, Atmospheric Chemistry and Physics, 11(15), 2011.
- Jain, A. K., Data clustering : 50 years beyond k-means, *ECML/PKDD (1)*, 5211, 3–4, 2008.
- Jain, A. K., and R. C. Dubes, Algorithms for clustering data, Prentice-Hall, Inc., 1988.
- Jickells, T., et al., Global iron connections between desert dust, ocean biogeochemistry, and climate, *science*, 308(5718), 67–71, 2005.
- Jing, L., M. K. Ng, and J. Z. Huang, An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data, *Knowledge and Data Engineering*, *IEEE Transactions on*, 19(8), 1026–1041, 2007.
- Jollois, F.-X., Contribution de la classification automatique à la fouille de données, Ph.D. thesis, Metz, 2003.
- Kaly, F., N. N. Keita, M. Ouattara, N. Awa, S. Thiria, B. Marticorena, and S. Janicot, Two step soft subspace som : une méthode de classification multi-bloc avec sélection de variables., in 6émes Journées thématiques 'Apprentissage Artificiel et Fouille de Données' AAFD'14, 29 et 30 avril 2014 Université Paris 13, Institut Galilée., vol. 10, p. 13, 2014 b.

- Kaufman, Y., I. Koren, L. Remer, D. Tanré, P. Ginoux, and S. Fan, Dust transport and deposition observed from the terra-moderate resolution imaging spectroradiometer (modis) spacecraft over the atlantic ocean, *Journal of Geophysical Research : Atmospheres* (1984–2012), 110(D10), 2005.
- Kiladis, G. N., C. D. Thorncroft, and N. M. Hall, Three-dimensional structure and dynamics of african easterly waves. part i : Observations, *Journal of the atmospheric sciences*, 63(9), 2212–2230, 2006.
- Klose, M., Y. Shao, M. K. Karremann, and A. H. Fink, Sahel dust zone and synoptic background, *Geophysical Research Letters*, 37(9), 2010.
- Knippertz, P., Dust emissions in the west african heat trough-the role of the diurnal cycle and of extratropical disturbances, *Meteorologische Zeitschrift*, 17(5), 553–563, 2008.
- Knippertz, P., and J.-B. W. Stuut, *Mineral Dust : A Key Player in the Earth System*, Springer, 2014.
- Knippertz, P., and M. C. Todd, Mineral dust aerosols over the sahara : Meteorological controls on emission and transport and implications for modeling, *Reviews of Geophysics*, 50(1), 2012.
- Knippertz, P., C. Deutscher, K. Kandler, T. Müller, O. Schulz, and L. Schütz, Dust mobilization due to density currents in the atlas region : Observations from the saharan mineral dust experiment 2006 field campaign, *Journal of Geophysical Research : Atmospheres* (1984–2012), 112(D21), 2007.
- Kocha, C., Interactions entre poussières désertiques et convection profonde en afrique de l'ouest : Observations et modélisation à échelle convective, Ph.D. thesis, Université Paul Sabatier-Toulouse III, 2011.
- Kohonen, T., The self-organizing map, Neurocomputing, 21(1), 1-6, 1998.
- Lafore, J.-P., and M. W. Moncrieff, A numerical investigation of the organization and interaction of the convective and stratiform regions of tropical squall lines, *Journal of the atmospheric sciences*, 46(4), 521–544, 1989.
- Laurent, B., B. Marticorena, G. Bergametti, J. Léon, and N. Mahowald, Modeling mineral dust emissions from the sahara desert using new surface properties and soil database, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 113(D14), 2008.
- Le Borgne, J., La pluviométrie au Sénégal et en Gambie, Université Cheikh Anta Diop, Département de géographie, laboratoire de climatologie, 1988.
- Lebart, L., A. Morineau, and M. Piron, Statistique exploratoire multidimensionnelle, 1997.
- Legrand, M., C. Pietras, G. Brogniez, M. Haeffelin, N. K. Abuhassan, and M. Sicard, A high-accuracy multiwavelength radiometer for in situ measurements in the thermal infrared. part i : Characterization of the instrument, *Journal of Atmospheric and Oceanic Technology*, 17(9), 1203–1214, 2000.
- Léon, J.-F., Y. Derimian, I. Chiapello, D. Tanré, T. Podvin, B. Chatenet, A. Diallo, and C. Deroo, Aerosol vertical distribution and optical properties over m'bour (16.96Ű w; 14.39Ű n), senegal from 2006 to 2008, Atmospheric Chemistry and Physics, 9(23), 9249–9261, 2009.
- Leroux, S., N. M. Hall, and G. N. Kiladis, A climatological study of transient-mean-flow interactions over west africa, *Quarterly Journal of the Royal Meteorological Society*, 136(S1), 397–410, 2010.

- Levin, Z., E. Ganor, and V. Gladstein, The effects of desert particles coated with sulfate on rain formation in the eastern mediterranean, *Journal of Applied Meteorology*, 35(9), 1511–1523, 1996.
- Lothon, M., F. Saïd, F. Lohou, and B. Campistron, Observation of the diurnal cycle in the low troposphere of west africa, *Monthly Weather Review*, 136(9), 3477–3500, 2008.
- Mahowald, N. M., and L. M. Kiehl, Mineral aerosol and cloud interactions, *Geophysical Research Letters*, 30(9), 2003.
- Mahrt, L., Stratified atmospheric boundary layers, *Boundary-Layer Meteorology*, 90(3), 375–396, 1999.
- Marticorena, B., and G. Bergametti, Modelling the atmospheric dust cycle, Journal of Geophysical Research, 100(8), 16,415–16,430, 1995.
- Marticorena, B., et al., Temporal variability of mineral dust concentrations over west africa : analyses of a pluriannual monitoring from the amma sahelian dust transect, Atmospheric Chemistry and Physics, 10(18), 8899–8915, 2010.
- Martiny, N., and I. Chiapello, Assessments for the impact of mineral dust on the meningitis incidence in west africa, *Atmospheric Environment*, 70, 245–253, 2013.
- Mauritsen, T., and G. Svensson, Observations of stably stratified shear-driven atmospheric turbulence at low and high richardson numbers, *Journal of the atmospheric sciences*, 64(2), 645–655, 2007.
- Middleton, N., Dust storms in the middle east, Journal of Arid Environments, 1986.
- Miller, R., and I. Tegen, Climate response to soil dust aerosols, *Journal of climate*, 11(12), 3247–3267, 1998.
- Molesworth, A. M., L. E. Cuevas, S. J. Connor, A. P. Morse, and M. C. Thomson, Environmental risk and meningitis epidemics in africa, *Emerging infectious diseases*, 9(10), 1287, 2003.
- Moulin, C., and I. Chiapello, Impact of human-induced desertification on the intensification of sahel dust emission and export over the last decades, *Geophysical Research Letters*, 33(18), 2006.
- Moulin, C., C. E. Lambert, F. Dulac, and U. Dayan, Control of atmospheric export of dust from north africa by the north atlantic oscillation, *Nature*, 387(6634), 691, 1997.
- Nadif, M., and G. Govaert, Binary clustering with missing data, Applied stochastic models and data analysis, 9(1), 59–71, 1993.
- N'Tchayi Mbourou, G., J. Bertrand, M. Legrand, and J. Baudet, Temporal and spatial variations of the atmospheric dust loading throughout west africa over the last thirty years, in *Annales Geophysicae*, vol. 12, pp. 265–273, Springer, 1994.
- N'Tchayi Mbourou, G., J. Bertrand, and S. Nicholson, The diurnal and seasonal cycles of wind-borne dust over africa north of the equator, *Journal of Applied Meteorology*, 36(7), 868–882, 1997.
- Ogunjobi, K., Z. He, and C. Simmer, Spectral aerosol optical properties from aeronet sun-photometric measurements over west africa, *Atmospheric Research*, 88(2), 89–107, 2008.

- Ouattara, M., Développement et mise en place d'une méthode de classification multiblocs : application aux données de l'oqai., Ph.D. thesis, Conservatoire national des arts et metiers-CNAM, 2014.
- Ozer, P., Les lithométéores en région sahélienne. revue internationale d'écologie et de géographie tropicales., 24, 1-317., 2001.
- Pandithurai, G., R. Pinker, O. Dubovik, B. Holben, and T. Aro, Remote sensing of aerosol optical characteristics in sub-sahel, west africa, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 106(D22), 28,347–28,356, 2001.
- Parker, D., R. Burton, A. Diongue-Niang, R. Ellis, M. Felton, C. Taylor, C. Thorncroft, P. Bessemoulin, and A. Tompkins, The diurnal cycle of the west african monsoon circulation, *Quarterly Journal of the Royal Meteorological Society*, 131(611), 2839–2860, 2005.
- Parsons, L., E. Haque, and H. Liu, Subspace clustering for high dimensional data : a review, ACM SIGKDD Explorations Newsletter, 6(1), 90–105, 2004.
- Prospero, J. M., Long-term measurements of the transport of african mineral dust to the southeastern united states : Implications for regional air quality, *Journal of Geophysical Research : Atmospheres (1984–2012), 104* (D13), 15,917–15,927, 1999.
- Prospero, J. M., and R. T. Nees, Impact of the north african drought and el nino on mineral dust in the barbados trade winds, 1986.
- Prospero, J. M., P. Ginoux, O. Torres, S. E. Nicholson, and T. E. Gill, Environmental characterization of global sources of atmospheric soil dust identified with the nimbus 7 total ozone mapping spectrometer (toms) absorbing aerosol product, *Reviews of geophysics*, 40(1), 2–1, 2002.
- Prospero, J. M., E. Blades, R. Naidu, G. Mathison, H. Thani, and M. C. Lavoie, Relationship between african dust carried in the atlantic trade winds and surges in pediatric asthma attendances in the caribbean, *International journal of biometeorology*, 52(8), 823–832, 2008.
- Pruppacher, H., and J. Klett, Microphysics of clouds and precipitation 1997.
- Rajot, J.-L., Wind blown sediment mass budget of sahelian village land units in niger, Bulletin de la Société Géologique de France, 172(5), 523–531, 2001.
- Schepanski, K., I. Tegen, B. Laurent, B. Heinold, and A. Macke, A new saharan dust source activation frequency map derived from msg-seviri ir-channels, *Geophysical Re*search Letters, 34(18), 2007.
- Schepanski, K., I. Tegen, M. Todd, B. Heinold, G. Bönisch, B. Laurent, and A. Macke, Meteorological processes forcing saharan dust emission inferred from msg-seviri observations of subdaily dust source activation and numerical models, *Journal of Geophysical Research : Atmospheres (1984–2012), 114* (D10), 2009.
- Scheuvens, D., L. Schütz, K. Kandler, M. Ebert, and S. Weinbruch, Bulk composition of northern african dust and its source sediments-a compilation, *Earth-Science Reviews*, 116, 170–194, 2013.
- Schmechtig, C., B. Marticorena, B. Chatenet, G. Bergametti, J.-L. Rajot, and A. Coman, Simulation of the mineral dust content over western africa from the event to the annual scale with the chimere-dust model, *Atmospheric Chemistry and Physics*, 11(14), 7185– 7207, 2011.

- Schulz, M., and L. J. Lanzerotti, Particle diffusion in the radiation belts, vol. 7, Springer Science & Business Media, 2012.
- Schütz, L., R. JAENICKE, and H. PIETREK, Saharan dust transport over the north atlantic ocean, *Geological Society of America Special Papers*, 186, 87–100, 1981.
- Sivakumar, M., Predicting rainy season potential from the onset of rains in southern sahelian and sudanian climatic zones of west africa, Agricultural and Forest Meteorology, 42(4), 295–305, 1988.
- Sokolik, I. N., and O. B. Toon, Incorporation of mineralogical composition into models of the radiative properties of mineral aerosol from uv to ir wavelengths, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 104(D8), 9423–9444, 1999.
- Sultan, B., and S. Janicot, The west african monsoon dynamics. part ii : The 'preonset' and 'onset' of the summer monsoon, *Journal of climate*, 16(21), 3407–3427, 2003 b.
- Sultan, B., K. Labadi, J.-F. Guégan, S. Janicot, et al., Climate drives the meningitis epidemics onset in west africa, *PLoS Med*, 2(1), e6, 2005.
- Sultan, B., S. Janicot, and P. Drobinski, Characterization of the diurnal cycle of the west african monsoon around the monsoon onset, *Journal of climate*, 20(15), 4014–4032, 2007.
- Swap, R., M. Garstang, S. Greco, R. Talbot, and P. Kållberg, Saharan dust in the amazon basin, *Tellus B*, 44(2), 133–149, 1992.
- TALBOT, R. W., J. HEPPELL, R. R. DOZOIS, and R. W. BEART, Vascular complications of inflammatory bowel disease, in *Mayo Clinic Proceedings*, vol. 61, pp. 140–145, Elsevier, 1986.
- Thomson, M. C., A. M. Molesworth, M. H. Djingarey, K. Yameogo, F. Belanger, and L. E. Cuevas, Potential of environmental models to predict meningitis epidemics in africa, *Tropical Medicine & International Health*, 11(6), 781–788, 2006.
- Todd, M., et al., Quantifying uncertainty in estimates of mineral dust flux : An intercomparison of model performance over the bodélé depression, northern chad. j. geophys. res., 113, d24107, doi :10.1029/2008jd010476., 2008.
- Todd, M. C., R. Washington, J. V. Martins, O. Dubovik, G. Lizcano, S. M'bainayel, and S. Engelstaedter, Mineral dust emission from the bodélé depression, northern chad, during bodex 2005, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 112(D6), 2007.
- Tompkins, A., C. Cardinali, J.-J. Morcrette, and M. Rodwell, Influence of aerosol climatology on forecasts of the african easterly jet, *Geophysical research letters*, 32(10), 2005.
- Torres, O., P. Bhartia, J. Herman, Z. Ahmad, and J. Gleason, Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation : Theoretical basis, *Journal of Geophysical Research : Atmospheres (1984–2012)*, 103(D14), 17,099– 17,110, 1998.
- Tucker, C. J., H. E. Dregne, and W. W. Newcomb, Expansion and contraction of the sahara desert from 1980 to 1990, *Science*, 253(5017), 299–301, 1991.
- Vesanto, J., J. Himberg, E. Alhoniemi, and J. Parhankangas, SOM toolbox for Matlab 5, Citeseer, 2000.

- Washington, R., and M. C. Todd, Atmospheric controls on mineral dust emission from the bodélé depression, chad : The role of the low level jet, *Geophysical Research Letters*, 32(17), 2005.
- Washington, R., M. Todd, N. J. Middleton, and A. S. Goudie, Dust-storm source areas determined by the total ozone monitoring spectrometer and surface observations, *Annals* of the Association of American Geographers, 93(2), 297–313, 2003.
- Washington, R., et al., Links between topography, wind, deflation, lakes and dust : The case of the bodele depression, chad, geophys. res. lett., 33, 109401, doi :10.1029/2006gl025827., 2006.
- Weisman, M. L., and R. Rotunno, A theory for strong long-lived squall lines, Journal of the atmospheric sciences, 61(4), 361–382, 2004.
- Yacoub, M., F. Badran, and S. Thiria, A topological hierarchical clustering : Application to ocean color classification, in *Artificial Neural Networks-ICANN 2001*, pp. 492–499, Springer, 2001.
- Yahi, H., B. Marticorena, S. Thiria, B. Chatenet, C. Schmechtig, J.-L. Rajot, and M. Crepon, Statistical relationship between surface pm10 concentration and aerosol optical depth over the sahel as a function of weather type, using neural network methodology, *Journal of Geophysical Research : Atmospheres*, 118(23), 13–265, 2013.
- Yu, H., et al., Quantification of trans-atlantic dust transport from seven-year (2007–2013) record of calipso lidar measurements, *Remote Sensing of Environment*, 159, 232–249, 2015.

## Annexe A

## Quelques figures

#### A.1 Figures régimes de temps bruts



#### A.1.1 régimes de temps bruts

Figure A.1 – Cycle Diurne d'occurrence des types de temps bruts en saison sèche pour les trois stations (A : Banizoumbou; B : Cinzan; C : Mbour).

	Janvier	Février	Mars	Avril	Mai	Octobre	Novembre	Déce
Oh	-		-	-	-	-		-
Зh					0	-		Canal
6 h		0		-	0	-	0	0
9 h	-	0	0		-	0	Carlos and a	Conception in the second
12 h	0	0		-	-	•	-	-
15 h	0		-		-	-	-	
18 h	0	-	-	-	-	-	0	0
21 h	0	-	-		0	•	0	0
Banizoumbou								
	Janvier	Février	Mars	Avril	Mai	Octobre	Novembre	Décen
0 h	-	-	-	•	0	-	-	-
Зh	-		-	0	-	-	-	-
6 h	-	0	-	•	-	-	-	9
9 h	9	-	-	-	0	-	0	9
12 h	-	-	-	•	0	-	•	9
15 h			-	-	-	-	-	9
18 h	-		-	-	•	-	-	9
21 h	9	-	-	-	•	-	-	-
Cinzana								
	Jan	Fév	Mars	Avril	Mai	Oct	Nov	Déc
Oh	Name of Street o	- Coccast	-	-	-	-	-	-
3 h			-	-	-	-	-	-
6 h			-	-	0			
9 h				-	0	-		
12 h	-		-	-	-	-	0	0
15 h	9		-	-	0	-	-	
18 h	-		-	-	0	-	0	0
21 h			-	0	0	-	0	-
V	VR1 💻	WR2	W	R3 🔼	WR4	- W	R5 -	WR6
				Mbour				

Figure A.2 –  $\,$  : Diagramme diurne-mensuel des régimes de temps sur les trois stations (Banizoumbou, Cinzana, Mbour).



#### A.1.2 régimes de temps desaisonalisés

Figure A.3 – Cycle Diurne d'occurrence des types de temps bruts en saison sèche pour les trois stations (a : Banizoumbou; b : Cinzan; c : Mbour).

#### A.1.3 Système d'alerte

Afin de voir les AI correspondant au 15% des pm bruts les plus fort sur l'ensemble des trois stations, on a tracé les figures. Quel que soit le régime, on retrouve des localisations communes des zones de fort AI. La zone où les AI sont les plus forts est localisée è l'extrême est du domaine et centrée à environ 16°N. Elle se situe en aval de la dépression de Bolélé, une zone reconnue comme la plus active du Sahara [*Prospero et al.*, 2002].

Pour toutes les trois stations, les AI les plus forts correspondent aux régimes WR1 et WR2 caractéristique du flux de mousson comme vu sur les régimes de temps bruts. Ces forts AI sont localisées sur la partie nord du domaine, en zone saharienne, en accord avec *Engelstaedter and Washington* [2007] ) qui ont montrée que les AI maximum sont maximum en été.



A: Banizoumbou



B : Cinzana



Figure A.4 – Carte (vent, geopotentiel (iso lignes)) à 1000 hPa et AI (Couleur) pour15% des pm bruts les plus fort sur l?ensemble des trois stations.



A : Banizoumbou





Figure A.5 – : Carte (vent, geopotentiel (iso lignes)) à 1000 hPa et AI (Couleur) pour 15% des pm les plus fort sur la base désaisonnalisée pour l?ensemble des trois stations (A :Banizoumbou; B : Cinzana; C : Mbour ).

#### A.2 Poster et Article

**POSTER 1** : Weather types identification at M'Bour (Senegal) and implication on the relationships between PM10 concentrations and aerosols optical thickness (AOT) Conférence Internationale AMMA Toulouse, France, du 02 Juillet 2012 au 06 Juillet 2012.



Figure A.6 – Weather types identification at M'Bour (Senegal) and implication on the relationships between  $PM_{10}$  concentrations and aerosols optical thickness (AOT) Conférence Internationale AMMA Toulouse, France, du 02 Juillet 2012 au 06 Juillet 2012.

POSTER 2 : Multi-annual variability of mineral dust concentration over West Africa monitored by the AMMA Sahelian Dust Transect Conférence : European Geosciences Union avril 2013 à Vienne/Autriche (EGU 2013).



Figure A.7 – Multi-annual variability of mineral dust concentration over West Africa monitored by the AMMA Sahelian Dust Transect Conférence : European Geosciences Union avril 2013 à Vienne/Autriche (EGU 2013).

## Two step soft subspace SOM : une méthode de classification multi-bloc avec sélection de variables

F. Kaly<sup>\*</sup>, <sup>\*\*,\*\*\*</sup> Ndèye Niang<sup>\*\*\*\*</sup> Mory Ouattara<sup>\*\*\*\*</sup> Awa Niang<sup>\*</sup> Sylvie Thiria<sup>\*\*\*</sup> Beatrice Marticorena<sup>\*\*</sup> Serge Janicot<sup>\*\*\*</sup>

\*LTI, UCAD, Dakar, Sénégal

\*\* LISA, UMR CNRS 7583 ; IPSL ; UPEC ; UPD, Créteil, France

\*\*\*LOCEAN, UMR 7159 CNRS-IRD-UPMC-MNHN; IPSL, Paris, France

\*\*\*\*Statistique Appliquée, CNAM 292, rue Saint Martin, 75141 Paris Cedex 03, France

**Résumé.** Nous proposons une stratégie de classification de données multi-blocs basée sur l'utilisation de la méthode de soft subspace clustering 2S-SOM dans un processus hiérarchique à deux niveaux combiné à des tests statistiques. Une première application de la méthode 2S-SOM fournit un système de poids évaluant les contributions relatives des variables et des blocs aux groupes d'observations. Nous proposons une procédure de test statistique permettant de sélectionner les variables significativement pertinentes. 2S-SOM est à nouveau utilisée sur ces dernières pour déterminer la partition finale des observations. La méthode est évaluée sur des données simulées et réelles. En particulier, l'application sur des données météorologiques montre que la sélection des variables au niveau 1 facilite l'interprétation des classes obtenues.

#### **1** Introduction

Les méthodes de classification non-supervisées (ou clustering) permettent d'explorer des données non-labélisées dans le but de trouver des groupes d'observations homogènes et bien séparés. Les récentes avancées technologiques en termes de capacité de stockage d'informations d'une part, et la multiplication des sources d'informations d'autre part, contribuent à la mise en place de bases de données complexes et de grande dimension. De plus ces données peuvent présenter une structure en plusieurs blocs de variables caractérisant chacune une vue particulière sur les données recueillies selon une thématique spécifique, on parle de données multi-vues ou multi-blocs. Or, la majorité des mesures de distance perdent leur pouvoir discriminant au fur et à mesure que la dimension augmente; les observations étant pratiquement toutes équidistantes les unes par rapport aux autres, Parsons et al. (2004). En outre, en l'absence d'une structure globale de corrélation entre les variables (ce qui est souvent le cas en grande dimension à cause de la présence possible de variables souvent distribuées uniformément), la similarité entre deux observations est souvent portée par un nombre limité de variables. Les classes sont donc à rechercher dans des sous-espaces de l'espace initial, on parle alors de méthode de subspace clustering. Le principe des méthodes de subspace clustering reposent sur la recherche de sous espaces de l'espace initial permettant une meilleure détection et interprétation des groupes d'individus Agrawal; Kriegel et al. (2009). Des approches récentes basées sur l'introduction dans la méthode des K-moyennes d'une pondération des variables ou des blocs permettent de prendre en compte en plus de la grande dimension, la structure multi-blocs (Jing et al., 2007; Chen et Ye, 2012). Ce sont des méthodes de soft subspace clustering. Plus récemment, ces approches ont été étendues, à travers une nouvelle méthode 2S-SOM Ouattara et al. (2014), aux cartes auto-organisées ou self organizing maps (SOM) Kohonen (1998) permettant ainsi d'exploiter les propriétés de visualisation de SOM.

Nous proposons une approche hiérarchique de sélection de variables en classification fondée sur une double utilisation de 2S-SOM. Dans une première application, 2S-SOM fournit un système de poids à partir duquel on recherche des variables ou des blocs pertinents à l'aide d'un test d'hypothèse statistique. Ensuite, une deuxième utilisation de 2S-SOM sur les variables sélectionnées fournit la partition recherchée. La méthode proposée est présentée en section 2 après les notations et la description de 2S-SOM. La section 3 présente son illustration sur des données réelles. La section 4 est consacrée à la conclusion.

#### 2 Soft-Subspace clustering basé sur SOM : 2S-SOM

Nous disposons de N observations  $z_i$  décrites par p variables divisées en B blocs. On recherche une partition des observations en K classes.

Les notations suivantes seront utilisées :

- $\mathcal{V} = \{z^j, j = 1, ..., p\}$  l'ensemble des variables divisé en *B* blocs de  $p_b$  variables tels que  $p_1 + ... + p_b + ... + p_B = p$ .
- $\alpha$  est une matrice  $K \times B$  où K désigne le nombre de classes c dans Z,  $\alpha_{cb}$  est le poids du bloc b dans la classe c de Z.
- $\beta = [\beta_1, \dots, \beta_B]$  est une matrice  $K \times p$  où  $\beta_b$  est une matrice de dimension  $K \times p_b$  définissant les poids  $\beta_{cbj}(j = 1, \dots, p_b)$  sur les variables du bloc *b* pour chaque *c* de *Z*.

#### 2.1 2S-SOM

Les cartes topologiques auto-organisées sont utilisées pour quantifier et visualiser des données numériques de grande dimension dans un espace de faible dimension, généralement 1 ou 2 dimensions, appelé carte topologique. De manière générale, la méthode suppose l'existence d'une carte discrète C ayant Kcellules c structurées par des graphes non-orientés permettant de définir a priori une distance entre les cellules. Dans la suite, nous utiliserons indifféremment les termes cellule ou classe. Chaque cellule de la carte est représentée par un vecteur référent ou prototype  $w_c$  synthétisant l'information de la cellule. L'algorithme SOM initial des cartes topologiques consiste à minimiser de manière itérative en deux phases la fonction de coût (Kohonen, 1998) :

$$\mathcal{J}_{SOM}^T(\mathcal{Z}, \mathcal{W}) = \sum_{i=1}^N \sum_{c \in C} \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c)) \sum_{j=1}^p (z_{ij} - \omega_{cj})^2$$
(1)

Dans cette expression,  $\mathcal{X}(z_i) = \underset{c \in \mathcal{C}}{\operatorname{argmin}} (\sum_{r \in \mathcal{C}} \mathcal{K}^T(\sigma(r, c)) ||z_i - w_c||^2)$  re-

présente une fonction d'affectation des observations  $z_i$  à la cellule c dont le vecteur référent est le plus proche, W est l'ensemble des vecteurs référents  $\omega_c$  des cellules c.  $\mathcal{K}^T$  et le paramètre T associé définissent respectivement une fonction décroissante de contrainte de voisinage définie entre deux cellules c et r de la carte et la taille du voisinage d'une cellule.

Dans le cas des données en bloc, l'approche de type subspace clustering 2S-SOM repose sur une modification de la fonction de coût de SOM en introduisant un double système de poids  $\alpha_{cb}(b = 1, ..., B)$  et  $\beta_{cbj}(j = 1, ..., p_k)$ 

définis respectivement sur les blocs et sur les variables pour chaque cellule  $c \in C$ . La classification et les poids relatifs à la pertinence des blocs et des variables sont donc obtenus par minimisation de la fonction objectif  $J_{2SSOM}^T$  définie par la relation suivante :

$$\mathcal{J}_{2S-SOM}^{T}(\mathcal{X}, \mathcal{W}, \alpha, \beta) = \sum_{c \in \mathcal{C}} \left( \sum_{b=1}^{B} \left( \sum_{i=1}^{N} \alpha_{cb} \mathcal{K}^{T}(\sigma(\mathcal{X}(z_{i}), c)) d_{\beta_{cb}} + J_{cb} \right) + I_{c} \right)$$
(2)

avec  $d_{\beta_{cb}} = \sum_{j=1}^{p_b} \beta_{cbj} (z_{ibj} - \omega_{ckj})^2$  et sous les contraintes :

$$\begin{cases} \sum_{j=1}^{p_b} \beta_{cbj} = 1, \ \beta_{cbj} \in [0,1], \forall \ c \in \mathcal{C}, \forall b \\ \\ \sum_{b=1}^{B} \alpha_{cb} = 1, \ \alpha_{cb} \in [0,1], \ \forall \ c \in \mathcal{C} \end{cases}$$
(3)

 $I_c = \lambda \sum_{b=1}^{B} \alpha_{cb} log(\alpha_{cb})$  et  $J_{cb} = \eta \sum_{j=1}^{p_b} \beta_{cbj} log(\beta_{cbj})$  représentent des termes entropies négatives pondérées et associées aux vecteurs poids relatifs aux blocs et aux variables pour une cellule c. La minimisation de la fonction de coût  $J_{2S-SOM}^T$  s'effectue de façon alternée en quatre étapes dont les deux premières phases d'affectation des observations aux classes et d'actualisation des vecteurs référents sont identiques à celles de la méthode SOM. Les valeurs des poids sont supposées connues et fixées à leur valeur courante, on a alors :

— étape 1 : Les référents W sont connus et fixés, les observations sont affectées aux cellules en respectant l'équation (4) :

$$\mathcal{X}(z_i) = \operatorname{argmin}_{c \in \mathcal{C}} \left( \sum_{r \in \mathcal{C}}^{B} \left( \sum_{b=1}^{B} \alpha_{cb} \mathcal{K}^T(\sigma(r, c)) d_{\beta_{cb}} \right) \right)$$
(4)

— étape 2 : Actualisation des centres de classe à l'aide de : (5)

$$\omega_c^T = \frac{\sum_{i=1}^N \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c)) z_i}{\sum_{i=1}^N \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c))}$$
(5)

A l'aide du lagrangien de la fonction  $J_{2S-SOM}^T$  relativement aux quantités  $\alpha$  et  $\beta$  on détermine les poids  $\alpha$  et  $\beta$  associés respectivement aux blocs et aux

variables. Ainsi, on obtient :

— Etape 3 : les paramètres  $\mathcal{X}$ ,  $\omega$  et  $\beta$  connus et fixés à leurs valeurs courantes alors on a :

$$\alpha_{cb} = \frac{exp(\frac{-\Psi_{cb}}{\lambda})}{\sum_{b=1}^{B} exp(\frac{-\Psi_{cb}}{\lambda})}$$
(6)

avec

$$\Psi_{cb} = \sum_{z_i \in r, r \neq c} \mathcal{K}^T(r, c) d_{\beta_{cb}} + \mathcal{K}^T(c, c) \sum_{z_i \in c} d_{\beta_{cb}}$$
(7)

— Etape 4 : de manière identique, si les paramètres  $\mathcal{X}$ ,  $\omega$  et  $\alpha$  sont connus et fixés à leurs valeurs courantes alors , on a :

$$\beta_{cbj} = \frac{\exp(\frac{-\Phi_{cbj}}{\eta})}{\sum_{j=1}^{p_b} \exp(\frac{-\Phi_{cbj}}{\eta})}$$
(8)

avec

$$\Phi_{cbj} = \sum_{z_i \in r, r \neq c} \alpha_{cb} \mathcal{K}^T(r, c) (z_{ibj} - \omega_{cbj})^2 + \mathcal{K}^T(c, c) \sum_{z_i \in c} \alpha_{cb} (z_{ibj} - \omega_{cbj})^2 \quad (9)$$

Le poids d'une variable ou d'un bloc sera donc d'autant plus important qu'il minimise simultanément la somme des écarts entre les référents  $w_c$  et les observations appartenant à la cellule c et aux cellules r du voisinage T de la cellule c. Les coefficients de pondération  $\alpha_{cb}$  et  $\beta_{cbj}$  définis par 2S-SOM indiquent respectivement l'importance relative des blocs et des variables dans les classes. Ainsi, plus le poids d'un bloc b ou d'une variable j est important, plus le bloc ou la variable contribue à la définition de la classe au sens où elle permet de réduire la variabilité des observations dans la cellule et dans son voisinage proche. Finalement, à la convergence, 2S-SOM fournit d'une part une carte topologique permettant de visualiser les données et d'autre part des systèmes de poids pour les classes de la classification.

#### 2.2 Sélection des variables pertinentes

Au niveau des cellules, la pertinence d'une variable ou d'un bloc est fournie directement par son poids défini par 2S-SOM.

Les poids  $\alpha_{cb}$  et  $\beta_{cbj}$  sont définis pour un bloc b et pour une cellule c sous les contraintes  $\sum_{j}^{p_b} \beta_{cbj} = 1$  et  $\sum_{b}^{B} \alpha_{cb} = 1$ . La contribution moyenne d'une variable à une cellule de la carte est donc  $\frac{1}{p_b}$  que l'on peut utiliser comme un seuil de sélection des variables pertinentes. Une variable telle que  $\beta_{cbj} < \frac{1}{p_b}$   $\forall c \in C$  sera non-pertinente donc éliminée.

Dans certains cas, la taille de la carte peut conduire à un grand nombre de cellules. Il est alors possible d'appliquer un algorithme de classification ascendante hiérarchique (CAH) sous contrainte de voisinage sur la matrice composée des vecteurs référents pour réduire ce grand nombre de cellules en un nombre restreint K' de classes contenant  $n_{k'}$  cellules par classe k' (Gordon, 1996).

Ce regroupement des cellules engendre la nécessité de définir le poids et la pertinence d'une variable dans les classes. Nous proposons de prendre la moyenne des poids des cellules constituant la classe, soit  $\gamma_{k'bj} = \frac{1}{n_{k'}} \sum_{l=1}^{n_{k'}} \beta_{c_l^{k'}bj}$  pour la variable *j* du bloc *b* et la classe *k'* composée des cellules  $c_l^{k'}$ .

De même, on définit  $\delta_{k'b} = \frac{1}{n_{k'}} \sum_{l=1}^{n_{k'}} \alpha_{c_l^{k'}b}$  la moyenne des poids du bloc b sur les cellules de la classe k'.

Pour évaluer la pertinence d'une variable dans une classe de la CAH, nous proposons d'utiliser un test statistique basé sur le principe de la valeur-test proposée par Lebart et al. (1997). On désigne par  $\gamma_{bj} = \frac{1}{N_{cell}} \sum_{c \in C} \beta_{cbj}, s_{bj}^2$ ,  $s_{k'bj}^2 = \frac{N_{cell} - n_{k'}}{N_{cell} - 1} \frac{s_{bj}^2}{n_{k'}}$  respectivement la moyenne, la variance des poids de la variable *j* du bloc *b* pour l'ensemble des cellules et la variance des poids de la variable *j* du bloc *b* dans la classe *k'* de la CAH. La valeur-test, pour les poids d'une variable *j* dans une classe *k'*, se définit alors par :

$$t_{k'bj} = \frac{\gamma_{k'bj} - \gamma_{bj}}{s_{k'bj}}$$

La valeur test  $t_{k'bj}$  peut se lire comme la statistique d'un test de comparaison de moyennes où, sous l'hypothèse nulle de tirage au hasard de  $n_{k'}$  cellules parmi  $N_{cell}$ , elle suivrait de manière asymptotique la loi normale centrée réduite. Pour les niveaux de risque usuels (5%), on considérera donc que la

différence entre le poids moyen dans la classe et le poids moyen sur la carte est significative lorsque la valeur absolue de la valeur test est supérieure à 2.

Ainsi, au niveau de la valeur test, la pertinence d'une variable ou d'un bloc dans une classe de la CAH est relative à la distribution des poids de 2S-SOM pour l'ensemble des cellules de la carte.

Pour une classe la contribution moyenne des variables est :

$$\frac{1}{p_b} \sum_{j=1}^{p_b} \gamma_{k'bj} = \frac{1}{p_b} \sum_{j=1}^{p_b} \frac{1}{n_{k'}} \sum_{l=1}^{n_{k'}} \beta_{c_l^{k'}bj} = \frac{1}{p_b}$$

Par ailleurs, compte tenu des propriétés de 2S-SOM (Ouattara et al., 2014), les variables de bruit qui ont en général des poids  $\gamma_{bj} < \frac{1}{p_b}$  peuvent également être sélectionnées par la valeur test. Ainsi, parmi les variables sélectionnées par la valeur test, seront éliminées celles dont le poids moyen  $\gamma_{bj} < \frac{1}{p_b}$ .

Il est alors possible de sélectionner l'ensemble des variables pertinentes pour les blocs à travers une première application de la méthode 2S-SOM. L'approche hiérarchique que nous proposons ici, consiste ensuite, à appliquer de nouveau 2S-SOM sur les variables sélectionnées pour obtenir la partition recherchée. La réduction du nombre de variables permet de simplifier d'une part l'interprétation des classes obtenues, d'autre part elle fournit des partitions de meilleure qualité en terme d'indices de pureté et de NMI comme cela est illustré ci-dessous sur des données labellisées. Les notions de pureté et de NMI sont detaillées dans l'annexe. sNous évaluons les performances de notre approche de sélection des variables sur un jeu de données réelles issus de l'UCI et sur un jeu de données simulées :

- Le jeu de données "Image Segmentation" (IS) contient 2310 observations et 19 variables décrivant les pixels de 7 images. Chaque observation représente un point d'une image décrite par deux blocs de 9 et 10 variables caractérisant le contraste de couleur de ce point sur l'image. Chaque observation possède une étiquette comprise entre 1 et 7.
- Les données simulées D contiennent 400 observations divisées en 4 classes de 100 observations décrites par 4 blocs de variables. Elle contiennent 4 blocs de 5 variables. Les blocs contiennent respectivement 2, 2, 4 et 4 variables de bruit.

Le tableau 1 présente la valeur moyenne des performances de 2S-SOM au niveau 1 (avant sélection), au niveau 2 (après sélection) grâce au test d'hypo-

Data	Index		SOM	2S-SOM	$2S-SOM_{VT}$	EWKM	FGKM
D	NMI	mu	0.11	0.82	0.81	0.43	0.32
		std	0.04	0.08	0.06	0.01	0.34
	Pureté	mu	0.35	0.89	0.90	0.32	0.56
		std	0.06	0.10	0.11	0.01	0.22
IS	NMI	mu	0.60	0.60	0.64	0.53	0.40
		std	0.02	0.08	0.06	0.07	0.14
	Pureté	mu	0.61	0.61	0.64	0.61	0.63
		std	0.03	0.06	0.05	0.05	0.05

thèse et par rapport à la sélection brute dans laquelle une variable est jugée importante si son poids est supérieur à  $1/p_b$ .

TAB. 1 – Comparaisons des performances des méthodes 2S-SOM avant (2S-SOM) et après la sélection des variables (2S-SOM<sub>VT</sub>) avec les performances des méthodes basées sur la méthode des K-moyennes (EWKM, FGKM)

	$C_1$	$C_2$	$C_3$	$C_4$
Bloc	(2,3)	3	4	1
Var	(2-3, 3)	0	1	3

TAB. 2 – Variables pertinentes

Le tableau 2 montre qu'on sélectionne effectivement les variables noninformatives pour la classification puisque la suppression des variables nonpertinentes ne dégrade pas les performances de classification sur les bases D et IS. Par ailleurs, les performances de la méthodes  $2S-SOM_{VT}$  restent meilleures que celles des méthodes EWKM, FGKM.

#### **3** Application

#### 3.1 Données

Les données utilisées, sont les champs météorologiques ré-analysées ERAInterim du centre européen de prévision météorologique (ECMWF). Les observations sont fournies toutes les 3 heures selon des grilles superposées en altitude correspondant aux 9 niveaux de pressions (1000 hpa, 925 hpa, 850 hpa,

700 hpa, 500 hpa, 400 hpa, 300 hpa, 250 hpa, 200hpa) et dont chaque maille carrée de 0.5° de coté décrit l'état de l'atmosphère. Ces données sont initialement constituées de 7 paramètres météorologiques que sont la température (T), l'humidité spécifique (Q), le géopotentiel (Z), le vent zonal (U) et le vent méridien (V), la vitesse verticale du vent (W) et la hauteur de la couche limite (BLH) qui servent à la classification des données. Chacun des 6 premiers blocs initiaux est constitué de 153 variables et 2549 observations correspondant à l'état de l'atmosphère en une heure donnée. Le bloc BLH est constitué de 17 variables. L'étude porte sur les données de la saison sèche (d'Octobre à Mai) des années allant de 2006 à 2010 de la station de Mbour (Sénégal). Ainsi, on a formé une base de données de 2549 observations sur 935 variables. Dans ce travail, les blocs sont structurés selon deux thématiques : l'une spécifique à la thermodynamique de l'atmosphère, il s'agit des paramètres T, U, V et Z composant un bloc de 612 variables et le deuxième bloc correspond aux mouvements de l'atmosphère (Q, W, BLH) et est composé de 323 variables. Une carte topologique a été réalisée avec 2S-SOM introduisant pour chaque bloc et pour chaque variable un système de poids adaptatifs. A partir de ces poids on identifie les variables et les blocs les plus pertinents dans la classification.

#### 3.2 Résultats

L'algorithme a été appliqué sur la base de données décrite ci-dessus afin de procéder d'abord à une sélection de variables et ensuite sur les variables sélectionnées pour évaluer la pertinence des résultats. Au niveau 1 de l'algorithme, plusieurs applications de 2S-SOM ont été réalisées en faisant varier les paramètres d'initialisation. La meilleure carte en termes de quantification vectorielle est retenue et on a obtenu une carte de  $21 \times 12$  soit 252 neurones. La projection des variables du paramètre BLH de la base de données sur la carte obtenue montre une bonne organisation de la topologie des observations sur la carte (Cf. Fig 1).



FIG. 1 – Représentation de la topologie des variables du paramètre BLH sur la carte topologique (la barre des couleurs représente l'échelle des données, des plus faibles (bleu) aux plus fortes (rouge))

Les autres variables n'ont pas été représentées à cause du nombre important de variables d'apprentissage, mais un travail préalable a été effectué pour s'assurer de leur bonne organisation avant de procéder à l'exploitation de cette carte topologique.

La figure 2 suivante présente le nombre de données captées par chaque neurone de la carte. Nous remarquons qu'il existe une distribution relativement homogène des données sur la carte.
F. Kaly et al.



FIG. 2 – Cardinalité des neurones de la carte topologique

La figure 3(a) donne une représentation graphique des poids  $\alpha_{cb}$  définis par 2S-SOM sur les blocs par rapport aux cellules de la carte sur les deux niveaux. Il ressort de l'analyse de cette dernière, que pour le niveau 1 comme au niveau 2, globalement le bloc 1 est plus important que le bloc 2 sur les cellules. Ce qui montre que la suppression des variables de bruit ne change pas fondamentalement l'importance relative des blocs.

Etant donné le grand nombre de neurones de la carte, nous avons regroupé les neurones, en appliquant une classification ascendante hiérarchique (CAH) avec le critère de Ward comme critère d'agrégation des neurones de la carte auto-organisatrice. La figure 3(b) suivante présente les indices de Davies Bouldin Davies et Bouldin (1979) pour plusieurs classifications en faisant varier le nombre de classes. Nous utilisons ce dernier pour choisir le bon nombre de classes, en l'occurrence 4 classes fig 3(b). les poids  $\beta$  de ces 4 classes servent ensuite à sélectionner les variables pertinentes pour déterminer la carte finale. Les figures 2 et 3(c) présentent la cardinalité des observations par cellule et la représentation des classes de la CAH. Le nombre de données de chaque classe présenté sur le tableau 3 montre une répartition homogène des données dans les classes.

Sélection de variables en classification



(a) Les poids des blocs au niveau 1 (à droite) et au niveau 2 (à gauche) pour les cellules de la carte



FIG. 3 – Caractéristiques des cellules et la répartition en classes

	classe1	classe2	classe3	classe4
Ν	650	525	556	818
Р	25.5%	20.6%	21,80%	32.09%

TAB. 3 – nombre de donn $\tilde{A}$  ces dans chaque classe et le pourcentage associ $\tilde{A}$  c)

La sélection des variables a été réalisée en deux étapes. Nous avons utilisé d'abord le principe des valeurs tests présenté dans la section 2.2 pour évaluer les variables pertinentes dans les 4 classes obtenues par la CAH afin de procéder à un premier filtre tenant compte de la variabilité intra classe des poids  $\beta$  de chaque variable. Ensuite sur les variables retenues par la valeur test, on a sélectionné les variables les plus pertinentes en se fixant un seuil à  $1/p_b$  avec  $p_b$  étant le nombre de variables du bloc b. On dira qu'une variable est importante

si son poids moyen dans une classe est significativement supérieur à  $1/p_b$ . La figure 4, montre le poids des variables de chaque bloc dans chaque classe au niveau 1. La sélection des données effectuée montre que 59% (548 variables dont 408 pour le bloc 1 et 150 pour le bloc 2) des variables ne sont pertinentes pour aucune classe. Les 41% (377 variables dont 204 pour le bloc 1 et 173 pour le bloc 2) des variables restantes ont servi à déterminer la partition finale au niveau 2.



FIG. 4 – Les poids des variables des blocs au niveau 1 (à gauche) et au niveau 2 (à droite) pour les cellules de la carte

#### **3.3** Description des classes

La figure 5 caractérise les 5 classes par rapport aux vents qui décrivent la direction et l'intensité du vent. On observe que les classes 1, 2, 5 sont caractérisées par des vents venant du Nord Est alors que les classes 3 et 4 sont caractérisées par les vents venant de l'Ouest.

Sélection de variables en classification



FIG. 5 – rose des vents de chaque classe

La figure 6, montre que les classes 1, 2 et 5 sont essentiellement constituées des données de la saison sèche (de Novembre à Mars), la classe 3 est majoritairement constituée des données du mois de Mai, mois d'intersection entre la saison sèche et la saison de pluie et la classe 4 est majoritairement constituée des données du mois d'Avril. Nous remarquons également que les classes correspondant aux mois de la saison sèche sont caractérisées par des vents venant du nord Est c'est à dire les vent d'harmattan (Figure 6). Les classes 3 et 4 correspondant majoritairement aux mois de début de la saison des pluies se caractérisent par des vents venant majoritairement de l'ouest. Ce qui indique que l'utilisation de 2S-SOM au niveau 2 a permis d'avoir des classes avec une saisonnalité et une direction de vent particulière chacune (Figure 6, 5) et qui

permet de dire qu'à la sortie, la sélection de variables permet d'avoir des résultats cohérents interprétable au point de vue géophysique.



FIG. 6 – Saisonnalité des classes obtenues au niveau 2

# 4 Conclusion

Nous avons proposé une approche de sélection des variables en classification. Les meilleurs résultats de classification obtenus sur les variables pertinentes sélectionnées au niveau 1 vis-à-vis des données étiquetées montre l'intérêt de ce processus de filtrage des données. De plus, l'application de la méthode sur les données météorologiques montre que nous avons proposé une méthode efficace de sélection de variables permettant de fournir en sortie non seulement des variables pertinentes mais aussi un résultat visuel et compréhensible des clusters identifiés.

# Références

Agrawal, R., J. G. D. G. e. P. R. Automatic subspace clustering of high dimensional data for data mining applications. pp. pp. 94–105.

Sélection de variables en classification

- Chen, X. et Y. Ye (2012). A feature group weighting method for subspace clustering of high-dimensional data. *Pattern Recogn*.
- Davies, D. L. et D. W. Bouldin (1979). A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (2), 224–227.
- Gordon, A. (1996). A survey of constrained classification. *Computational Statistics & Data Analysis 21*(1), 17 29.
- Jing, L., M. Ng, et J. Huang (2007). An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. knowledge and data engineering. *IEEE Transactions on 19 (8) 1026 –1041*.
- Kohonen, T. (1998). The self-organizing map. Neurocomputing 21(1-3).
- Kriegel, H.-P., P. Kröger, et A. Zimek (2009). Clustering high-dimensional data : A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans. Knowl. Discov. Data 3, 1:58.
- Lebart, L., A. Morineau, et M. Piron (1997). Statistique exploratoire multidimensionnelle.
- Ouattara, M., N. Niang, F. Badran, et C. Mandin (2014). une méthode de soft-subspace clustering pour données multi-blocs basée sur les cartes topologiques auto-organisées. *Revue des Nouvelles Technologies de l'information (RNTI)*.
- Parsons, L., E. Haque, et H. Liu (2004). Subspace clustering for high dimensional data : a review. *SIGKDD Explor. Newsl.* 6(1), 90–105.

### Annexe

Les indices de comparaison de deux partitions C et C' Nous désignons par :

 $N_{11}$ , le nombre de fois où deux observations sont dans une même classe dans C et dans une classe C' (accords positifs)  $N_{10}$ , le nombre de fois où deux observations sont dans la même classe de Cet dans des classes différentes dans C'.  $N_{01}$ , le nombre de fois où deux observations sont dans la même classe de C et des classes différentes C'  $N_{00}$ , le nombre de fois où deux observations sont dans des classes différentes de C et de C' (accords négatifs)

L'indice de précision indique la probabilité que deux objets soient regroupés dans la partition C s'ils le sont dans la partition C' :

F. Kaly et al.

$$Prcision(C, C') = \frac{N_{11}}{N_{11} + N_{01}}$$

Le coefficient de rappel évalue la probabilité que deux objets soient regroupés dans la partition C s'ils le sont dans la partition C' :

$$Rappel(C, C') = \frac{N_{11}}{N_{11} + N_{10}}$$

La pureté d'une partition s'évalue en quantifiant la cohérence d'une partition par rapport une autre. La manière la plus simple d'évaluer la pureté est de rechercher le label majoritaire de chaque classe et de sommer le nombre d'observations ayant le label majoritaire par classe. La pureté se définit alors simplement par l'expression suivante :

$$Purete(C, C') = 1/N \sum_{k=1}^{K} argmax_{cl}(n_{kl})$$

 $n_{kl}$  est le nombre d'observation dans la classe k de C et dans la classe l de C'.

L'indice de Rand indique la proportion de paires d'observations pour lesquelles deux partitions sont en accord.

$$Rand(C, C') = \frac{N_{11} + N_{00}}{N_{11} + N_{11} + N_{01} + N_{00}}$$

## Summary

We propose a method of feature selection in classification based on selforganized maps SOM. It uses the method of subspace clustering 2S-SOM in two hierarchical steps. The first level provides a system of weight, evaluating the variables and blocks relative contributions to the groups of observations. These weights allow the selection of relevant variables. 2S-SOM is again used on the selected variables to determine the final partition of the observations in the second step. The method is evaluated on simulated and real data. In particular, the application on meteorological data shows that the selection of variables at the level 1 facilitates the geophysical interpretation of the classes obtained at the level 2.