



Université de Toulouse

MASTER 2 GEOMATIQUE

« ScIences Géomatiques en environneMent et Aménagement » (SIGMA)

<http://sigma.univ-toulouse.fr>

MÉMOIRE DE RECHERCHE

***Calcul et cartographie d'indices climatiques sur
l'ensemble de l'Afrique de l'Ouest à l'échelle
locale et régionale***

Fabrice PIERRON

IRD Espace Dev



Maîtres de stage : Dimitri DEFRANCE (Espace Dev) /Jean-Emmanuel PATUREL (HSM)

Tuteur-enseignant : David SHEEREN (DYNAFOR)

Septembre 2018

Résumé

Les études portant sur le changement climatique sont nombreuses en Afrique de l'Ouest. Les expertises, menées par les différents scientifiques, tendent à démontrer que la région devrait connaître une perturbation dans les cycles de mousson. Cette dernière s'avère primordiale en Afrique de l'Ouest, car elle constitue bien souvent le seul apport en eau pour l'agriculture. Celle-ci tend néanmoins à varier de trois manières : Nord-Sud, Est-Ouest et interannuelle. C'est dans ce contexte que le projet AGRICORA (Agriculture et Gestion des Risques Climatiques : Outils et Recherches en Afrique) a vu le jour. Ce dernier a pour objectif de prendre en compte les risques climatiques dans la production agricole des pays de la région.

Afin, d'analyser ces risques, des modèles climatiques sont produits. Ces derniers visent à anticiper et à prévoir les impacts du changement climatique. Mais ils ne sont pas toujours adaptés pour les études d'agronomies et d'hydrologies. Ainsi, pour analyser la qualité de ces modèles, nous allons comparer ces derniers avec des données réelles (stations), issues de la base données Hydrosociences/SIEREM. Cette comparaison doit permettre d'étudier les écarts entre les données créées et les données des stations. Pour appréhender au mieux cette différence, une série d'indices climatiques va être calculés. Ceux-ci doivent permettre de prendre en compte l'ensemble des contextes climatiques de la région.

De plus, pour quantifier ces erreurs, des opérateurs statistiques (RMSE, RRMSE) ont également été réalisés. Ces derniers visent à mesurer les erreurs entre les différents jeux de données, en étudiant les variabilités en fonction des zones et des indicateurs climatiques.

Mots-clés : Afrique de l'Ouest, changements climatiques, mousson, modèles climatiques, stations météorologiques, indicateurs

Abstract

Studies on climate change are numerous in West Africa. Expertise, led by different scientists, tends to show that the region should experience a disturbance in the monsoon cycles. This is essential in West Africa because it is often the only water supply for agriculture. It nevertheless tends to vary in three ways : North-South, East-West and interannual. It is in this context that the AGRICORA project (Agriculture and Climate Risk Management: Tools and Research in Africa) was born. This one aims to take into account the climatic risks in the agricultural production in the countries of the region.

In order to analyze these risks, climate models are produced. These aim to anticipate and predict the impacts of climate change. But they are not always suitable for studies of agronomy and hydrology. Thus, to analyze the quality of these models, we will compare these with real data (weather stations), from the Hydrosociences / SIEREM database. This comparison should make it possible to study the differences between the data created and the data of the stations. To better understand this difference, a series of climate indices will be calculated. These must make it possible to take into account all the climatic contexts of the region.

In addition, to quantify these errors, statistical operators (RMSE, RRMSE) have also been made. These aim to measure the errors between the different data sets, by studying the variabilities according to the zones and the climatic indicators.

Keywords : West Africa, climate change, monsoon, climate models, weather stations, indicators

Sommaire

<i>Résumé</i>	3
<i>Sommaire</i>	4
<i>Remerciements</i>	5
<i>I Introduction</i>	6
<i>II La structure d'accueil</i>	7
<i>III Contexte d'étude</i>	8
1) Le projet Agricora.....	8
2) Vulnérabilités et variabilités climatiques.....	8
3) Délimitation de la zone d'étude.....	12
<i>IV Problématique objectifs et organisation du stage</i>	13
1) Problématique.....	13
2) Objectifs.....	13
3) Gestion du stage.....	14
<i>V Données et méthodes</i>	16
1) Les données disponibles.....	16
2) Outils et Méthodologies.....	25
<i>VI Résultats</i>	37
1) Sorties graphiques.....	37
2) Calcul du RMSE.....	38
3) Étude comparative des indices climatiques à l'aide du RRMSE.....	40
4) Étude de cas intra-maille.....	45
<i>VII Limites et perspectives</i>	50
<i>VIII Conclusion</i>	52
<i>Bilan personnel</i>	53
<i>Table des figures</i>	54
<i>Table des tableaux</i>	55
<i>Table des matières</i>	56
<i>Annexes</i>	58
<i>Bibliographie</i>	65
<i>Webographie</i>	67

Remerciements

Avant d'entamer ce rapport, je tiens à remercier les différentes personnes qui m'ont permis de réaliser ce stage.

Premièrement, je tiens à remercier **Dimitri Defrance**, mon maître de stage, pour m'avoir encadré tout au long de ce travail. Je tiens également à adresser mes remerciements, à mon tuteur-enseignant, **David Sheeren**.

Ensuite, je tiens à remercier l'ensemble des personnes qui m'ont apporté leurs expertises et leurs connaissances thématiques et techniques durant le stage : **Jean-Emmanuel Paturel, Jean-Christophe Desconnets, Benjamin Sultan, Thérèse Libourel** et **Gil Mahé**.

Je souhaite également remercier l'équipe **d'Hydrosciences/Sierem** qui m'a fourni des données précieuses pour cette étude. Mes pensées vont notamment à **Nathalie Rouché**

Je tiens aussi à présenter mes remerciements à mon binôme de stage, **Romain Bouvier**. Merci également à **Benjamin Pilot** pour son aide technique durant le stage.

Enfin, je tiens à remercier l'ensemble de ma famille, mes amis et mes collègues de la Maison de la Télédétection, qui m'ont soutenus et aidés tout au long de ce stage.

I Introduction

L'impact des activités humaines sur le climat est presque inéluctable. Ce dernier devrait entraîner une hausse des événements extrêmes telles que inondations et sécheresses (Hirabayashi and Kane, 2009 ; Brooks and Adger, 2003 ; Burton, Kates, White, 1993). Cette augmentation pourrait avoir des conséquences néfastes pour de nombreux êtres humains. Le secteur de l'agriculture devrait notamment connaître un profond bouleversement. En effet, des baisses de rendements et de productivités sont à prévoir du fait de ces changements. Certaines régions du globe devraient être plus fortement touchées par ces changements. C'est notamment le cas pour l'Afrique de l'Ouest.

En effet, les différentes études portant sur le climat estiment que les températures devraient augmenter de 2 à 6° C, selon les scénarios, d'ici à la fin du siècle (Groupe d'experts Intergouvernemental sur l'évolution du Climat, 2013). Ces changements climatiques pourraient également s'accompagner d'une modification des régimes pluviométriques dans la région. En outre, les cycles de la mousson, de mai à septembre, pourraient être décalés ou réduits (Biasutti, 2013). Ces possibilités de décalages sont primordiaux pour cette région car la mousson constitue bien souvent le seul apport en eau pour le secteur agricole (Sultan, 2011). A ce titre, Sultan *et al* (2015) rappelle que l'agriculture pluviale représente près de 93 % des terres cultivées de la région.

C'est dans ce contexte relativement complexe et inquiétant que le projet Agriculture et Gestion des Risques Climatiques : Outils et Recherches en Afrique (AGRICORA) a vu le jour. Ce dernier vise à « *contribuer à l'élaboration d'outils efficaces de gestion du risque climatique en co-construisant des stratégies innovantes basées sur les résultats de la recherche avec des réseaux de chercheurs et d'acteurs directement impliqués dans l'accompagnement de l'agriculture* ».

Pour évaluer les conditions climatiques, des stations météorologiques sont utilisées. Ces dernières s'avèrent relativement rares ou de mauvaises qualités en Afrique de l'Ouest. Ainsi, on utilise des données de modèles afin d'analyser et d'anticiper le climat sur la région. Mais, celles-ci ne sont pas systématiquement pertinentes pour les études d'impacts, d'agronomies et d'hydrologies. La spatialisation des modèles climatiques reste, bien souvent, trop large et imprécise pour effectuer une analyse à l'échelle locale. Ce constat est particulièrement vrai pour les données de précipitations qui sont fréquemment mal modélisées.

Ainsi, afin d'analyser les sorties des modèles climatiques, nous allons effectuer une comparaison entre des données réelles et des données modélisées. Pour ce faire, nous allons sélectionner et calculer une série d'indices climatiques. Ces derniers doivent permettre d'effectuer une comparaison entre des données réelles (stations) et des données créées (modèles). Afin, de représenter au mieux les disparités climatiques, ces indices doivent avoir des pas de temps, des variables et des seuils différents. Ces derniers doivent permettre par la suite d'évaluer la qualité des modèles climatiques, en s'appuyant sur différentes analyses statistiques.

Cette étude sur les modèles climatiques et leur comparaison avec des données réelles reste néanmoins soumise à une série de questionnement. Comment fournir des données appropriées et fiables ? Quels indicateurs climatiques intégrer afin de répondre aux vulnérabilités futures ? Comment quantifier et qualifier les pertes d'information lorsque l'on passe de l'échelle locale à l'échelle maillée ?

Pour répondre à ces différentes problématiques nous allons nous centrer sur une méthodologie et une série de données climatiques que nous présenterons dans le corps du rapport. Ces dernières

devraient permettre de calculer et d'analyser les sorties des modèles. Enfin, dans une dernière partie nous reviendrons sur les problèmes, les limites et les perspectives de ces différentes analyses.

II La structure d'accueil

1) Institut de Recherche pour le Développement (IRD)

L'Institut de Recherche pour le Développement (IRD) est un organisme public français de recherche qui existe depuis 1939. Ses recherches sont axées sur les relations entre l'homme et son environnement en Afrique, Méditerranée, Amérique latine, Asie et dans l'Outre-Mer tropical français. Les activités de l'IRD ont pour objectifs de contribuer au développement social, économique et culturel des pays du Sud. C'est un Institut qui s'inscrit dans une approche pluridisciplinaire. En 2017, l'IRD se compose de 2 013 agents, dont 833 chercheurs et 1 180 ingénieurs et techniciens, 66 unités de recherche ; 40 représentations à l'étranger et en Outre-Mer. Au sein, de la délégation régionale d'Occitanie, la structure d'accueil de l'IRD et la Maison de la Télédétection, basée à Montpellier.

2) Maison de la Télédétection (MTD)

La Maison de la Télédétection (MTD) a été construite en 1994 à Montpellier-Agropolis. Elle rassemble des équipes appartenant aux organismes de recherche et de formation AgroParisTech, le Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), l'IRD, l'Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (IRSTEA) et le Centre Nationale de Recherche Scientifique (CNRS). Ces équipes travaillent sur la télédétection et plus largement sur l'information spatialisée. Au sein de la MTD, deux unités mixtes de recherche travaillent en collaboration, l'UMR TETIS et l'UMR Espace-Dev.

3) Espace DEV

L'UMR Espace Dev, qui est sous la tutelle de l'IRD, est une unité de recherche qui développe ses activités sur les dynamiques spatiales caractérisant les éco-sociosystèmes. C'est dans le cadre de cet UMR que cette étude a été réalisée. Les objectifs de cette unité visent à établir des définitions d'indicateurs des dynamiques bio-géophysiques : évolutions des sociétés, risques liés aux maladies émergentes, changements et vulnérabilités des territoires aux changements globaux. Afin de répondre à ces problématiques, l'unité met au point des méthodologies en télédétection spatiale et en intégration des connaissances multidisciplinaires. Le site officiel de l'UMR stipule « *L'UMR fait le constat que l'accès aux données spatiales au Sud peut être une clé pour l'amélioration des connaissances et des prises de décision face aux changements globaux.* »

L'organisation fonctionnelle de l'UMR est basée sur trois équipes :

- [OSE](#) (Observation Spatiale de l'Environnement),
- [AIMS](#) (Approche Intégrée des Milieux et de Sociétés)
- [MICADO](#) (Modélisation, Ingénierie des Connaissances et Analyses de DOnnées spatiales).

L'équipe qui est basée sur ce projet est OSE.

III Contexte d'étude

1) Le projet Agricora

Ce stage est financé par le projet Agriculture et Gestion des Risques Climatiques : Outils et Recherches en Afrique (AGRICORA) . Ce dernier est soutenu par le Ministère de l'Europe et des affaires étrangères. Il s'agit d'un Fonds de Solidarité Priorité (FSP) pour une durée de 3 ans, sous la coordination de l'IRD. AGRICORA s'appuie sur des partenariats existants entre les chercheurs français et africains. L'objectif principal du projet est « *de contribuer à l'élaboration d'outils efficaces de gestion du risque climatique en co-construisant des stratégies innovantes basées sur les résultats de la recherche avec des réseaux de chercheurs et d'acteurs directement impliqués dans l'accompagnement de l'agriculture* ».

Trois priorités thématiques sont mises en avant par le projet :

1. **Services climatiques pour l'agriculture : APTE-21** = Applications des prévisions climatiques et des pratiques agricoles dans la traduction des événements pluviométriques extrêmes du 21ème siècle en zones semi-arides inondables.
2. **Intensification écologique : ECO-FERT-CLIM** = Gestion écologique de la fertilité des sols dans les agrosystèmes de polyculture élevage pour une agriculture intelligente face au climat.
3. **Ressources en eau pour l'agriculture : GENERIA** = Gestion des nouveaux risques et des opportunités des terres inondables pour l'agriculture africaine.

L'ambition du projet est d'offrir la possibilité aux acteurs de l'agriculture (organisations paysannes, filières, secteur privé agricole, banque de développement agricole, fournisseurs d'intrants, services agricoles et de météorologie) « *de s'approprier de nouvelles connaissances et de nouveaux outils pour une meilleure prise en compte des risques climatiques dans la gestion des systèmes de production* ». Ce stage est transversal entre les différentes thématiques et vise à une évaluation des données climatiques nécessaires aux différents acteurs.

2) Vulnérabilités et variabilités climatiques

2.1) Le contexte climatique de la zone d'étude

L'ensemble des analyses réalisées dans le cadre d'AGRICORA est centré sur l'Afrique de l'Ouest. Cette région est caractérisée par un climat bi-saisonnier avec une saison sèche et une saison humide :

- La saison des pluies, appelée mousson, se déroule de mai à septembre. C'est durant cette saison, que la très grande majorité des événements pluvieux de la région est recensée. Les mois de mousson en Afrique de l'Ouest, sont un élément essentiel pour les populations africaines. Ces derniers constituent souvent le seul apport en eau pour le secteur agricole (Sultan, 2011). Cet apport favorise le développement agropastoral dans la région. Sultan (2011) rappelle, à ce propos, que l'agriculture pluviale représente près de 93 % des terres cultivées de la région, qui sont le plus souvent non irriguées
- La saison sèche s'établit d'octobre à avril. Cette dernière est caractérisée , comme son nom l'indique, par la très faible propension d'événements pluvieux.

Les conditions climatiques dans la région restent néanmoins hétérogènes. On observe ainsi une plus forte pluviométrie au sud qu'au nord (Figure 1). Plus, on descend vers le sud, plus la saison des pluies est longue et intense. Cette différence nord/sud est liée à la Zone de Convergence Inter-Tropicale (ZCIT). Cette dernière connaît deux latitudes d'équilibre. La première se situe à 5°N durant le mois de mai. Elle caractérise le début de la saison des pluies autour du bassin guinéen. La seconde se localise à 10°N durant les mois de juillet et/ou août. Elle est responsable du début de la saison des pluies en zone soudano-sahélienne et sahélienne (Sultan, 2002). La carte ci-dessous, issue des modèles WFDEI et inspirée des travaux de L'Hôte *et al* (2002), montre la somme annuelle moyenne de précipitations sur 35 ans (1979 – 2013).

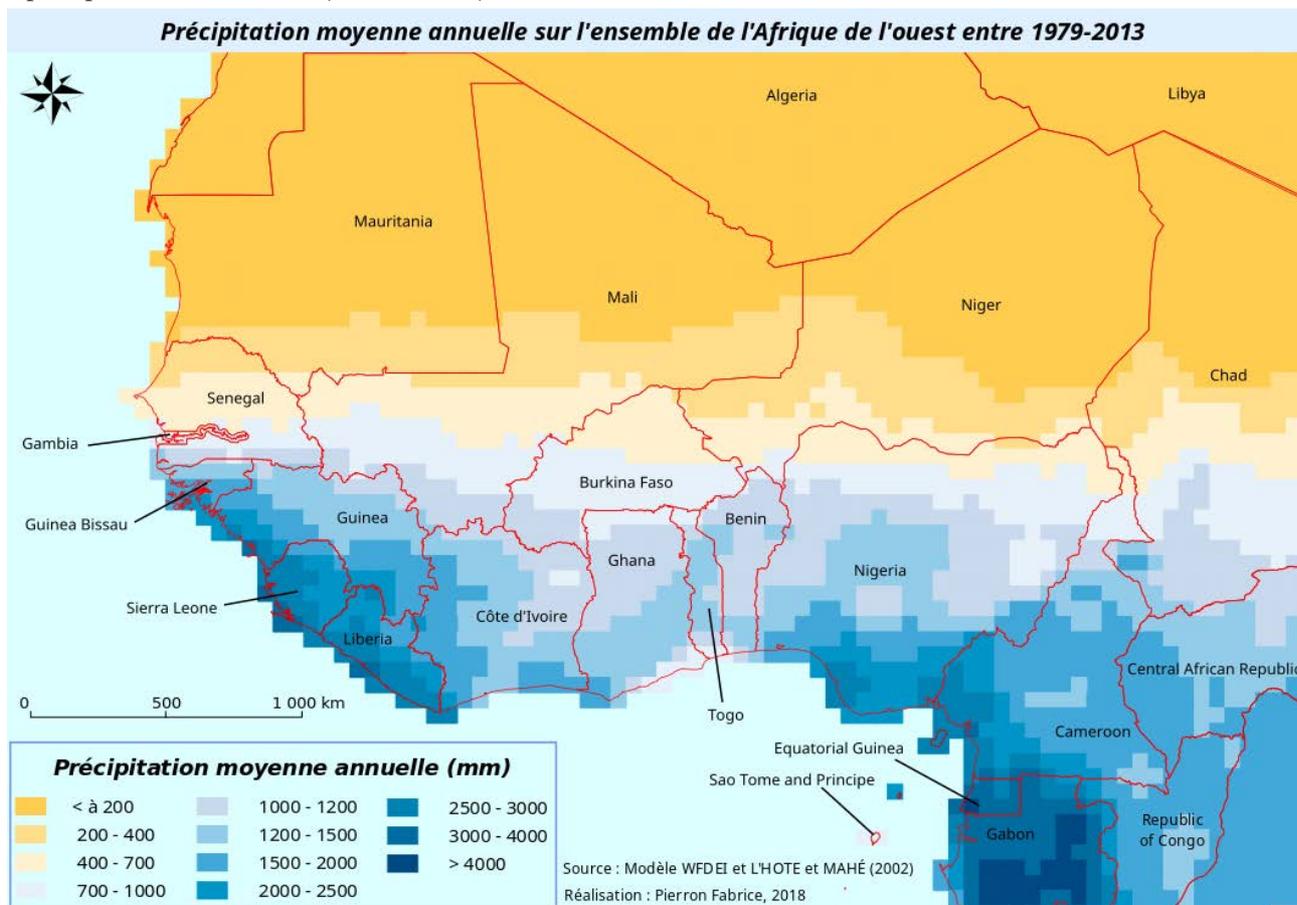


Figure 1 : Cumul interannuel moyen de précipitations (1979-2013) sur l'ensemble de l'Afrique de l'Ouest

Cette répartition des mois de mousson reste néanmoins fluctuante en fonction des années et des localisations. Les travaux menés par Sultan *et al* (2013) ont mis en évidence des variabilités dans les cycles de mousson. On observe en effet trois types de fluctuations dans le processus de mousson :

- Premièrement, on a des événements pluvieux qui diffèrent selon les années. La longueur et la fréquence des périodes de sécheresse ainsi que la durée des précipitations saisonnières affectent la mousson. Ces éléments sont marqués par une grande variabilité interannuelle (Janicot *et al*, 2011). Les très grandes sécheresses du siècle dernier (1909-1913, 1940-1944, 1969-1973, 1983-1985) avaient fait de très nombreuses victimes (club du sahel et de l'Afrique de l'ouest, 2007). Ces périodes de sécheresse affectent également le développement agricole de la région. Kandji *et al* (2006) rappellent, par exemple, que le PIB du Mali et du Niger ont respectivement chuté de 9 % et 18 % suite à la sécheresse de 1984.

- Deuxièmement, on peut observer des cycles intra-saisonniers dans la mousson. Ces cycles se caractérisent par des séquences sèches au sein des mois de mousson. Sultan *et al* (2003) ont montré que les mois de mousson sont marqués par deux modes de variabilités : le premier à courte échelle de temps avec des séquences humides et sèches autour de 15 jours et le second à de plus basses fréquences avec plusieurs jours marqués par une baisse de la pluviométrie au milieu du mois d'août. Sultan *et al* (2003), ont aussi mis en évidence la récurrence de ces deux modes de variabilités. Cette idée est également mise en avant par Janicot *et al* (2011) « *Les précipitations sur l'Afrique de l'Ouest sont très intermittentes dans l'espace et le temps. La saison des pluies est ponctuée de périodes sèches et humides se produisant à différentes échelles temporelles intra-saisonniers* ».
- Troisièmement, on a des différences en fonction des zones géographiques. Ainsi, outre une différenciation Nord/Sud et Est/Ouest, on peut observer des écarts climatiques dans des régions très proches. Balme *et al* (2005) explique notamment dans leurs travaux que des villages voisins connaissent parfois une même année des pluviométries et des dates de démarrages de mousson très différentes.

2.2) Vers un bouleversement climatique

Comme nous avons pu le voir précédemment, les conditions climatiques constituent un enjeu majeur en Afrique de l'Ouest, notamment pour le développement du secteur agricole. Ces conditions restent fluctuantes et tendent à être modifiées du fait du changement climatique. Celui-ci devrait entraîner une modification des phénomènes climatiques dans la région. Cette évolution affecterait notamment fortement la saison des pluies dans les années futures. Les différentes études menées par le Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC) insistent sur l'importance du changement climatique durant cette période de l'année. D'après ces derniers, l'évolution du climat durant les mois de mousson devrait se caractériser par un décalage dans le cycle saisonnier des pluies avec moins de pluies au début de la saison de mousson et plus de pluies sur la fin. En outre, la Zone de Convergence InterTropicale (ZCIT) devrait connaître des changements d'occurrence.

Sur la Figure 2, issue des travaux du GIEC, on peut observer les évolutions de températures et de précipitations en fonction des différents scénarii Representative Concentration Pathway (RCP) sur l'ensemble de l'Afrique. Le scénario RCP 2.6, le plus optimiste, tend à montrer une augmentation des températures de l'ordre de 1 à 2 °C d'ici la fin du 21^e siècle. Le scénario RCP 8.5, le plus pessimiste, montre une hausse de température comprise entre 4 et 6 °C d'ici la fin du siècle.

Les évolutions des précipitations tendent à montrer une variation plus contrastée, avec des régions qui devraient connaître une hausse des précipitations et d'autres qui devraient connaître une baisse. Cette variation s'avère particulièrement perceptible pour le scénario RCP 8.5. Néanmoins, les résultats des scénarii pour les précipitations doivent être atténués à cause des divergences selon les différents modèles climatiques étudiés. En effet, certains modèles ont tendance à montrer une baisse des précipitations sur l'Afrique, tandis que d'autres montrent une tendance à la hausse. De plus, cette différence tend à être variable selon les zones climatiques.

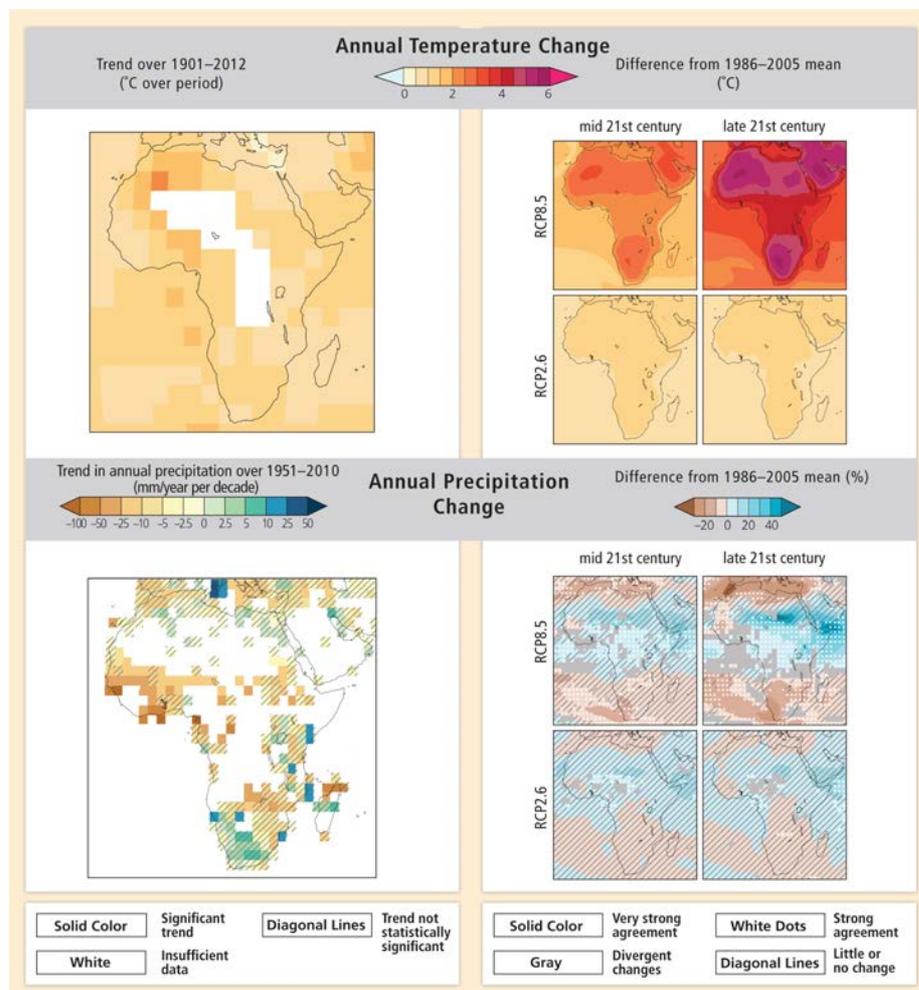


Figure 2 : Projection des évolutions de températures et de précipitations en fonction des scénarii climatiques
 Source : GIEC (2014)

2.3) Les modèles climatiques

Les modèles climatiques sont un moyen d'évaluer les conditions passées et futures du climat. Ils permettent d'effectuer des simulations du climat de quelques années à plusieurs millénaires. Ils ont pour objectifs de prévoir et d'anticiper les impacts liés aux changements climatiques. Le Centre National de Recherches Météorologiques (CNRM) stipule :

« les composantes du système climatique peuvent être représentées par des modèles numériques, que développent des scientifiques et ingénieurs spécialisés. Ces modèles s'appuient sur les lois de la physique et sont régulièrement testés de manière à ce que chacun d'entre eux simule correctement le milieu qu'il représente ».

Ces modèles restent néanmoins des simulations et ne peuvent donc pas représenter les spécificités climatiques à l'échelle locale. Ainsi, la question de la qualité et de la précision de ces derniers demeure présente. En outre, la résolution spatiale utilisée par ces modèles n'est pas toujours adaptée pour les études d'agronomie ou d'hydrologie. De plus, ces derniers varient selon les scénarii et les variables prises en compte dans le cadre des modèles. C'est pour cette raison que de plus en plus

d'outils d'analyses du climat s'appuient sur une combinaison de plusieurs modèles et scénarii climatiques (GIEC, 2013).

3) Délimitation de la zone d'étude

La zone d'étude est localisée sur l'ensemble des pays de l'Afrique de l'Ouest. Mais cette dernière a été peaufinée en fonction de la spatialisation et de la disponibilité des données selon les pays. Ainsi, certains pays localisés dans cette zone ne seront pas analysés dans cette étude. C'est le cas notamment de la Sierra Leone, du Liberia ou de la Guinée-Équatoriale.

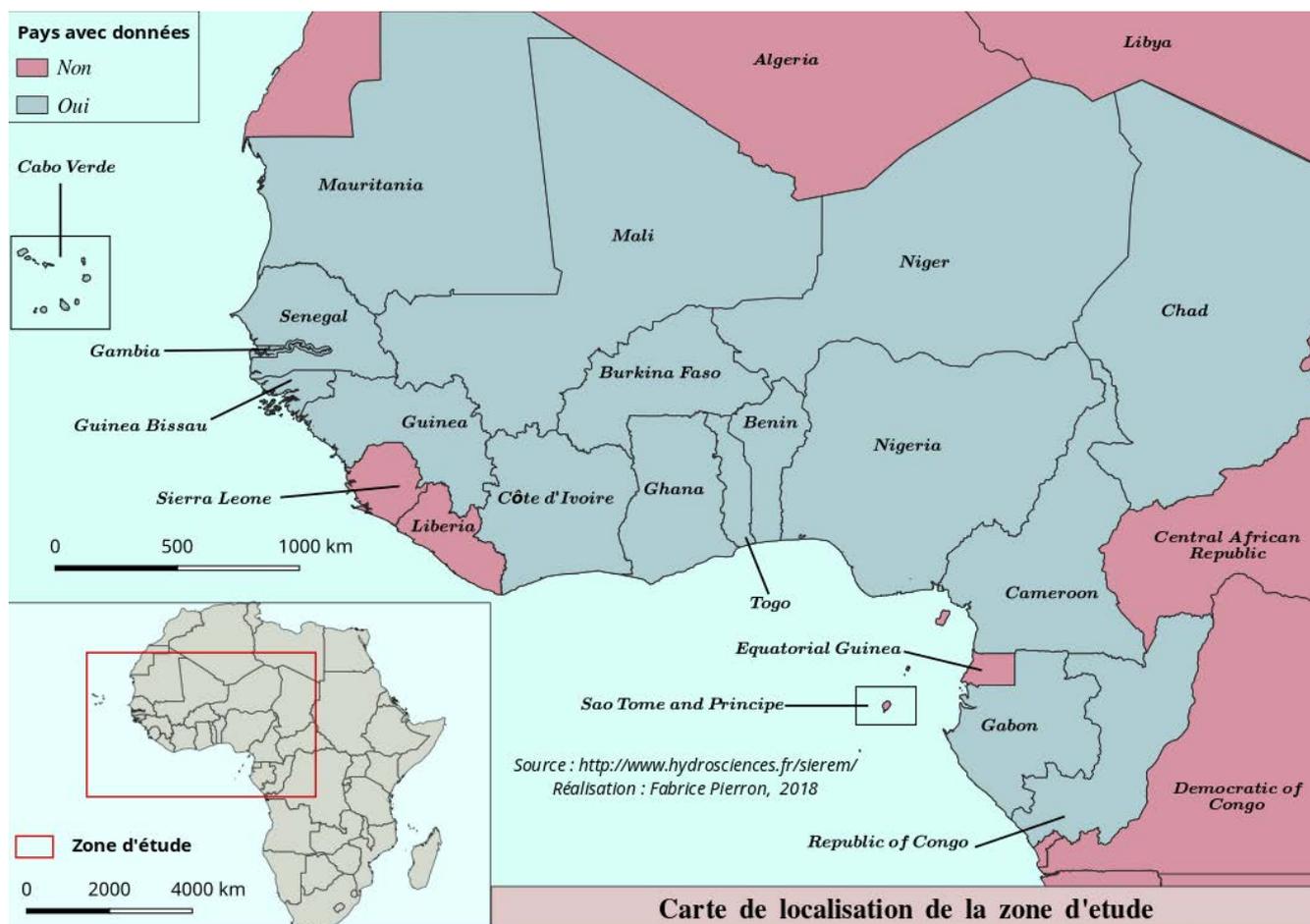


Figure 3 : Carte de localisation de la zone d'étude

La zone d'étude comprend donc 18 pays avec des données disponibles (Mauritanie, Mali, Niger, Tchad, Sénégal, Cap-Vert, Gambie, Guinée, Guinée-Bissau, Côte-d'Ivoire, Ghana, Burkina Faso, Togo, Bénin, Nigeria, Cameroun, Gabon, République du Congo). Cette disponibilité varie en fonction des pays, comme nous le verrons par la suite dans la partie consacrée aux données disponibles.

IV Problématique, objectifs et organisation du stage

1) Problématique

Le sujet de ce stage s'articule autour des thématiques précédemment mentionnées. Il s'agit en outre de s'interroger sur la qualité des données climatiques. Ces dernières étant relativement rares en Afrique de l'Ouest, les scientifiques et les chercheurs utilisent des modèles climatiques combinant des observations et des modélisations, afin d'étudier la variabilité des conditions climatiques. Ces derniers offrent la possibilité d'étudier et d'analyser de vastes régions. Néanmoins, les écarts entre les données modélisées et les données réelles restent importants.

Dans le cadre du projet AGRICORA, plusieurs groupes de chercheurs, au Burkina Faso et au Mali principalement, améliorent la modélisation de certaines méthodes d'adaptation agricole et les évaluent afin de sélectionner les meilleurs pour contrecarrer les effets négatifs liés aux changements climatiques.

Le but du stage est donc de comparer à l'aide de données locales (stations météorologiques) des résultats issus de réanalyse climatique. Afin de déterminer les écarts entre les modèles et les stations météorologiques, une série d'indicateurs climatiques va être calculée. Ces derniers ont pour objectif de quantifier et de qualifier les écarts entre les deux jeux de données. La problématique dans le cadre de ce stage est donc la suivante :

- Comment quantifier et qualifier les écarts de valeurs entre les données réelles et les données maillées, en s'appuyant sur des indices climatiques multiples et variés ?

2) Objectifs

Le stage s'articule autour des éléments thématiques précédemment mentionnés. Il s'agit en outre de fixer différents objectifs afin de répondre aux problématiques précédemment mentionnées. Les principaux objectifs sont récapitulés ci-dessous :

1 = Identifier et automatiser le calcul d'indicateurs climatiques pertinents

2 = Quantifier et qualifier les résultats des modèles climatiques en s'appuyant sur des données réelles (stations météorologiques)

3 = Spatialiser les données disponibles dans un SIG

4 = Évaluer les disparités interannuelles

5 = Cartographier les résultats des indices climatiques à l'échelle locale et/ou régionale

3) Gestion du stage

3.1) Organisation

Ce stage s'est déroulé à la Maison de la Télédétection (MTD) à Montpellier. Il a duré 6 mois du 13 mars au 13 septembre 2018. L'encadrant durant ce stage a été Monsieur [Dimitri Defrance](#) (Post-Doctorant, IRD/EspaceDev) et le tuteur-enseignant a été Monsieur [David Sheeren](#) (Maître de Conférences, UMR DYNAFOR). Monsieur [Jean-Emmanuel Paturel](#) (Directeur de recherche, HSM) Monsieur [Benjamin Sultan](#) (Directeur de recherche, IRD/EspaceDev) et Monsieur [Gil Mahé](#) (Directeur de recherche, HSM) ont apporté leurs expertises sur l'agronomie, le climat et l'hydrologie. Enfin, Monsieur [Jean-Christophe Desconnets](#) (Ingénieur de recherche, IRD/EspaceDev) a apporté ses compétences en termes de gestion et de conduite de projet. Le stage s'est déroulé en collaboration active avec [Romain Bouvier](#) (Master 2, SIGMA).

Le développement de ce projet a été réalisé à l'aide d'un ordinateur fourni par l'IRD. Ce dernier utilise un système d'exploitation Linux. Les outils de développement utilisés dans le cadre de ce stage ont été Python, QGIS. Une phase de pré-développement a également été envisagée sur R.

3.2) Prévisionnel

Le stage s'est déroulé en quatre grandes étapes de travail réparties au cours des six mois. L'ensemble des étapes et des phases du stage est développé dans le diagramme de Gantt (Figure 4) :

Étape 1 : Immersion

Découverte et appropriation du sujet avec une période de recherche bibliographique durant près d'un mois. Cette phase a aussi été l'occasion de prendre en main les nouveaux outils de travail (Système d'exploitation Linux, format multidimensionnel, modèles climatiques ...).

Étape 2 : Pré-traitement

Phase de collection, de nettoyage et de filtrage des données. Cette phase correspond aussi à la période durant laquelle les indices climatiques ont été choisis afin de répondre au mieux aux différentes problématiques.

Étape 3 : Traitement

C'est la phase la plus longue et la plus technique du projet. C'est durant cette étape, que les indices climatiques ont été calculés pour les stations et les modèles. Cette phase prend également en compte l'ensemble des étapes qui ont été nécessaires afin d'automatiser et de comparer le calcul des indices climatiques. Cette période correspond également à la réalisation des différentes cartes présentées dans ce rapport.

Étape 4 : Rendu

C'est la dernière phase de ce projet. Elle correspond à la période de rédaction et de préparation des rendus finaux. C'est la synthèse qui permet de mettre en avant le travail réalisé durant ce stage.

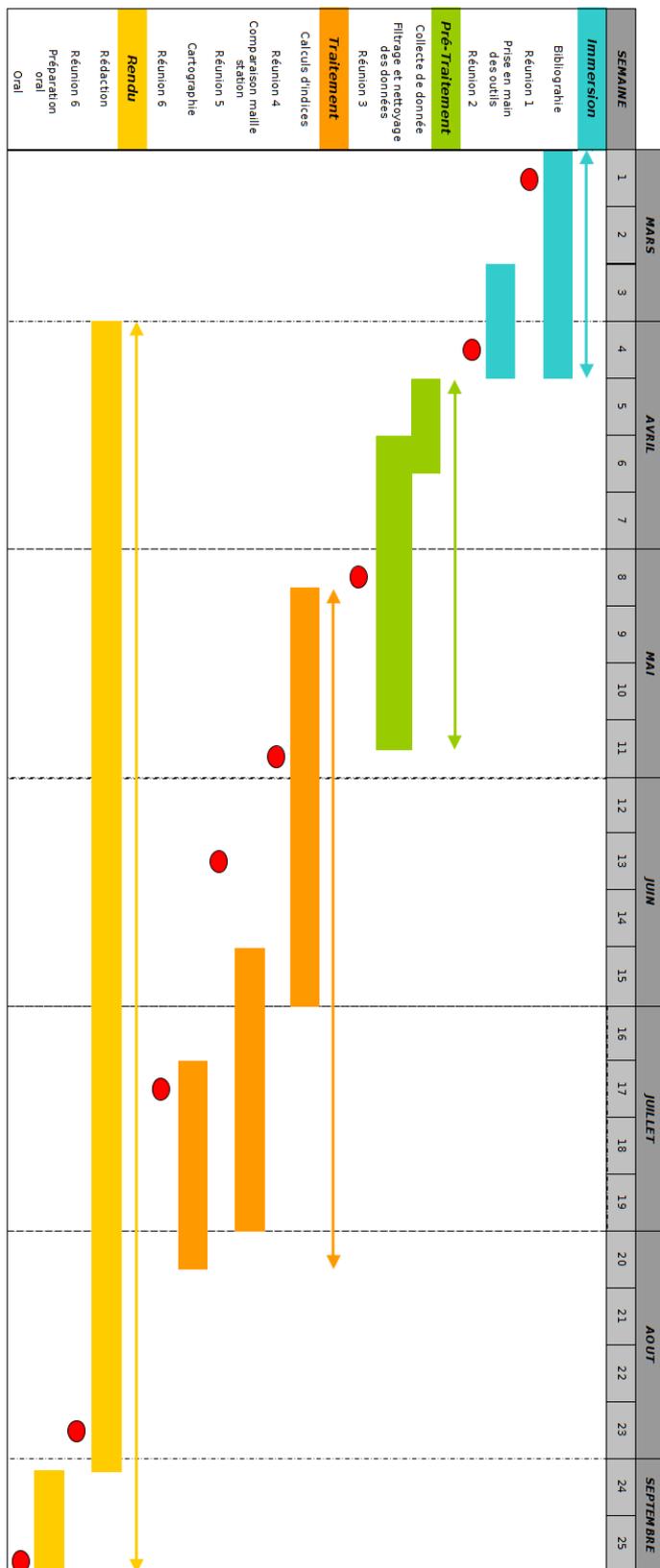


Figure 4: Diagramme de gantt

V Données et méthodes

1) Les données disponibles

1.1) Les données fournies par les études climatiques

1.1.1) Le projet WATCH

Les données des études climatiques que nous allons analyser dans le cadre de ce projet sont issues du projet WATCH (Water and Global Change). Le projet WATCH « est un projet financé par l'Union européenne pour améliorer notre compréhension du cycle de l'eau terrestre. Il réunit des scientifiques de 25 institutions de recherche européennes (ainsi que d'autres d'Amérique et du Japon) de nombreuses disciplines (hydrologie, climat, ressources en eau, télédétection, etc.) pour atteindre cet objectif commun. » (WATCH- Technical Report No.56, 2011)

Ce dernier regroupe des communautés hydrologiques, des ressources en eau et du climat afin d'analyser et d'étudier les composantes des cycles hydrologiques mondiaux actuels et futurs. La conduite de ce projet est exposée autour de neuf objectifs principaux (Annexe 1). Le but de ce projet vise à établir des modèles climatiques permettant d'analyser et d'étudier l'impact des évolutions climatiques actuelles et futures.

Afin, d'effectuer ces études de climat, le projet WATCH a mis en place la méthode WFDEI (WATCH Forcing Data ERA Interim). Cette dernière est basée sur les méthodes utilisées par le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT en anglais *European Centre for Medium-Range Weather Forecasts* ECMWF) qui étudie les données issues des travaux réalisés par le CRU (Climatic Research Unit) . Le CRU fournit un ensemble de données de séries chronologiques correspondant à des variations mensuelles du climat. Au total, la base de données du CRU comprend plus de 4 000 stations couvrant l'ensemble du globe. Les données du CRU reposent sur une grille de résolution de 0,5° x 0,5°.

La méthodologie WFDEI se base sur cette même grille d'analyse. Elle a permis un traitement identique pour tous les points de la grille terrestre (Weedon et al, 2014). Afin de réduire les biais, le WFDEI s'appuie sur différents indicateurs climatiques. Le tableau 1, issu des travaux de Weedon et al (2011) montre une série de variables obtenues à l'aide de la méthodologie WFDEI.

Tableau 1 : Exemple de variables climatiques obtenues à l'aide de la méthodologie WFDEI
Source: Weedon et al (2011)

Variable météorologique
La température à 2 mètres
Le taux de précipitation
L'humidité relative à 2 mètres
Le taux de neige
La vitesse du vent à 10 mètres

1.1.2) Le format NetCDF

Le projet WATCH a choisi d'utiliser le format Network Common Data Form (NetCDF). Ce dernier a été développé par Unidata à partir de 1988. Ce format est très utilisé en climatologie et en océanographie. Les fichiers NetCDF sont composés de trois éléments : les dimensions, les variables et les attributs. Ces derniers sont combinés afin d'obtenir les données finales.

Le NetCDF est un format multidimensionnel contenant trois ou quatre dimensions (latitude, longitude, temps et profondeur). Dans notre cas, les données WFDEI ne comprennent pas d'éléments relatifs à la profondeur. La donnée est donc composée de la latitude, de la longitude et du temps. Chaque fichier est composé de plusieurs variables qui représentent un ensemble de valeurs de même type (Figure 5). Les variables sont utilisées afin de stocker les données relatives aux différentes dimensions du fichier NetCDF. Ces dernières sont définies par un nom et un type de données. Les attributs permettent de stocker les métadonnées. Ceux-ci permettent de fournir des informations sur l'ensemble du fichier NetCDF (Figure 5). Les données correspondent à l'ensemble des valeurs de chacune des variables. Dans l'exemple ci-joint (Figure 5), on peut constater que la variable de précipitation (pr) est définie par un temps (time), une latitude (lat) et une longitude (lon).

Afin, de visualiser la structure des données d'un NetCDF nous avons employé l'outil ncdump (Figure 5). Ce dernier est utilisé directement en ligne de commande. Sur la Figure 5, on peut observer un exemple de visualisation des données de latitude à l'aide de cet outil (`ncdump -v lat NetCDF`).

```

master@ftsylvq1 ~/Bureau/WFDEI $ ncdump -v lat pr_daily_WFDEI_full.nc_africa.nc Visualisation du fichier NetCDF à l'aide de ncdump. Ici les
netcdf pr_daily_WFDEI_full.nc_africa {
dimensions:
  lon = 150 ;
  lat = 160 ;
  time = UNLIMITED ; // (12784 currently)
Variables:
  float lon(lon) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
  float lat(lat) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
  double time(time) ;
    time:standard_name = "time" ;
    time:long_name = "time" ;
    time:units = "days since 1860-01-01 00:00:00" ;
    time:calendar = "standard" ;
  float pr(time, lat, lon) ;
    pr:standard_name = "precipitation" ;
    pr:long_name = "Average of Mean rainfall rate over the previous 3 hours" ;
    pr:units = "kg/m2s" ;
    pr:FillValue = 1.e+20f ;
    pr:missing_value = 1.e+20f ;
    pr:actual_min = 0. ;
    pr:actual_max = 0.00371126846830308 ;

// global attributes:
  :CDI = "Climate Data Interface version 1.6.8 (http://mpimet.mpg.de/cdi)" ;
  :Conventions = "CF-1.0" ;
  :history = "Mon Oct 10 17:51:13 2016: cdo -O sellonlatbox,-20.0,55.0,-40.0,40.0 pr_daily_WFDEI_full.nc pr_daily_WFDEI_full.nc_africa.nc"
  :Title = "WATCH Forcing Data methodology applied \n",
    "to ERA-Interim data" ;
  :Note1 = " Daily averages of 3-hourly WFDEI file" ;
  :Note2 = " EI data regridded to half degree regular lat-lon" ;
  :Note3 = " Genuine land points from CRU grid only" ;
  :Note4 = "Rainf EIHD bias-corrected using CRU \n",
    "TS3.1 wet days & GPCCv5 precip \n",
    "totals, rain gauge correction and \n",
    "Fair threshold fixed" ;
  :Info1 = "www.ecmwf.int/products/data/archive \n",
    "/descriptions/ei/index.html" ;
  :Info2 = "http://badc.nerc.ac.uk/browse/badc/cru \n",
    "/data/cru_ts_3.10" ;
  :Methodology = "Weedon et al 2011 J.Hydromet. 12, 823-848 \n",
    "doi:10.1175/2011JHM1369.1" ;
  :Institution = "Met Office, JCHMR, Wallingford, UK" ;
  :History = "created July 2012" ;
  :Contact1 = "smgomes@fc.ul.pt" ;
  :Contact2 = "graham.weedon@metoffice.gov.uk" ;
  :CDO = "Climate Data Operators version 1.6.0 (http://mpimet.mpg.de/cdo)" ;

data:
  lat = -39.75, -39.25, -38.75, -38.25, -37.75, -37.25, -36.75, -36.25,
  -35.75, -35.25, -34.75, -34.25, -33.75, -33.25, -32.75, -32.25, -31.75,
  -31.25, -30.75, -30.25, -29.75, -29.25, -28.75, -28.25, -27.75, -27.25,
  -26.75, -26.25, -25.75, -25.25, -24.75, -24.25, -23.75, -23.25, -22.75,
  -22.25, -21.75, -21.25, -20.75, -20.25, -19.75, -19.25, -18.75, -18.25,
  -17.75, -17.25, -16.75, -16.25, -15.75, -15.25, -14.75, -14.25, -13.75,
  -13.25, -12.75, -12.25, -11.75, -11.25, -10.75, -10.25, -9.75, -9.25,
  -8.75, -8.25, -7.75, -7.25, -6.75, -6.25, -5.75, -5.25, -4.75, -4.25,
  -3.75, -3.25, -2.75, -2.25, -1.75, -1.25, -0.75, -0.25, 0.25, 0.75, 1.25,
  1.75, 2.25, 2.75, 3.25, 3.75, 4.25, 4.75, 5.25, 5.75, 6.25, 6.75, 7.25,
  7.75, 8.25, 8.75, 9.25, 9.75, 10.25, 10.75, 11.25, 11.75, 12.25, 12.75,

```

Les Dimensions du NetCDF. Dans notre cas, longitude (lon), latitude (lat), temps (time)

Les Variables du fichier NetCDF (lon, lat, time). Dans ce cas, on a un exemple avec la variable précipitation (pr). Une donnée de pr est défini par un temps (time), une latitude (lat) et une longitude (lon). On observe également les informations auxiliaires pour chacune des variables

Les attributs globaux correspondent aux métadonnées valables pour l'ensemble du fichier NetCDF. L'historique des traitements effectué sur le fichier est également observable

Exemple des données de latitudes pour le modèle WFDEI

Figure 5 : Exemple de fichier NetCDF pour les données de précipitations WFDEI

1.1.3) Description des données WFDEI

Les données maillées utilisées dans le cadre de cette étude sont issues de WFDEI. Ce dernier a une résolution spatiale de $0,5^\circ \times 0,5^\circ$. Cela signifie que la distance entre les longitudes et les latitudes est de 50 km. Une maille WFDEI représente, dans la majorité des cas, une surface de 2 500 km² (50 km x 50 km ou $0,5^\circ * 0,5^\circ$). L'annexe 2 montre la répartition des mailles WFDEI sur l'ensemble de la zone d'étude.

Les données WFDEI, nous ont permis d'obtenir les valeurs estimées pour six variables climatiques sur une période de près de 35 ans. Chaque variable correspond à un fichier NetCDF. Ces données réparties sur l'ensemble du monde, ont un pas de temps journalier. Ce qui signifie que chaque maille est définie par une longitude et une latitude à une date donnée. Le tableau 2 ci-dessous récapitule les différentes variables potentiellement utilisables pour le calcul des indices climatiques.

Tableau 2 : Variables au format NetCDF potentiellement exploitables pour le calcul des indices climatiques

Variables	Date de début	Date de fin	Nombre de jours
Précipitations (pr)	01/01/1979	31/12/2013	12784
Températures (tas)	01/01/1979	31/12/2014	13149
Températures minimales (tasmin)	01/01/1979	31/12/2014	13149
Températures maximales (tasmax)	01/01/1979	31/12/2014	13149
Radiations (rsds)	01/01/1979	31/12/2014	13149
Vents (wind)	01/01/1979	31/12/2014	13149

Les données météorologiques étant principalement centrées autour des précipitations et des températures, nous avons choisi de nous axer sur les variables de précipitations (pr), de températures (tas), de températures minimales (tasmin) et de températures maximales (tasmax), pour calculer les indices climatiques.

1.2) Données issues de stations météorologiques

1.2.1) Répartition spatiale de l'ensemble des stations potentiellement exploitables

Les stations météorologiques sont issues de la base de données Hydrosicences/SIEREM (Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation). Le lien de la base de données est le suivant : <http://www.hydrosciences.fr/sierem/>. Cette base possède l'avantage d'avoir une grande quantité de données potentiellement exploitables pour les calculs d'indices climatiques (Figure 6). Cette dernière étant principalement centrée autour des précipitations, nous avons choisi de privilégier les stations avec ce type de données.

La base de données présente également l'avantage de posséder une majorité de données avec un pas de temps journalier. Cela permet d'effectuer des analyses sur plusieurs pas de temps (journalier,

mensuelle, annuelle). De plus, les stations ont des mesures qui s'étalent sur une grande période (1850-2012).

La figure ci-dessous montre la répartition spatiale des stations météorologiques potentielles. Une cartographie sous forme de cluster a été réalisée sur QGIS (Annexe 3). Cette dernière permet de faciliter la lecture en regroupant les stations localisées dans une même zone. Néanmoins, nous verrons par la suite qu'une partie des données n'a pas été exploitée dans le cadre de cette étude. Celles-ci ne répondant pas toujours aux critères de filtrages et de nettoyages de données imposés dans le cadre de cette étude.

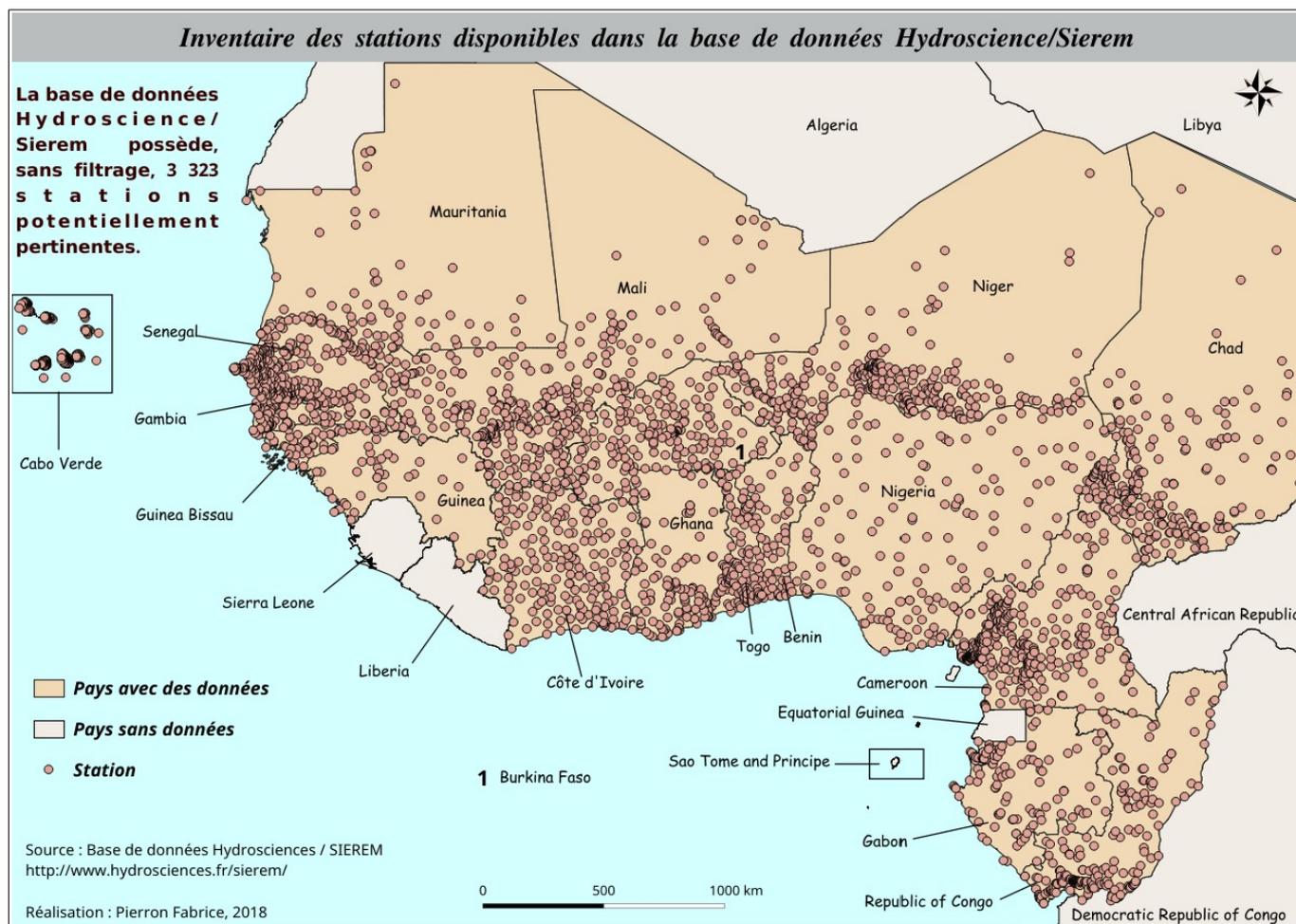


Figure 6 : Inventaire des stations potentiellement disponibles dans la base de données Hydrosiences/Sierem

1.2.2) Répartition spatiale des stations potentiels par pays et par zones climatiques

Au total et sans filtrage de la donnée, la base Hydrosiences/Sierem possède près de 3 323 stations potentiellement pertinentes pour notre étude. Cette grande quantité de données disponibles va nous offrir la possibilité d'effectuer une comparaison entre des données réelles et les données issues de WFDEI. Ces stations restent néanmoins réparties de manière hétérogène. La densité de stations par pays montre bien cette grande diversité (Annexe 4). On peut ainsi observer que certains pays comme le Cameroun, le Cap-Vert ou le Sénégal possèdent un grand nombre de stations. A l'inverse, les pays

localisés autour de la bande sahélienne (Mauritanie, Mali, Niger, Tchad) possèdent une plus faible densité de stations tous les "N" km².

Cette carte ne prenant pas en compte les disparités de densité par pays, nous avons fait le choix d'analyser la répartition des stations par niveau administratif de zone deux (Figure 7). Cette dernière offre la possibilité d'effectuer une étude de la répartition des stations à une échelle plus fine. Ainsi, on peut observer avec plus de justesse la spatialisation des stations pour des pays comme la Mauritanie, le Mali, le Niger ou le Tchad. Pour ces quatre pays, on constate une plus forte concentration des stations dans le sud que dans le nord.

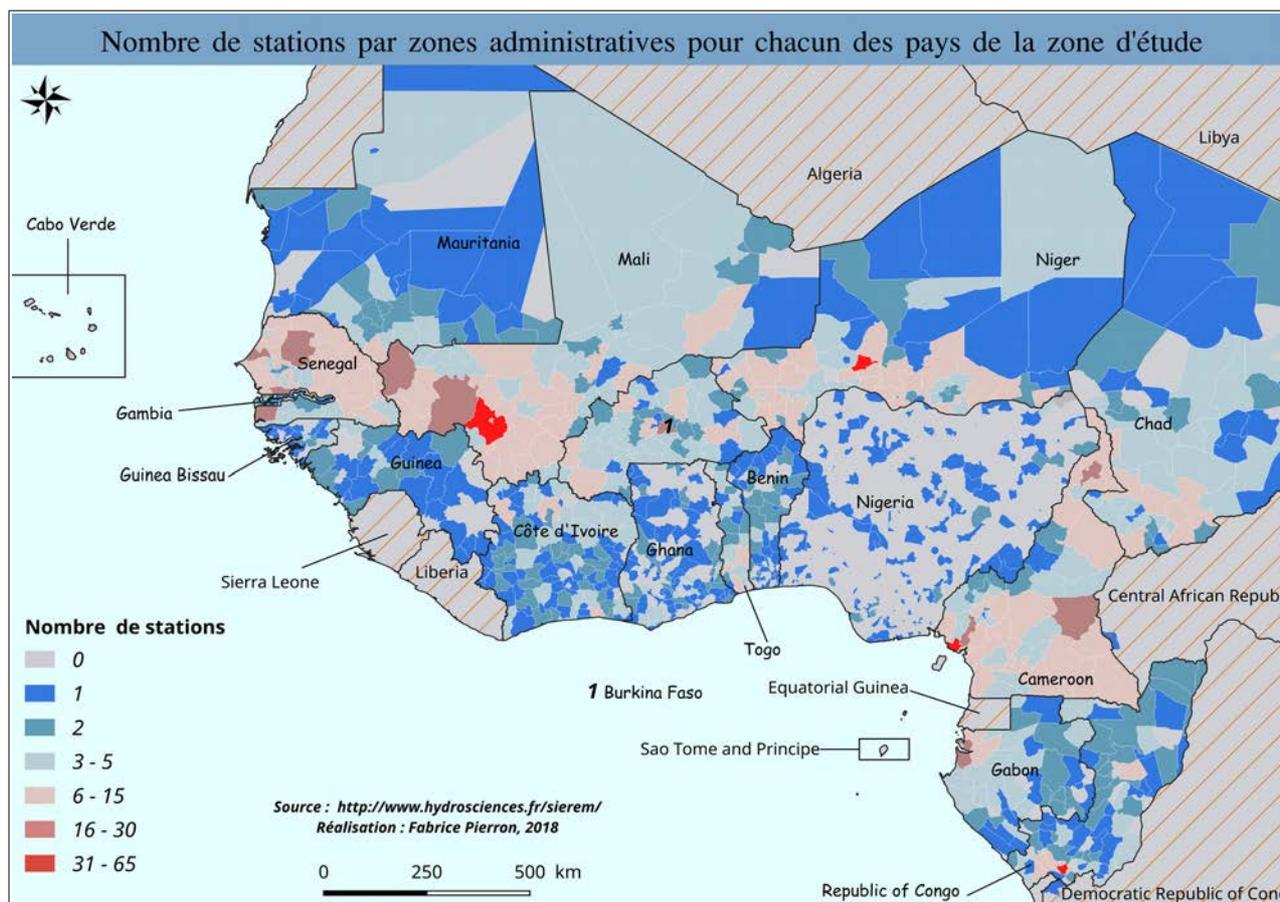


Figure 7 : Nombre de stations par zones administratives de niveau 2 pour chacun des pays de la zone d'étude

Une troisième alternative de représentation de la spatialisation de la donnée a aussi été réalisée (Figure 8). Cette dernière correspond à la densité de stations en fonction des niveaux de précipitations. Ce type de cartographie offre la possibilité d'effectuer une analyse sur une zone climatique spécifique. Ainsi, certaines zones peuvent être étudiées de manière préférentielle. Cette cartographie peut s'avérer pertinente pour les membres du projet AGRICORA, essentiellement focalisés sur les données du Burkina Faso et du Mali.

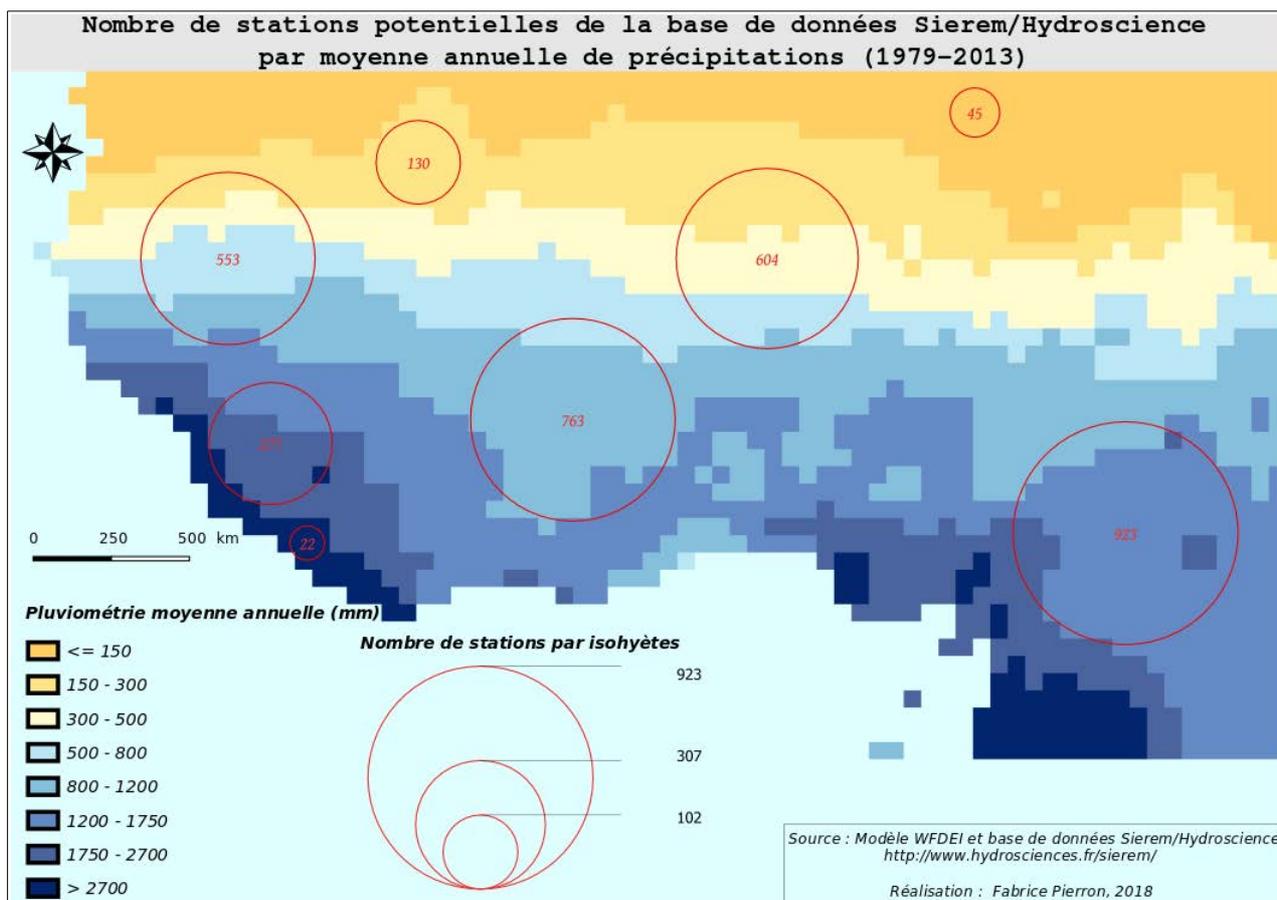


Figure 8 : Répartition des stations potentielles par zones climatiques

1.2.3) Filtrage des stations météorologiques

Comme nous avons pu le voir précédemment, la base de données Hydrosciences/Sierem possède 3 323 stations potentiellement exploitables pour cette étude. Mais, les données de stations n’ont été fournies que pour 243 stations. Les calculs d’indices climatiques ont donc été effectués pour ces stations. La méthodologie réalisée dans le cadre de cette étude reste néanmoins extrapolable pour l’ensemble des données de la base.

Les 243 stations fournies par la base de données Hydrosciences/Sierem ont été filtrées et nettoyées afin d’automatiser l’ensemble des calculs d’indices climatiques (Annexe 5). Un filtre correspondant aux stations avec moins de 30 % de données manquantes à partir de 1950 (date de début des modèles) a été choisi. Ce seuil a été sélectionné après une série de tests (20 %, 30 %, 50 % et 70 %). Ce dernier nous a offert la possibilité d’avoir un maximum de compromis entre les données disponibles et l’impact des données manquantes. Cette sélection a, en outre, permis de limiter l’influence des stations avec un surplus de données manquantes. La Figure 9 montre la répartition spatiale des données fournies par la base de données Hydrosciences/Sierem, soit 243 stations. Cette carte permet également de mettre en avant la qualité des stations fournies en représentant la part de données manquantes pour chacune des stations. Les stations avec une symbologie en rouge ne seront pas prises en compte pour le calcul des indices climatiques. Elles correspondent aux stations avec plus de 30 % de données manquantes.

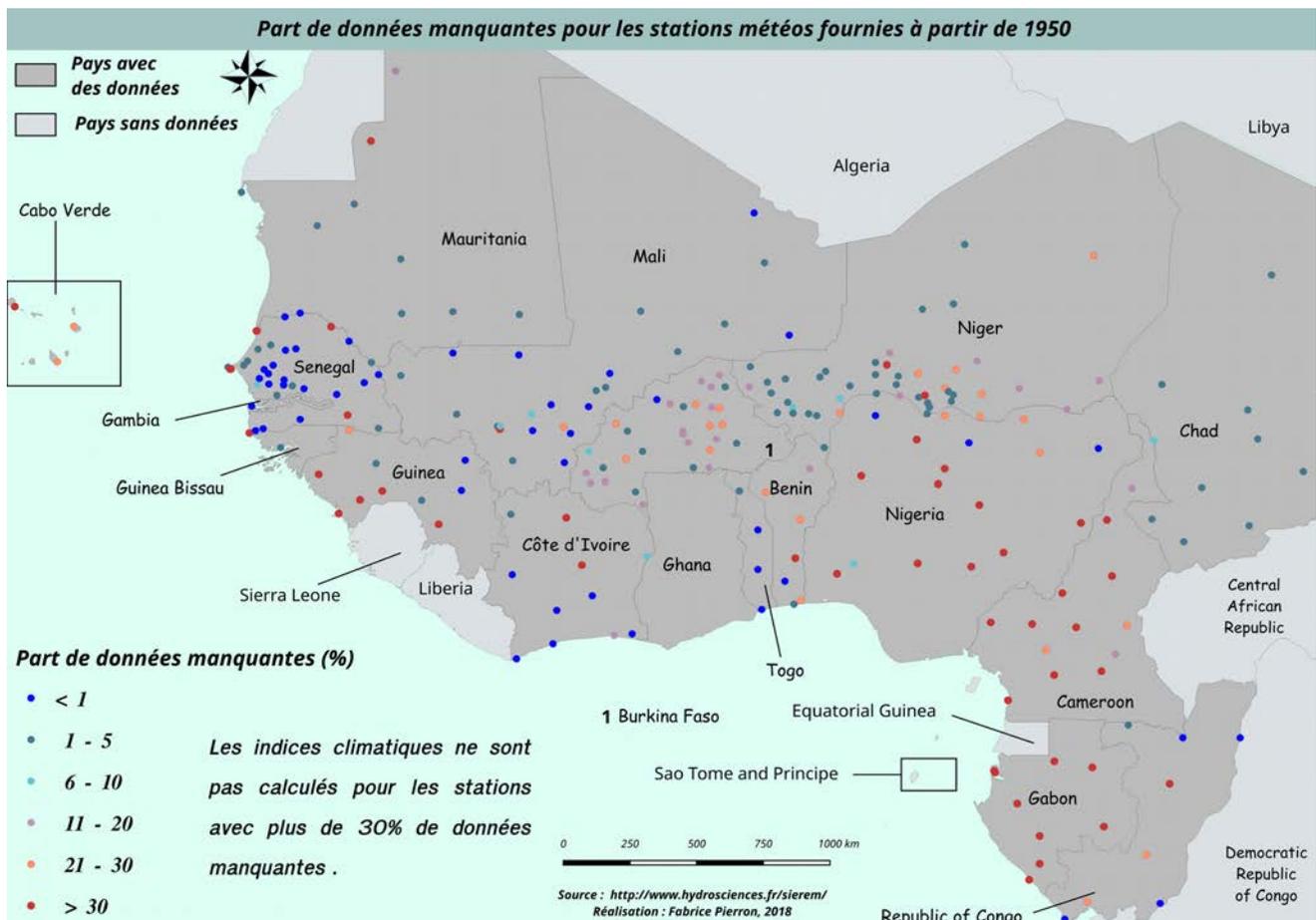


Figure 9 : Part de données manquantes pour les stations fournies par la base de données Hydrosciences/Sierem

On peut alors observer que certains pays comme le Nigeria, le Cameroun ou le Gabon ont une forte concentration de stations avec plus de 30 % de données manquantes. Ces pays possèdent respectivement 58,82 %, 69,23 % et 100 % de stations avec des parts de données manquantes supérieures à 30 %. À l'inverse, la qualité des données est bien meilleure pour des pays comme le Mali, le Togo, le Niger, le Burkina Faso ou le Tchad, qui ont une part de données manquantes supérieures à 30 %, de respectivement 4,17 %, 0,00 %, 4,35 %, 0,00 %, 0,00 % . Ces différentes informations sont récapitulées dans le tableau 3. Ce dernier montre également les données manquantes en fonction des différents seuils. Ainsi, on peut observer que 21,81 % des données fournies, ce qui représente 53 stations, ont plus de 30 % de données manquantes. Celles-ci ne seront pas prises en compte pour le calcul des indices climatiques. Ces derniers seront donc effectués pour les 190 stations restantes (243 – 53 = 190 stations). L'annexe 6 récapitule le nombre de stations, par pays, avec des indices climatiques calculés.

Tableau 3 : Part de données manquantes par pays et par seuils pour les 243 stations fournies

Pays	Stations par pays *	Part de stations par seuils de données manquantes					
		0 %	1 % - 5 %	6 % et 10 %	11 % et 20 %	20 % - 30 %	> 30 %
Bénin	8	12,50%	12,50%	0,00%	12,50%	37,50%	25,00%
Burkina Faso	31	3,23%	25,81%	3,23%	45,16%	22,58%	0,00%
Cameroun	13	0,00%	0,00%	0,00%	15,38%	15,38%	69,23%
Cap-Vert	3	0,00%	0,00%	0,00%	0,00%	33,33%	66,67%
Congo-Brazzaville	8	37,50%	0,00%	0,00%	0,00%	25,00%	37,50%
Cote-d'Ivoire	11	54,55%	9,09%	9,09%	9,09%	0,00%	18,18%
Gabon	9	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
Gambie	1	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Ghana	1	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Guinée	10	20,00%	20,00%	0,00%	0,00%	10,00%	50,00%
Guinée-Bissau	1	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%
Mali	24	41,67%	41,67%	8,33%	0,00%	4,17%	4,17%
Mauritanie	8	0,00%	87,50%	0,00%	0,00%	0,00%	12,50%
Niger	46	0,00%	60,87%	4,35%	15,22%	15,22%	4,35%
Nigeria	17	17,65%	0,00%	5,88%	0,00%	17,65%	58,82%
Sénégal	37	51,35%	24,32%	2,70%	2,70%	0,00%	18,92%
Tchad	11	9,09%	81,82%	9,09%	0,00%	0,00%	0,00%
Togo	4	75,00%	25,00%	0,00%	0,00%	0,00%	0,00%
Total	243	20,99%	31,69%	3,70%	10,70%	11,11%	21,81%
Stations par seuil	190**	51	77	9	26	27	53

* Le nombre de stations ne prend pas en compte le filtrage effectué sur les stations

** Nombre de stations avec moins de 30% de données manquantes (avec filtrage)

1.3) Données des stations et maille WFDEI

Une fois la cartographie effectuée pour les données Hydrosociences/Sierem, on peut analyser la répartition des stations selon les mailles. Pour ce faire, on va calculer et cartographier le nombre de stations par maille, à l'aide de QGIS (Figure 10). Cette carte nous offre la possibilité d'effectuer une analyse maille par maille. Elle permet également de montrer les mailles possédant plusieurs stations météorologiques. Ces dernières sont particulièrement intéressantes, car elles offrent la possibilité de réaliser une comparaison pour les stations localisées dans une même maille. Ainsi, on peut observer

une première différence entre les valeurs d'une maille et de plusieurs stations localisées en son sein. Outre, une spatialisation des données, cette carte permet d'apporter deux nouvelles informations concernant les stations.

Premièrement, on peut observer, le nombre de stations par maille. Cette répartition des stations en fonction des mailles va permettre d'effectuer une comparaison entre les données réelles (stations météo) et les données issues des modèles, pour chacun des indices climatiques calculés. Pour les stations avec des indices climatiques, on peut constater que 162 mailles possèdent 1 station, 7 mailles contiennent 2 stations, 2 mailles contiennent 3 stations et 2 mailles contiennent 4 stations. Le cas des mailles avec plusieurs stations permet d'évaluer si la variabilité intra-maille perdure avec WFDEI

Deuxièmement, la Figure 10 permet d'apprécier le nombre d'années communes entre les modèles climatiques et les stations fournies et filtrées. Cette information offre la possibilité d'effectuer une analyse plus poussée de la qualité des données disponibles. En effet, plus la station a des années en commun avec les modèles climatiques plus la comparaison entre les deux types de données sera pertinente. À l'inverse, des stations avec peu d'années communes ne permettront pas d'analyser suffisamment bien les différences entre les stations et les mailles. Les années communes entre la station et WFDEI varient fortement en fonction des stations et des pays. Ainsi, on peut observer que l'ensemble des stations de la République du Congo ont moins de 10 ans en commun avec WFDEI. Le nombre d'années communes pour les stations du Nigeria du Tchad et du Cameroun est également relativement faible. À l'opposé, des pays comme le Togo, le Bénin, le Sénégal, le Mali ou le Burkina Faso possèdent une majorité de stations avec un grand nombre d'années communes avec WFDEI.

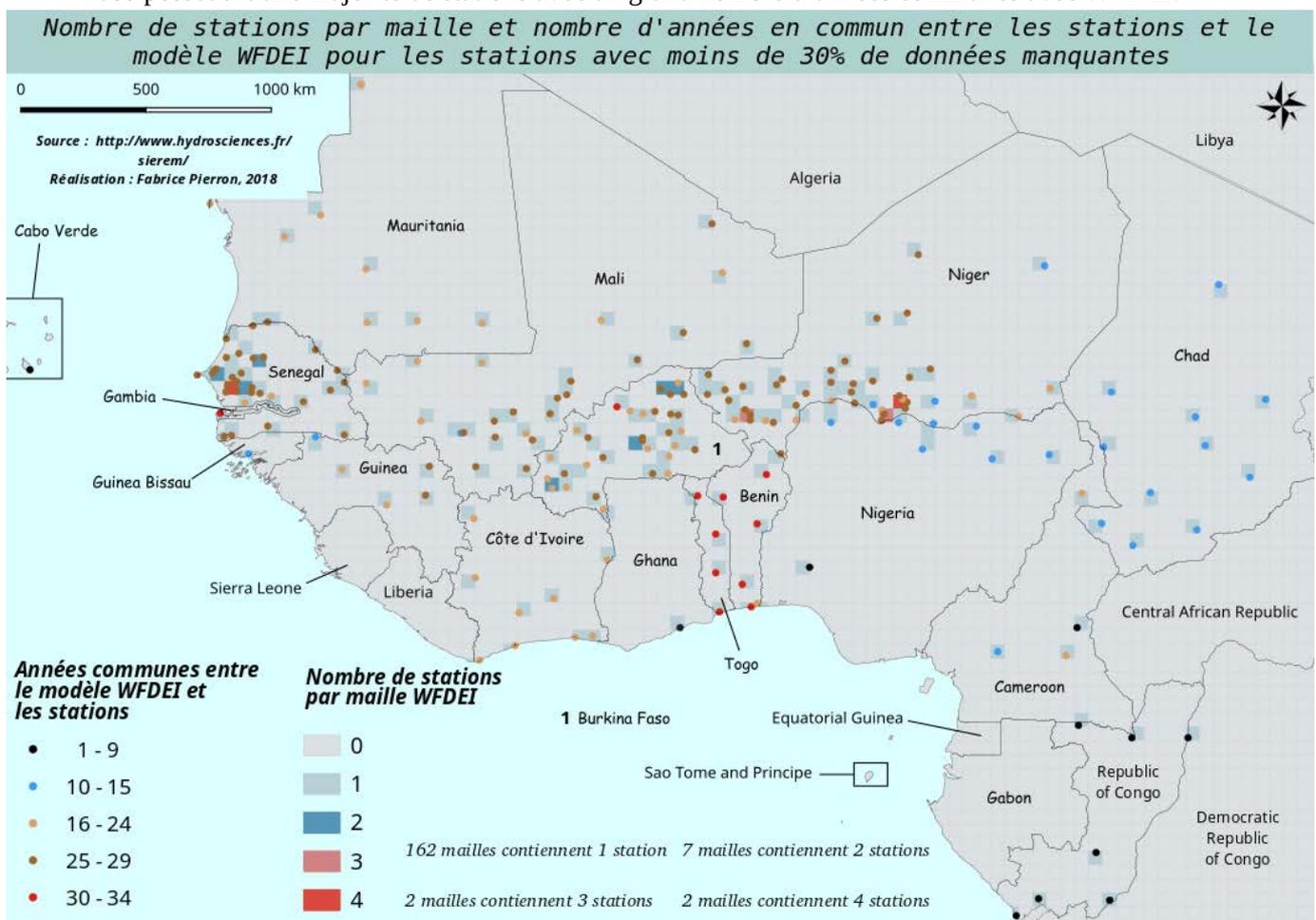


Figure 10 : Nombre de stations par maille et nombre d'années en commun entre les stations et WFDEI pour les stations avec moins de 30% de données manquantes

2) Outils et Méthodologies

2.1) Choix et calculs d'indices climatiques

Afin de réaliser une comparaison entre les données des stations météorologiques et les données maillées nous avons sélectionné une série d'indices climatiques. Ces derniers ont pour objectifs de quantifier et de qualifier les pertes d'informations lorsque l'on passe de l'échelle locale à l'échelle maillée. Il s'agit en outre de comparer les données réelles (stations) et les données WFDEI .

La sélection des indices climatiques a été réalisée à l'aide de plusieurs ressources bibliographiques et webographiques (Chaudhary *et al*, (2017) ; Weedon *et al*, (2011) ; AMMA-2050 - Technical Report No. 1, (2011) ; Karl *et al* (1999) ; Peterson *et al* (2001) ; Zhang *et al* (2005) ; http://etccdi.pamcificclimate.org/list_27_indices.shtml) . Cette phase de recherche a pour objectif d'identifier les indices les mieux adaptés aux conditions climatiques en Afrique de l'Ouest. Il s'agit en outre d'effectuer des calculs correspondant à l'ensemble des conditions climatiques présentes dans la région, en prenant en compte les « normes » et les extrêmes climatiques. Les indices climatiques choisis se sont également centrés autour de la donnée disponible pour les stations météorologiques. Ces dernières étant principalement axées autour des précipitations et des températures, nous avons privilégié les indices correspondant à ce type de variables. Ces indices devaient également permettre d'évaluer WFDEI dans l'optique d'études d'impacts agronomiques et hydrologiques.

Le tableau ci-dessous (Tableau 4) montre les 26 indices climatiques qui ont été calculés dans le cadre de ce projet :

Tableau 4 : Indices climatiques calculés dans le cadre de ce projet

Nom de l'indice	Description	Pas de temps
Annual sum	Somme des précipitations pour chaque année	Annuel
Monthly sum	Somme des précipitations pour chaque mois	Mensuel
R1 mm	Nombre de jours de précipitation par an	Annuel
R20 mm	Nombre de jours de précipitation supérieure à 20 mm par an	Annuel
R50 mm	Nombre de jours de précipitation supérieure à 50 mm par an	Annuel
Monthly T1 mm	Nombre de jours de précipitation par mois	Mensuel
Monthly T20 mm	Nombre de jours de précipitation supérieure à 20 mm par mois	Mensuel
Monthly T50 mm	Nombre de jours de précipitation supérieure à 50 mm par mois	Mensuel
Sum 3 days	Somme des précipitations tous les 3 jours	3 jours
Sum 5 days	Somme des précipitations tous les 5 jours	5 jours
Sum 10 days	Somme des précipitations tous les 10 jours	10 jours
Monsoon days	Nombre de jours de mousson chaque année	Annuel

Monsoon precipitation	Somme de précipitations durant la mousson chaque année	Annuel
Average monsoon precipitation	Moyenne de précipitations durant la mousson chaque année	Annuel
« Wetspell »	Nombre de jours consécutifs avec des précipitations	Journalier
« Dryspell »	Nombre de jours consécutifs sans précipitations	Journalier
« SPII (Simple Precipitation Intensity Index) »	Moyenne de précipitations durant les jours consécutifs de précipitations	Journalier
Cumulative Rainfall	Différence entre la somme annuelle de chaque année et la somme moyenne annuelle de précipitations de toutes les années	Annuel
R95pTOT	95 eme percentile de précipitations chaque année	Annuel
R99pTOT	99 eme percentile de précipitations chaque année	Annuel
Tas 35	Nombre de jours par an avec des températures supérieures à 35°C	Annuel
Tas 40	Nombre de jours par an avec des températures supérieures à 40°C	Annuel
Daily Temperature Range (DTR)	Différence entre les températures maximales et les températures minimales	Journalier
Diagramme Ombrothermique	Somme moyenne mensuelle de précipitations et moyenne mensuelle de températures sur l'ensemble des années	Mensuel
Indice d'aridité	Rapport entre les sommes annuelles de précipitations et la moyenne annuelle des températures (Échelle de De Martonne)	Annuel
WSDI (Warm Spell Duration)	Nombre de jours sans pluie et avec des températures supérieures ou égales aux 90eme percentile	Mensuel

En bleu = les indices avec des données de précipitations, en rouge= les indices avec des données de températures, en orange = les indices avec des données de précipitations et de températures

2.2) Les outils utilisés

Afin, d'automatiser les calculs et la cartographie des données climatiques plusieurs outils ont été utilisés. Ceux-ci vont être décrits dans les différentes sous-parties.

2.2.1) Python et ses librairies

Une grande partie des calculs, des traitements et des graphiques des données issues des stations météorologiques ainsi qu'une partie des opérations réalisées sur WFDEI ont été effectuées à l'aide du langage de programmation Python. Ce dernier offre la possibilité d'accomplir une automatisation des

traitements pour une grande partie des opérations réalisées dans le cadre de ce travail. La grande attractivité de Python réside dans la richesse de ses librairies. Le tableau 5 montre dans une première colonne les noms des librairies utilisées dans le cadre de cette étude. La deuxième colonne indique, quant à elle, le rôle de ces librairies pour les opérations effectuées.

Tableau 5 : Librairies python utilisées pour les indices climatiques

Nom de la librairie	Rôle de la librairie
glob	- Regrouper les fichiers issus d'un même chemin dans une même variable - Ouverture des données choisies
netCDF4	- Ouverture des fichiers NetCDF - Traiter les données WFDEI
os	- Création de dossier afin d'enregistrer les données des indices climatiques - Création de dossier afin d'enregistrer les sorties graphiques.
csv	- Ouverture des fichiers aux formats csv (stations)
pandas	- Nettoyage des données - Filtrage des données - Calcul des indices climatiques
numpy	- Nettoyage des données - Calcul des indices climatiques - Mesurer l'écart entre les stations et les mailles - Enregistrement des données (np.save)
scipy	- Calcul des indices climatiques - Sélection des coordonnées de maille les plus proches de ceux de la station choisie
matplotlib.pyplot	- Représentation graphique des calculs d'indices pour les stations météorologiques
plotly	- Représentation graphique des comparaisons des indices climatiques pour les stations et les WFDEI.
basemap	- Représentation cartographique des données NetCDF

En rouge = les librairies qui ont été utilisées pour l'exploitation et l'ouverture des fichiers de données, En orange = les librairies utilisées pour le traitement de la donnée, En bleu = les librairies utilisées pour la cartographie et les sorties graphiques.

Les indices et les stations météorologiques n'étant pas stockés dans une base de données, nous avons fait le choix d'utiliser des dictionnaires afin de sauvegarder les différentes données. Ces derniers se sont avérés très utiles et ont pu pallier en grande partie l'absence de base de données.

Les dictionnaires sont des objets qui peuvent en contenir d'autres. Chaque objet est contenu dans une clé. Ces dernières sont uniques, un dictionnaire ne peut pas contenir deux clés identiques. Chaque clé peut contenir des données de différents types (str, float, DataFrame...). Les appels des données stockées dans un dictionnaire se font à l'aide de ces clés. Dans notre cas, les clés correspondent au nom des indices climatiques et aux différentes métadonnées. Les Figures 11 & 12 montrent des exemples d'utilisation et de structure de dictionnaire pratiqués dans le cadre de ce projet.

Indice_climatique - Dictionnaire (28 éléments)				
Clé	Type	Taille	Valeur	
Donnees_de_depart	DataFrame	(20758, 2)	Column names: DATE, precipitation	
deficit_annuel_pr	DataFrame	(57, 2)	Column names: DATE, cumul_precipitation	
dryspell	DataFrame	(19228, 2)	Column names: DATE, consecutif_day	
jours_mousson	DataFrame	(56, 2)	Column names: years, jours_mousson	
latitude	float64	1	14.7333	
longitude	float64	1	-17.5	
mean_pr_mousson	DataFrame	(56, 2)	Column names: years, pr_mean_mousson	
month_jour_pr	DataFrame	(684, 2)	Column names: DATE, month_jour_pr	
month_pr_20mm	DataFrame	(684, 2)	Column names: DATE, month_pr_20mm	
month_pr_50mm	DataFrame	(684, 2)	Column names: DATE, month_pr_50mm	
moyenne_annuelle	DataFrame	(57, 2)	Column names: DATE, precipitation	
moyenne_mensuelle	DataFrame	(684, 2)	Column names: DATE, precipitation	
nom_station	str	1	DAKAR-YOFF	
pays	str	1	Senegal	
pr_sup_95_quantile	DataFrame	(1045, 4)	Column names: DATE, precipitation, quantile_95, consecutif_95	
pr_sup_99_quantile	DataFrame	(208, 4)	Column names: DATE, precipitation, quantile_99, consecutif_99	
somme_annuelle	DataFrame	(57, 2)	Column names: DATE, precipitation	
somme_mensuelle	DataFrame	(684, 2)	Column names: DATE, precipitation	
somme_pr_10_jours	DataFrame	(2082, 2)	Column names: DATE, precipitation	
somme_pr_3_jours	DataFrame	(6940, 2)	Column names: DATE, precipitation	
somme_pr_5_jours	DataFrame	(4164, 2)	Column names: DATE, precipitation	
spii	DataFrame	(1530, 2)	Column names: DATE, spii	
sum_pr_mousson	DataFrame	(56, 2)	Column names: years, pr_mousson	
type_station	str	1	Synoptique	
wetspell	DataFrame	(1530, 2)	Column names: DATE, consecutif_day	
year_jour_pr	DataFrame	(57, 2)	Column names: DATE, year_jour_pr	
year_pr_20mm	DataFrame	(57, 2)	Column names: DATE, year_pr_20mm	
year_pr_50mm	DataFrame	(57, 2)	Column names: DATE, year_pr_50mm	

Figure 11 : Exemple d'un dictionnaire pour la station de Dakar-Yoff au Sénégal.

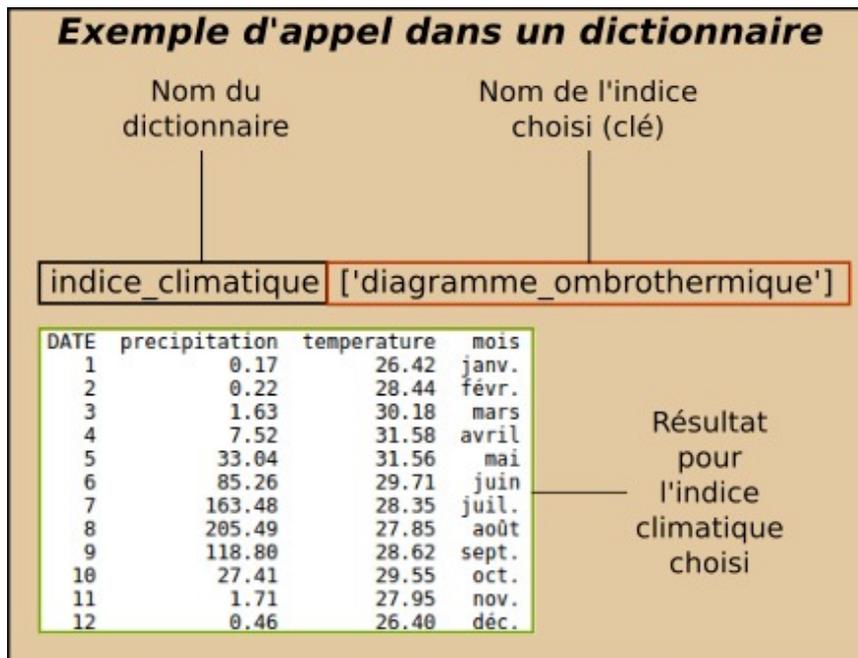


Figure 12 : Exemple d'appel de données stockées dans un dictionnaire

2.2.2) CDO et NCVIEW pour les fichiers NetCDF

La grande majorité des indices climatiques calculés sur les mailles WFDEI a été réalisée à l'aide de l'outil Climate Data Operators (CDO). CDO est une collection d'opérateurs en ligne de commande pour manipuler et analyser le climat (<https://code.mpimet.mpg.de/projects/cdo>). Cet outil prend en charge plusieurs types de formats, dont le NetCDF. Plus de 600 opérateurs sont disponibles sur CDO. . La Figure 13 ainsi que la Figure 14 montrent des exemples d'utilisation de CDO.

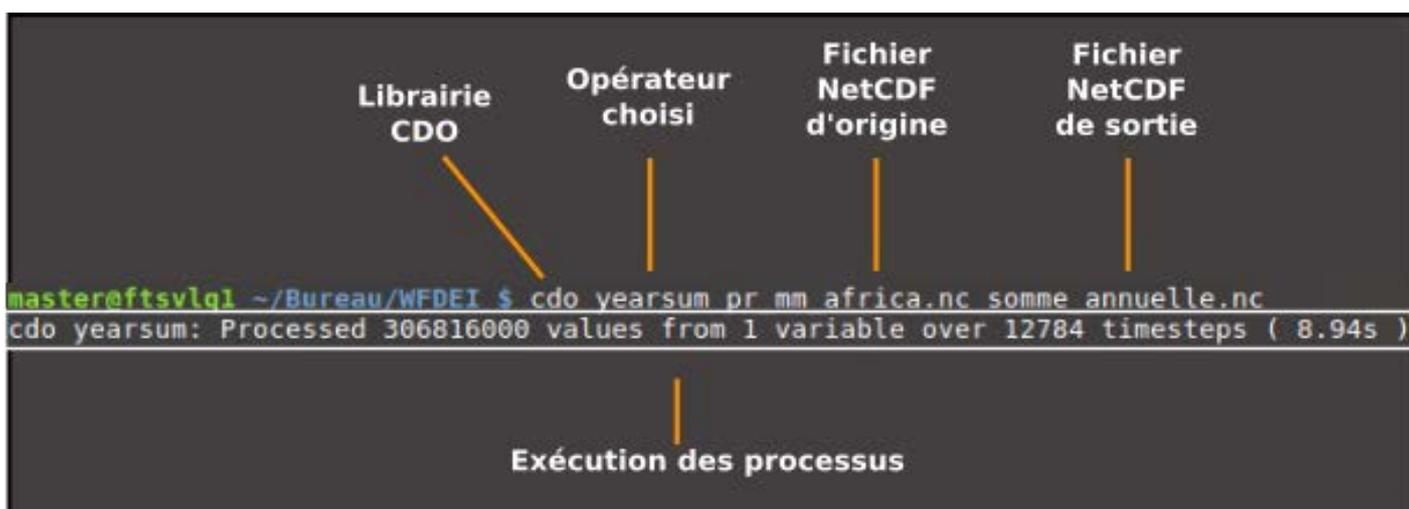


Figure 13 : Exemple d'utilisation de l'outil CDO pour le calcul de l'indice « Somme annuelle »

L'avantage de CDO réside également dans le fait qu'une combinaison d'opérateurs peut être effectuée sur une seule ligne de commande (Figure 14). Ceux-ci permettent d'accélérer et de faciliter le calcul des indices climatiques. De plus, cela évite de stocker des fichiers NetCDF temporaires,

uniquement nécessaires pour le calcul des indices finaux. Enfin, les calculs sont réalisés rapidement, ce qui constitue un autre avantage de cette librairie.

```

master@ftsvlq1 ~/Bureau/WFDEI $ cdo yearsum -gtc,50 pr mm africa.nc year pr 50mm.nc
cdo yearsum: Started child process "gtc,50 pr mm africa.nc (pipel.1)".
cdo(2) gtc: Processed 306816000 values from 1 variable over 12784 timesteps ( 20.35s )
cdo yearsum: Processed 306816000 values from 1 variable over 12784 timesteps ( 20.35s )
  
```

Librairie CDO 1er Opérateur (yearsum) 2eme Opérateur (gtc*,50) Fichier NetCDF d'origine Fichier NetCDF de sortie

Exécution des processus

Les opérations les unes après les autres

*gtc = Greater Than Constant

Figure 14 : Exemple d'utilisation de l'outil CDO avec deux opérateurs (indice R50 mm)

Lorsque les indices étaient trop complexes pour être calculés en CDO, nous sommes passés sur Python. C'est notamment le cas, pour l'indice Simple Precipitation Intensity Index (SPII) et pour les trois indices concernant le calcul de mousson (jours mousson, moyenne mousson et somme mousson).

Afin d'afficher sommairement les résultats des données issues de WFDEI, nous avons utilisé l'outil NCVIEW (http://meteora.ucsd.edu/~pierce/ncview_home_page.html). Ce dernier permet de visualiser les fichiers au format NetCDF. Cet outil s'avère très utile et très pratique pour un aperçu rapide et simple des données issues de fichiers NetCDF. Il reste néanmoins limité dès lors que l'on veut effectuer une analyse plus poussée de nos données. Le choix des représentations et des discrétisations, des fichiers NetCDF, est imposé par l'outil. Il est donc impossible d'effectuer des traitements manuellement. L'annexe 7 montre un exemple d'utilisation de NCVIEW. L'outil NCDUMP a également été utilisé afin de décrire et de connaître les données d'un fichier NetCDF (<https://www.unidata.ucar.edu/software/netcdf/netcdf-4/newdocs/netcdf/ncdump.html>).

2.2.3) QGIS pour la cartographie

L'ensemble des cartes visibles dans ce rapport ont été réalisées sur QGIS. Cette cartographie a permis de spatialiser les données des stations à l'aide du gestionnaire des sources de données « Texte Délimité ». Ce dernier nous a offert la possibilité de localiser les stations grâce aux champs X et Y, correspondant respectivement à la longitude et à la latitude, issue du fichier « inventaire stations » (cf : Figure 15). La majorité des cartes ont été réalisées sur la version 2.18 de QGIS. La symbologie sous la forme de cluster n'étant pas présente sur la version 2.18 de QGIS, nous sommes passés sur la version 3.0 pour réaliser la carte de répartition des stations (Annexe 3 : Répartition des stations potentiellement disponible sous la forme de cluster).

2.3) Chaîne de traitement pour les stations météorologiques

Les données météorologiques et les indices climatiques étant choisis, nous avons ensuite réfléchi à la structure et à la méthodologie à utiliser pour automatiser le calcul des indices climatiques pour l'ensemble des stations météorologiques. Plusieurs éléments ont dû être pris en compte dans le cadre de cette phase de traitement.

Tout d’abord, il a fallu étudier les données fournies par la base de données Hydrosiences/Sierem. Cette dernière nous a transmis les données pour 243 stations.

Ensuite, il a fallu réfléchir à la structure des données fournies par la base Hydrosiences/Sierem. Cette dernière est composée d’un fichier « **inventaire_station** » correspondant aux métadonnées. C’est notamment au sein de ce fichier que l’on retrouve la longitude et la latitude de chacune des stations (Figure 15). Les données des stations sont, quant à elles, regroupées par pays. Un fichier CSV/TXT réunit les stations d’un pays. On a donc 18 fichiers CSV/TXT (Mauritanie, Mali, Niger, Tchad, Sénégal, Cap-Vert, Gambie, Guinée, Guinée-Bissau, Côte-d’Ivoire, Ghana, Burkina Faso, Togo, Bénin, Nigeria, Cameroun, Gabon, République du Congo). Le lien entre chacune des données de stations et ses métadonnées se fait à l’aide de la clé de la station (Figure 15 & Figure 16) . On a ensuite des fichiers pour chacune des variables (précipitations, températures, températures maximales et températures minimales). Ces derniers varient selon les disponibilités des données Hydrosiences/Sierem. En effet, au sein de cette dernière, on observe une plus faible proportion de stations avec des données de températures. Ces données de températures ont été fournies pour seulement 7 pays (Bénin, Burkina Faso, Côte-d’Ivoire, Gambie, Mali, Mauritanie et Niger).

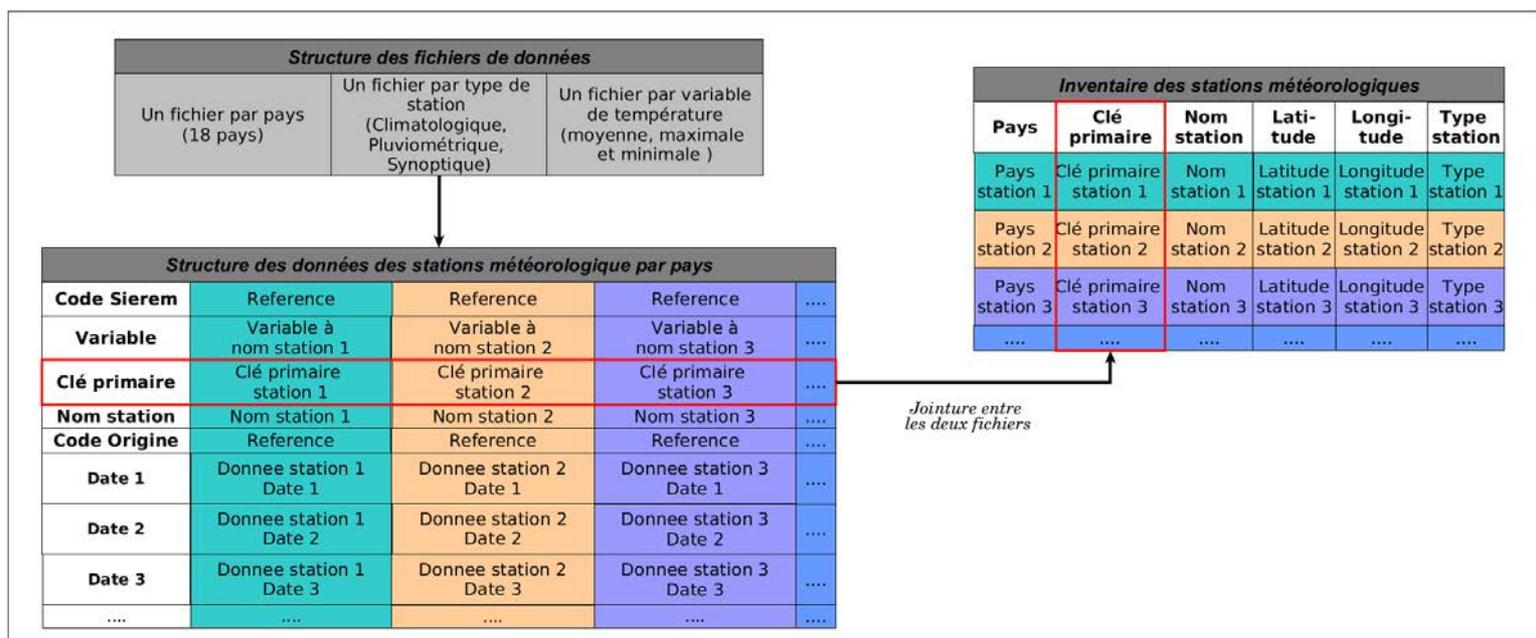


Figure 15 : Structure des données de la base Hydrosiences/Sierem

Une fois la structure des données appréhendée une étape de nettoyage et de filtrage des données a été nécessaire. Cette étape a pour objectif de faciliter et d’accélérer l’automatisation des calculs d’indices pour les 243 stations fournies. Les différentes opérations effectuées dans le cadre de cette étape sont visibles en Annexe (**Annexe 5 : Filtrage et nettoyage des données des stations météorologiques**).

Le filtrage des données a pour objectif d’accorder les données des stations avec les données des modèles climatiques. Le premier filtrage réalisé dans le cadre de cette étape a consisté à sélectionner les données à partir de l’année 1950. Cette année correspond aux dates de début des modèles (outre WFDEI). Ainsi, les calculs des indices climatiques peuvent être extrapolés à d’autres modèles. Nous verrons ultérieurement que la présence de données manquantes a, néanmoins, put modifier les comparaisons et les calculs d’indices.

La dernière partie de la chaîne de traitement a été l’automatisation du calcul des 26 indices climatiques, sur python, pour chacune des stations . Cette phase a probablement été la plus longue du

stage. La première étape a été la création d'un dictionnaire vide. Ensuite, les métadonnées de la station choisie ont été implémentées au dictionnaire (Figure 16). Puis, les indices ont été calculés un par un et ont été progressivement insérés dans le dictionnaire. Une fois l'ensemble des indices climatiques introduit dans le dictionnaire, nous l'avons sauvegardé à l'aide de la fonction **np.save** au format « npy ». La donnée sauvegardée contient l'ensemble des indices climatiques d'une station ainsi que ses métadonnées (Figure 11). Cette opération a ensuite été répétée pour l'ensemble des stations avec moins de 30% de données manquantes.

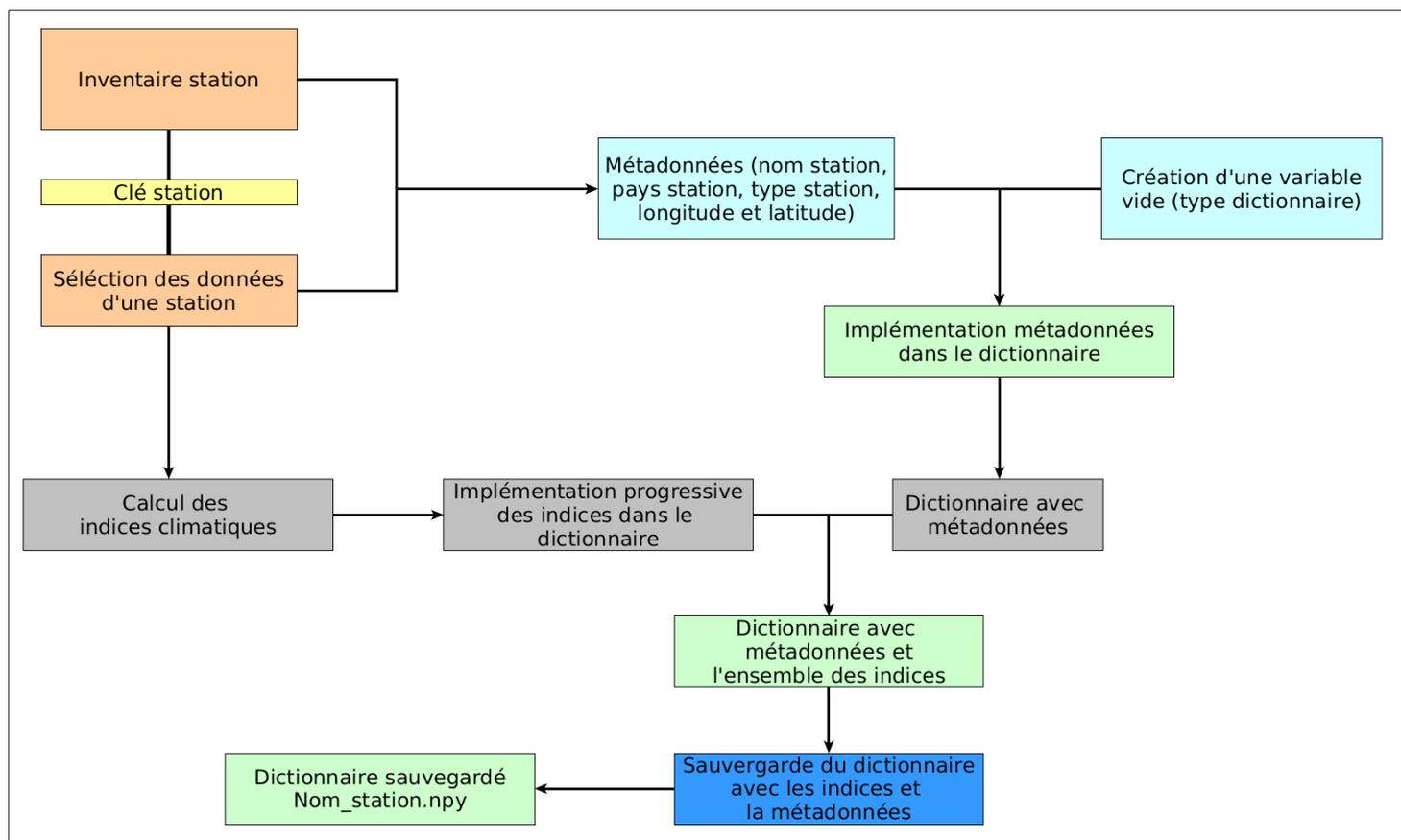


Figure 16 : Calcul des indices climatiques sur Python

La chaîne de traitement ci-jointe (Figure 17) récapitule l'ensemble des étapes qui ont été réalisées dans le cadre de l'automatisation du calcul des indices climatiques sur les stations météorologiques.

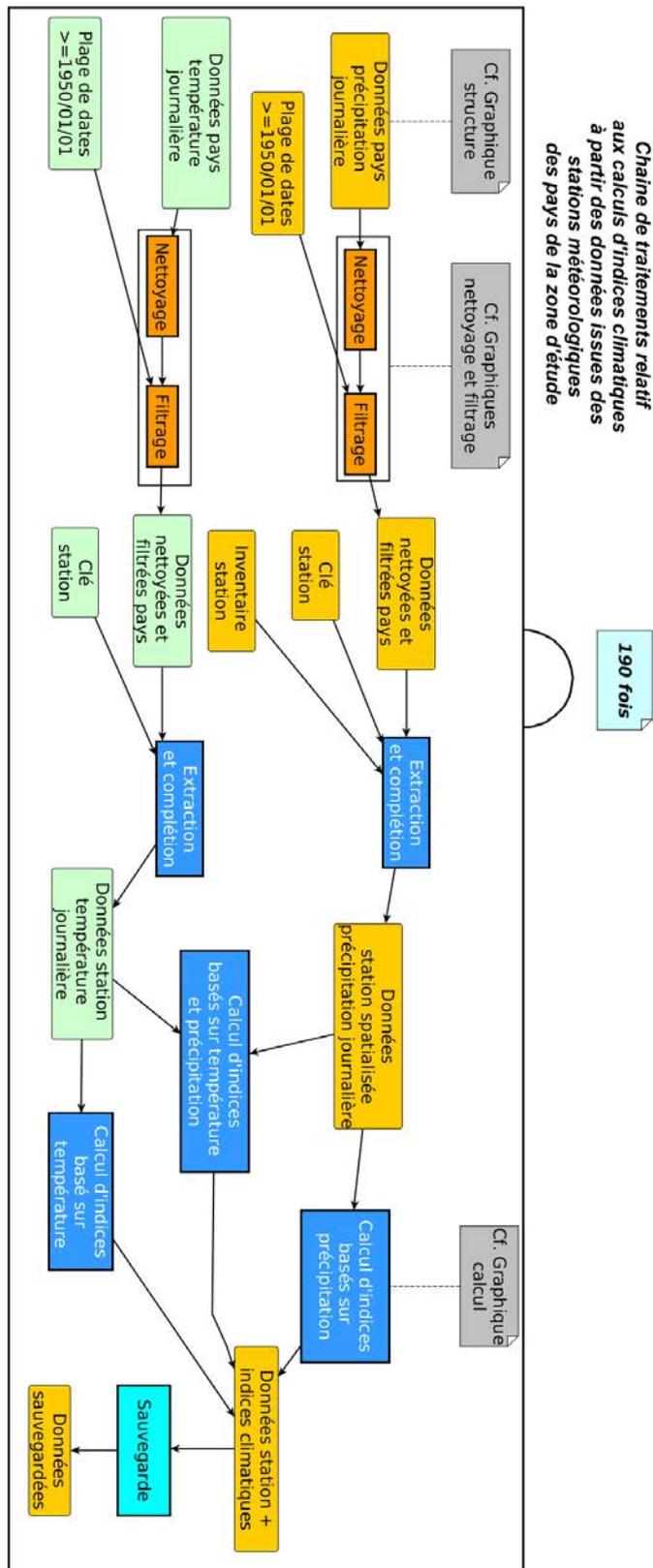


Figure 17 : Chaîne de traitement pour le calcul des indices climatiques pour les stations météorologiques

2.4) Chaîne de traitement pour WFDEI

Le calcul des indices climatiques et de la chaîne de traitement a été plus rapide et plus simple pour les données maillées. Cette différence s'explique en partie par la plus grande homogénéité des données pour WFDEI. Ceux-ci ont notamment nécessité moins de post-traitement, les données manquantes et les dates de début des modèles étant déjà quasiment adaptées aux calculs des indices. Le filtrage des données a « uniquement » consisté à déterminer la localisation des mailles par rapport aux stations météorologiques, l'objectif final étant d'avoir l'ensemble des stations et des mailles dans une même zone d'étude (Figure 10). De plus, les étapes de traitement ont été plus rapide pour les modèles grâce à l'utilisation de l'outil **CDO**. La Figure 18 montre les étapes de traitement qui ont été nécessaire afin de calculer les indices climatiques.

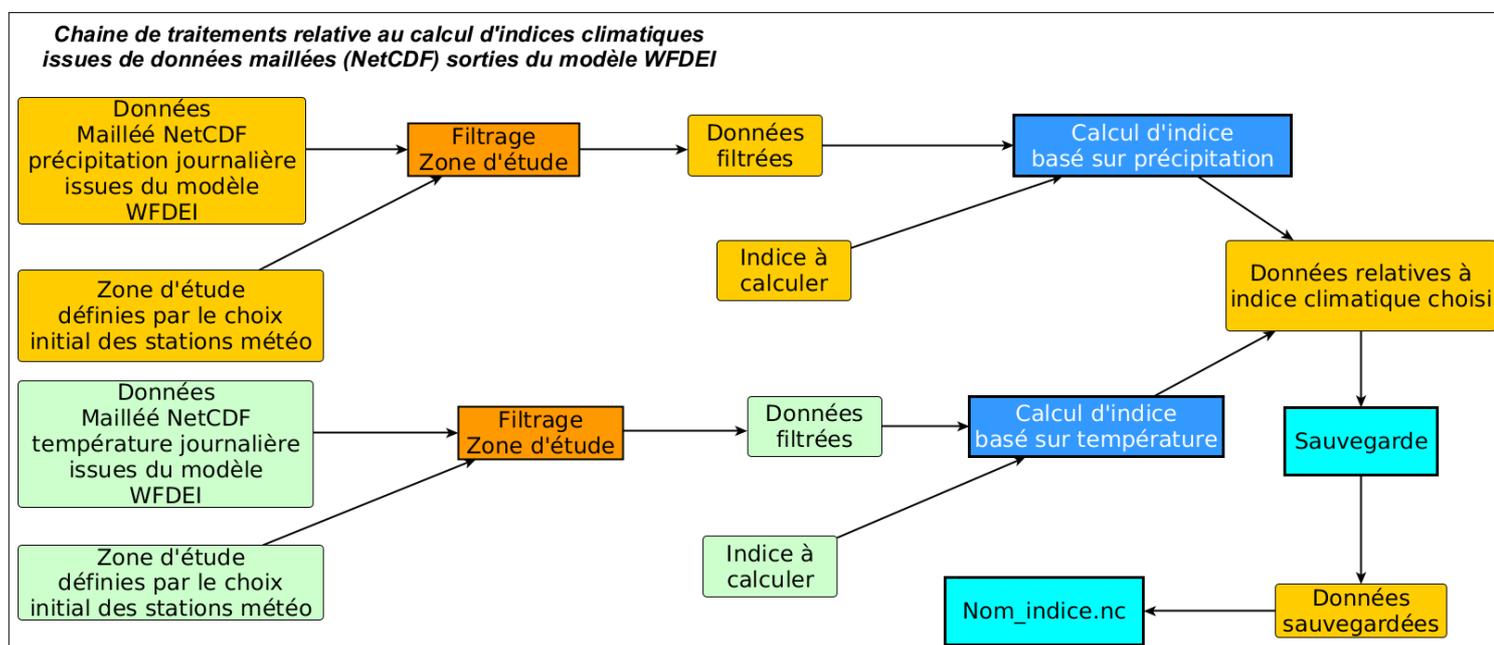


Figure 18: Chaîne de traitement relative aux données maillées (NetCDF) sorties de WFDEI

2.5) Comparaison des données de stations et des données WFDEI

2.5.1) Automatisation des sorties graphiques

La comparaison des données météorologiques et des données issues de WFDEI a été effectuée à l'aide de python. La structure et la chaîne de traitement sont composées de plusieurs étapes (Figure 19). Il s'agit dans un premier temps de sélectionner la station souhaitée, puis, de récupérer la longitude et la latitude de la maille correspondant à la station. Ensuite, on choisit l'indice climatique pour lequel on souhaite effectuer l'analyse. La sélection de l'indice va alors être automatiquement opérée pour WFDEI. Une fois l'indice climatique choisi pour la station et la maille, un graphe est automatiquement créé et sauvegardé.

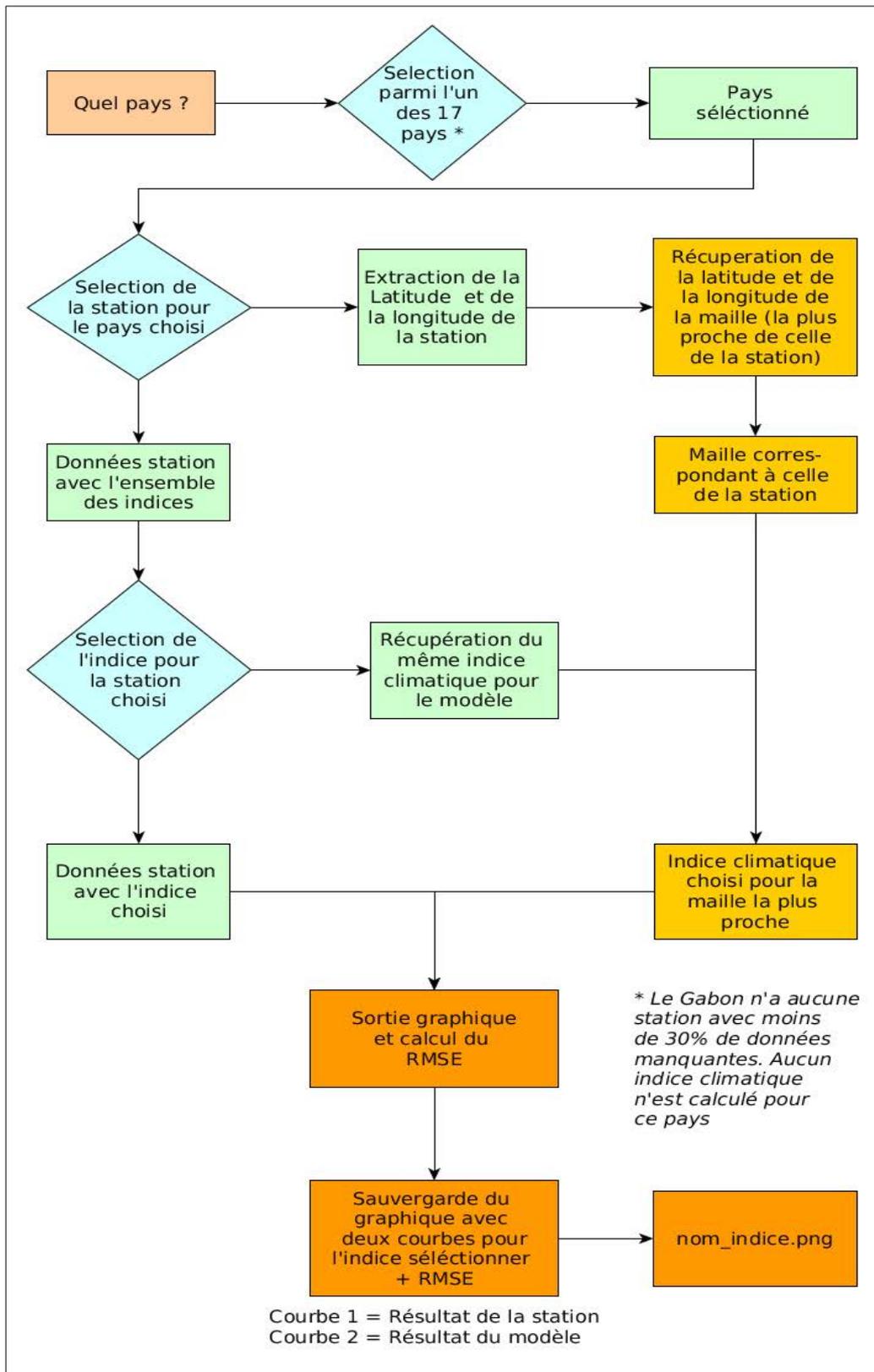


Figure 19 : Chaîne de traitement pour l'automatisation des sorties graphiques

2.5.2) Calcul d'opérateurs statistiques

Pour chacune des sorties graphiques, le Root-Mean-Square Error (RMSE) est calculé. Ce dernier correspond à la racine carrée de l'erreur quadratique moyenne. Celui-ci nous donne une mesure synthétique de l'erreur globale dans une seule valeur. Une erreur quadratique moyenne faible montre une petite variance de l'erreur entre les deux variables étudiées (Bouthevillain et Mathis, 1995). Le RMSE permet donc d'évaluer l'écart moyen entre les résultats de l'indice climatique choisi pour la station sélectionnée et pour la maille la plus proche.

$$\text{RMSE} = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

Y_{pred} = Les résultats de l'indice climatique issus de WFDEI

Y_{ref} = Les résultats de l'indice climatique issus de la station

N = Durée de la période de l'indice climatique (jours, mois ou années)

Afin de prendre en compte le mieux possible ces variabilités, nous avons également calculer le Relative Root Mean Square Error (RRMSE). Cet indicateur est calculé en divisant le RMSE par la valeur moyenne des données de la station mesurée. La formule de ce dernier est la suivante :

$$\text{RRMSE} = \frac{\text{RMSE}}{\bar{I}} * 100$$

RMSE = Écart moyen entre la station et le modèle

\bar{I} = La moyenne de l'indice climatique pour la station choisie

VI Résultats

1) Sorties graphiques

L'automatisation des sorties graphiques nous permet de visualiser l'écart entre les deux jeux de données pour les 190 stations et les 26 indices. Il est donc possible d'effectuer 4 940 sorties graphiques ($190 \times 26 = 4\,940$). Ces sorties nous permettent d'effectuer une comparaison entre la station choisie et le maille WFDEI la contenant. Deux courbes sont alors créées pour le même indice (Figure 20). La première courbe correspond aux données de la station et la seconde correspond à la maille WFDEI contenant la station.

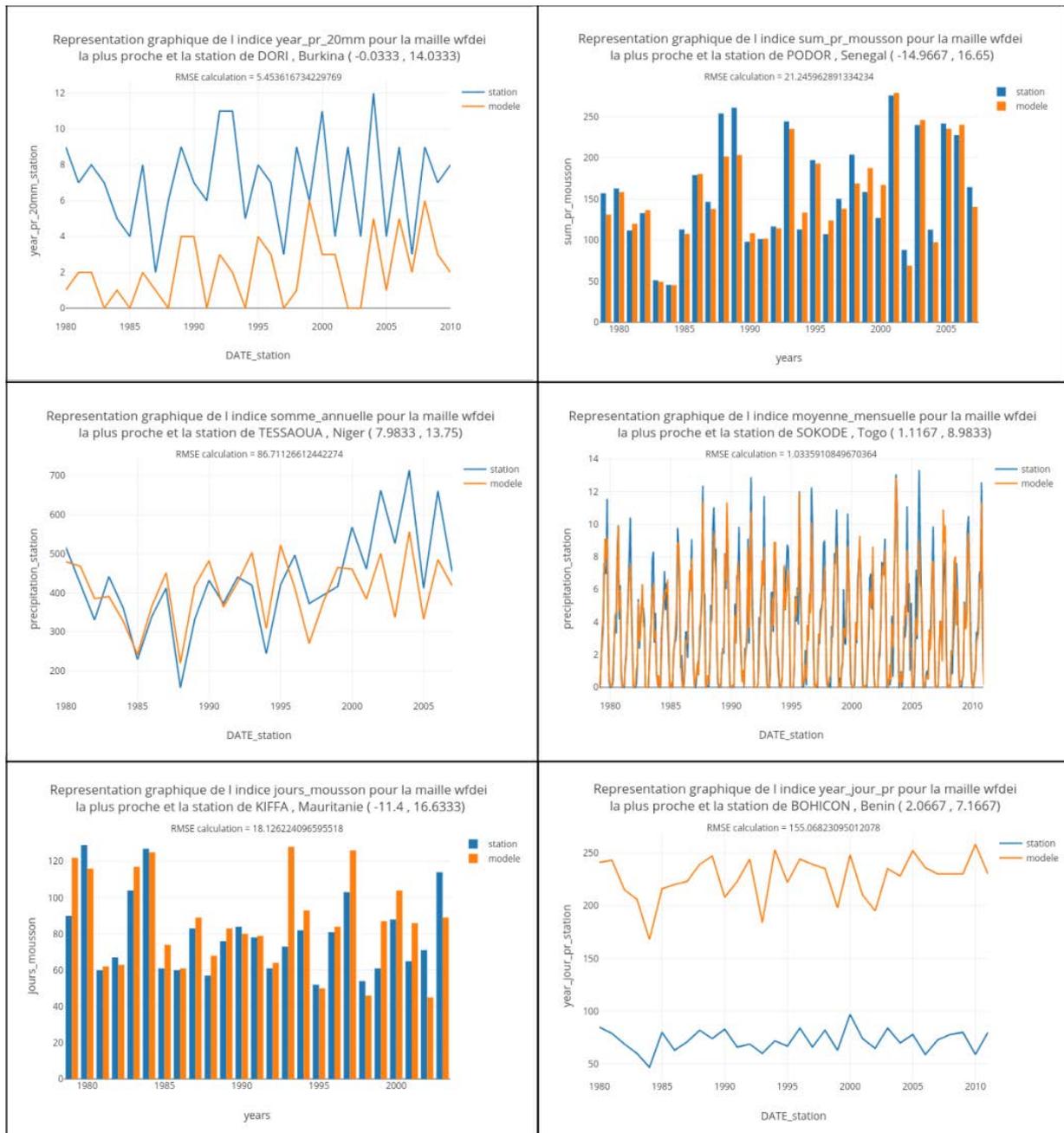


Figure 20: Exemple de sorties graphiques pour plusieurs stations et plusieurs indices climatiques

2) Calcul du RMSE

Une moyenne des RMSE pour 16 indices a été calculée sur l'ensemble des stations. Ces 16 indices ont été choisis pour deux raisons. Premièrement, il s'agissait des indicateurs les plus pertinents et deuxièmement, pour une question de temps, il n'a pas été possible d'effectuer une analyse sur l'ensemble des indices. Dans la suite du projet, il serait pertinent de calculer les RMSE pour l'ensemble des stations. Par ailleurs, certaines stations ont été supprimées lors du calcul du RMSE. Ces suppressions s'expliquent parce que les données manquantes influençaient trop fortement les valeurs moyennes du RMSE et ce malgré le filtrage à 30%. Au final, on a 173 stations avec des RMSE calculés.

Afin, d'évaluer les stations dans leur globalité, nous avons fait le choix de regrouper ces dernières par zones climatiques. Ainsi, chaque zone climatique possède les valeurs de RMSE pour chacun des 16 indices et pour l'ensemble des stations avec une pluviométrie relativement proche. Le nombre de stations finales étant limité, il a fallu définir des zones climatiques avec au moins cinq stations. La répartition des stations par seuils de somme de précipitations moyennes annuelles sur 35 ans diffère donc légèrement par rapport aux seuils définis pour la Figure 1.

Une fois cette classification effectuée, nous avons réalisé une moyenne des RMSE, pour chaque indice, pour l'ensemble des stations localisées dans une même zone climatique. Le tableau 6 montre la répartition des stations par zone climatique

Tableau 6 : Répartition des stations afin de calculer le RMSE moyen par zone climatique

Zones climatiques	Nombre de stations
< 150 mm	5
150 – 300 mm	13
300 – 500 mm	37
500 – 800 mm	56
800 – 1200 mm	38
1200 – 1750 mm	23
> 1750 mm	1*

* La station de Tabou en Côte d'Ivoire est seule dans sa zone climatique (> 1750 mm). Cette zone là ne sera par conséquent pas étudiée séparément. La station a néanmoins été conservé pour le calcul de la moyenne sur les 173 stations.

Le Tableau 7 montre le résultat des RMSE pour l'ensemble des 16 indices et pour les 173 stations. On peut notamment y observer les différents écarts entre les stations et les mailles par zones climatiques. Ce tableau donne une première information sur l'écart qui peut exister entre les deux jeux de données.

Tableau 7: RMSE sur 173 stations et pour 16 indices climatiques

Nom Indice	Unité	Zones climatiques (précipitations annuelles basées sur 35 ans 1979-2013)						
		Moyenne (173 stations)	inf 150mm (5 stations)	150-300mm (13 stations)	300-500mm (37 stations)	500-800mm (56 stations)	800-1200mm (36 stations)	1200-1750mm (23 stations)
Moyenne Annuelle	mm	0,40	0,11	0,26	0,28	0,40	0,52	0,54
Somme Annuelle	mm	151,95	31,04	68,06	91,39	147,01	206,62	243,78
Year Jour_pr	jour	69,58	7,51	23,34	44,75	58,58	95,34	130,57
Year Pr_20mm	jour	8,42	0,96	2,31	4,99	7,12	11,89	16,46
Year Pr_50mm	jour	1,81	0,17	0,56	1,06	1,44	2,34	3,79
Month Jour_pr	jour	8,40	1,58	3,94	6,52	7,78	10,69	13,00
Month Pr_20mm	jour	1,35	0,24	0,56	0,95	1,24	1,76	2,20
Month Pr_50mm	jour	0,40	0,06	0,16	0,27	0,35	0,51	0,70
Moyenne Mensuelle	mm	0,89	0,23	0,54	0,71	0,88	1,08	1,26
Somme Mensuelle	mm	30,59	7,13	16,08	21,86	31,06	38,24	44,68
Deficit Annuel	mm	139,14	31,83	53,68	89,72	139,17	170,07	243,65
Jours Mousson	jour	36,30	96,72	25,66	30,24	42,07	29,48	36,40
Somme Mousson	mm	108,88	30,63	45,65	69,40	134,07	121,07	144,29
Tas 35	°C	9,04	15,51	18,84	11,54	10,67	2,85	Pas de Station
Tas 40	°C	0,63	1,02	0,80	0,11	0,04	1,07	Pas de Station
Indice Aridité	Seuil de DeMartonne	3,37	0,93	1,39	2,06	3,86	4,45	Pas de Station

Néanmoins, cette analyse des RMSE par zones climatiques et par indices est rendue difficile pour plusieurs raisons :

Premièrement, les indices n'ont pas les mêmes pas de temps. Ainsi, les valeurs des RMSE pour les indicateurs avec des pas de temps différents ne sont pas significatifs. Il est en effet difficile de comparer des données annuelles et des données mensuelles.

Deuxièmement, les unités et les variables étudiées sont différentes selon les indices. Il est ainsi difficile d'effectuer des comparaisons entre des indices en mm, en jours ou en °C.

Troisièmement, les zones climatiques étant basées sur des précipitations moyennes annuelles différentes, il est complexe de comparer les zones avec des faibles précipitations et les zones avec des fortes précipitations.

Afin, de pallier à l'ensemble de ces difficultés, nous avons fait le choix de comparer les indices et les stations par zones climatiques à l'aide du RRMSE.

3) Étude comparative des indices climatiques à l'aide du RRMSE

3.1) Résultats du RRMSE pour six indices climatiques

Le calcul du RRMSE va nous offrir la possibilité d'effectuer des comparaisons entre les différents indices climatiques, tout en négligeant les difficultés précédemment citées pour le RMSE. Pour effectuer cette comparaison nous allons nous appuyer sur les travaux de Despovic *et al* (2016). Ces derniers ont établi une échelle qui permet de qualifier la précision de modèle. Pour eux, la précision du modèle est considérée comme excellente lorsque le RRMSE est inférieur à 10 %, bonne si le RRMSE est entre 10 % et 20 %, juste si le RRMSE est entre 20 % et 30 % et mauvaise si le RRMSE est supérieur à 30%. C'est cette échelle de Despovic *et al* (2016) que l'on va utiliser afin d'évaluer l'écart entre les données des stations et les données WFDEI.

Les RRMSE ont donc été mesurés pour les 16 indices avec des RMSE calculés. Néanmoins, afin de ne pas avoir trop d'informations nous avons fait le choix de nous concentrer sur certains indices. Ces derniers vont nous permettre de représenter les variabilités des résultats pour l'ensemble des stations. Les indices choisis pour cette partie d'étude sont les suivants :

Somme annuelle, Nombre de jours de précipitation par an, Moyenne mensuelle,

Jours moussons, Nombre de jours de précipitation supérieure à 20 mm par an, indice d'aridité

Ces indices présentent l'avantage d'avoir des pas de temps, des variables et des seuils différents. Ainsi, ils peuvent permettre de caractériser au mieux les erreurs entre les données issues des stations et les données WFDEI. La Figure 21 montre la moyenne des RRMSE pour les 173 stations précédemment mentionnées pour les six indices sélectionnés. L'indice d'aridité étant partiellement basé sur des données de températures, 34 stations seulement ont été utilisées pour le calcul du RRMSE. Les stations fournies par la base de données Hydrosociences/Sierem ne possèdent pas toutes des données de températures.

En s'appuyant, sur l'échelle définie par Despovic *et al* (2016), on peut observer que certains indices sont justes alors que d'autres sont mauvais. Ainsi, les indices **Somme annuelle, Jours moussons** et **l'indice d'aridité** sont relativement proches pour les 173 stations et pour la maille WFDEI. On a un RRMSE respectif de 27,78 %, 29,02 %, et 20,48 %. À l'inverse, les indices **Nombre de jours de précipitation supérieure à 20 mm par an, Nombre de jours de précipitation par an** et **Moyenne mensuelle** sont mauvais avec des valeurs respectives de 163,09 %, 82,75%, et 60,06%

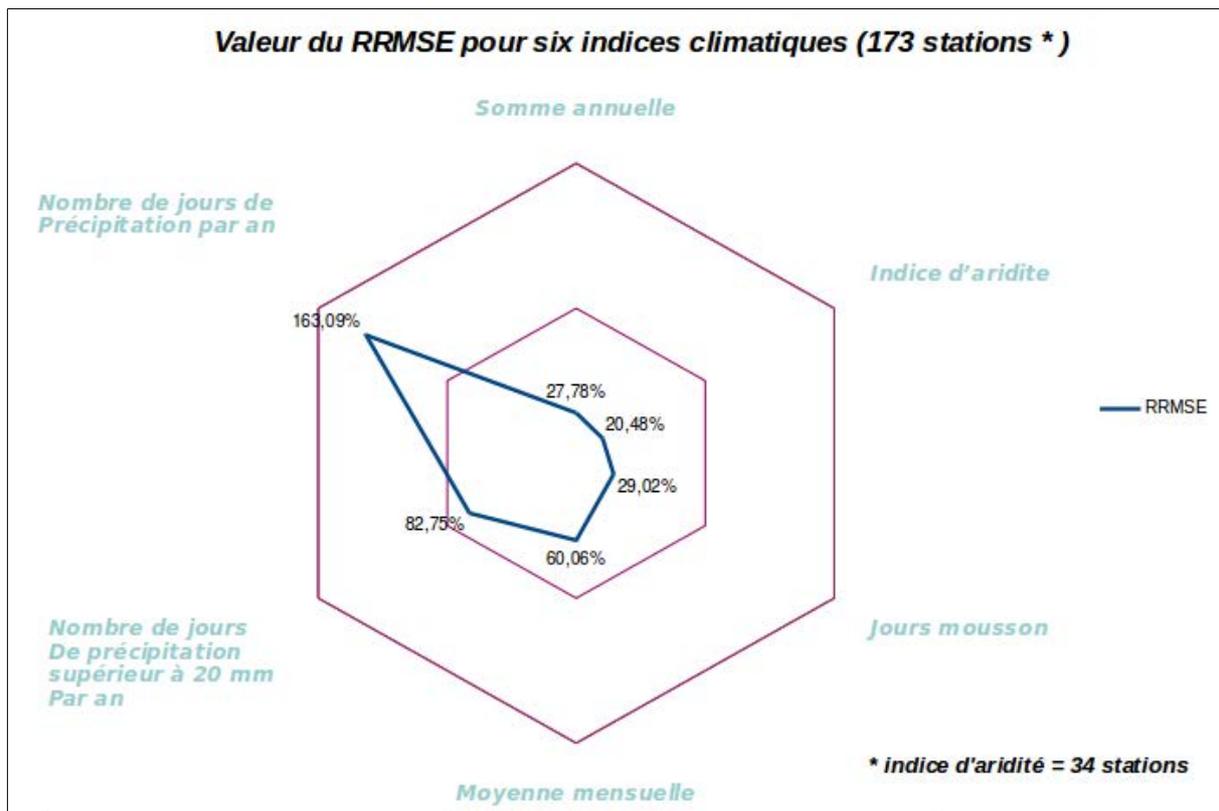


Figure 21: RRMSE moyen pour six indices climatiques

Cette première figure du RRMSE donne un premier aperçu des erreurs qui peuvent exister entre les stations et WFDEI. Néanmoins, il y a une grande diversité de contexte climatique pour ces résultats de RRMSE. Ainsi, nous allons effectuer la même analyse par zones climatiques pour chacun des six indices climatiques. Pour ce faire, nous allons nous appuyer sur les mêmes zones climatiques que pour le calcul des RMSE (Tableau 6 & 7).

3.2) Résultats du RRMSE par zones climatiques

L'analyse menée dans cette sous-partie vise à étudier les résultats des RRMSE en fonction des zones climatiques. Le tableau 8 montre les résultats pour les six indices précédemment sélectionnés, Sur ce dernier, on peut observer plusieurs informations importantes pour cette étude :

- Premièrement, on peut constater que la zone climatique < 150 mm a des valeurs de RRMSE importants pour l'ensemble des indices climatiques sélectionnés. Les résultats pour cette zone sont nettement plus importants que pour les autres. Cette différence tend à montrer que WFDEI ne parvient pas à représenter les données climatiques pour les zones de faibles précipitations.
- Deuxièmement, il semblerait que les zones climatiques avec des précipitations relativement importantes (800-1200 mm et 1200-1750 mm) soient mieux représentées par WFDEI que les autres. On observe en effet des valeurs de RRMSE globalement plus faibles pour ces deux zones climatiques.
- Troisièmement, on peut observer des résultats variables par zone climatique en fonction des indices. Comme sur la figure 21, on observe des résultats fluctuants en fonction de l'indice choisi.

Sur une même zone climatique, certains indices auront tendance à être bons alors que d'autres auront tendance à être mauvais.

Tableau 8 : RRMSE par zones climatiques pour six indices

Zones climatiques	Nombre de Stations	Somme Annuelle	Year_jour_pr	Year_pr_20mm	Moyenne Mensuelle	Jours Mousson	Indice Aridité
<150 mm	5	52,21%	139,10%	195,63%	149,70%	111,24%	38,37%
150-300 mm	13	34,09%	127,62%	80,97%	94,67%	32,40%	25,56%
300-500 mm	37	27,17%	174,52%	83,64%	70,32%	26,95%	20,03%
500-800 mm	56	30,72%	168,12%	79,57%	59,81%	28,74%	22,35%
800-1200 mm	38	24,06%	165,44%	76,55%	42,51%	19,94%	15,36%
1200-1750 mm	23	19,49%	154,67%	77,28%	35,92%	28,84%	19,67%
Toutes les stations	173	27,78%	163,09%	82,75%	60,06%	29,02%	20,48%

La dernière ligne correspond aux valeurs que l'on a pu voir sur la figure 21

3.3) Étude de l'influence des données manquantes

Afin, d'étudier du mieux possible les écarts entre les données des stations et WFDEI, nous avons fait le choix d'analyser l'influence des données manquantes pour les résultats des six RRMSE précédemment observés. Pour ce faire, nous avons réalisé la même étude que pour le tableau 8, mais uniquement avec les stations ayant une part de données manquantes inférieure ou égale à 5 % (Tableau 9). Ce dernier montre des valeurs plus faibles pour la majorité des indices et pour une majorité des zones climatiques.

Tableau 9 : RRMSE par zones climatiques pour les stations avec une part de données manquantes inférieure ou égale à 5 % pour six indices

Zones climatiques	Nombre de Stations	Somme Annuelle	Year_jour_pr	Year_pr_20mm	Moyenne Mensuelle	Jours Mousson	Indice Aridité
<150	4	48,40%	88,83%	153,91%	124,63%	108,96%	38,37%
150-300	11	29,86%	121,52%	81,61%	85,03%	30,54%	25,56%
300-500	24	23,21%	155,49%	80,67%	71,14%	23,57%	20,03%
500-800	34	22,23%	139,76%	75,38%	58,66%	25,47%	13,24%
800-1200	23	17,99%	146,39%	71,82%	42,94%	17,65%	14,17%
1200-1750	18	14,92%	140,45%	75,01%	33,88%	31,49%	19,67%
Toutes les stations	115	22,00%	141,00%	78,81%	58,71%	27,75%	18,61%

Afin de quantifier cet écart de manière plus synthétique, nous avons fait un calcul de différence en nous basant sur les données issues des tableaux 8 et 9 (Tableau 10). Pour ce faire nous avons calculé la différence entre les valeurs du Tableau 9 et du Tableau 8 (Tableau 9 - Tableau 8 = Tableau 10). Dans la très grande majorité des cas, on peut observer des valeurs négatives. Cela signifie que les RRMSE ont tendance à être plus faibles dès lors que l'on considère uniquement les stations avec une part de données manquantes inférieure ou égale à 5 %.

Tableau 10 : Différence entre les RRMSE pour les stations avec une part de données manquantes inférieures ou égale à 5 % et les RRMSE pour l'ensemble des 173 stations

Zones climatiques	Somme Annuelle	Year_jour_pr	Year_pr_20mm	Moyenne Mensuelle	Jours Mousson	Indice Aridité
<150	-3,81%	-50,27%	-41,72%	-25,07%	-2,28%	0,00%
150-300	-4,24%	-6,10%	0,64%	-9,64%	-1,87%	0,00%
300-500	-3,97%	-19,03%	-2,97%	0,83%	-3,39%	0,00%
500-800	-8,49%	-28,36%	-4,19%	-1,15%	-3,27%	-9,11%
800-1200	-6,07%	-19,04%	-4,73%	0,43%	-2,29%	-1,19%
1200-1750	-4,57%	-14,22%	-2,27%	-2,05%	2,65%	0,00%
Toutes les stations	-5,78%	-22,09%	-3,95%	-1,35%	-1,27%	-1,87%

En bleu = les indices avec des valeurs de RRMSE plus faibles pour les stations avec une part de données manquantes $\leq 5\%$, *en rouge* = les indices avec des valeurs de RRMSE plus élevées pour les stations avec une part de données manquantes $\leq 5\%$, *en orange* = les indices avec des valeurs de RRMSE égaux pour les stations avec une part de données manquantes $\leq 5\%$ et pour l'ensemble des stations

3.4) Synthèse des résultats des RRMSE

La Figure 22 montre les résultats de RRMSE avec et sans données manquantes pour les six indices climatiques et pour l'ensemble des zones. Cette dernière permet notamment de récapituler les trois informations principales découlant de l'analyse des RRMSE :

- Premièrement, on a pu constater des différences importantes entre les indices climatiques. Certains semblent montrer une relative proximité entre les données issues des stations et les données issues de WFDEI (Somme annuelle, indice d'aridité, jours mousson), alors que d'autres, comme le nombre de jours de précipitations par an, indiquent des grandes différences entre les deux jeux de données.
- Deuxièmement, on a observé des différences notables de RRMSE en fonction des zones climatiques. Il semblerait, en effet, que certaines régions ont une erreur plus faible. C'est notamment le cas des zones climatiques avec une pluviométrie moyenne comprises entre 800 et 1750 mm. À l'inverse, la zone climatique avec des précipitations moyennes annuelles < 150 mm est presque systématiquement moins bien représentée par rapport aux autres zones climatiques. Cette différence peut en grande partie s'expliquer par la faible propension d'événements pluvieux.

- Troisièmement, les résultats obtenus à partir de l'ensemble des données et des stations avec une part de données manquantes $\leq 5\%$ montrent un écart important. Il semblerait, en effet, que la présence de données manquantes est tendance à augmenter la valeur des RRMSE pour l'ensemble des indices et des zones climatiques. Ainsi, on peut constater que la présence de données manquantes tend à avoir un effet négatif lors des calculs d'indices climatiques.

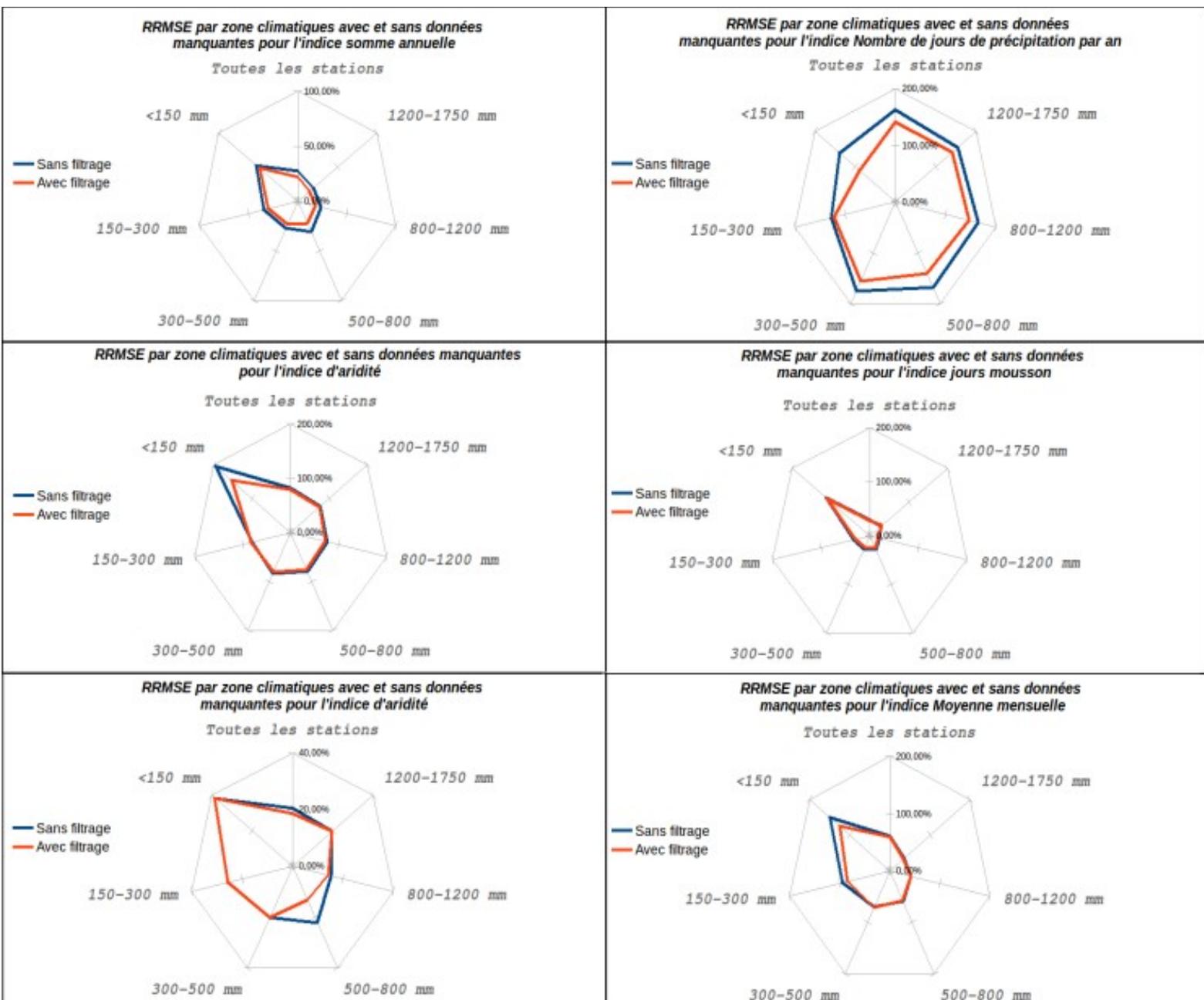


Figure 22 : RRMSE par zones de pluviométries moyennes annuelles avec et sans données manquantes pour six indices climatiques

4) Étude de cas intra-maille

4.1) Choix de la maille d'étude spécifique

Afin d'analyser les écarts de résultats entre les mailles et les stations, de manière moins générale, nous avons centré nos résultats sur une maille. Cette étude intra-maille vise à estimer la perte d'information liée à la résolution spatiale des données WFDEI. La maille que nous avons sélectionné afin de réaliser cette analyse, est localisée au Sénégal. Cette dernière a été choisie pour trois raisons.

Premièrement, elle possède en son sein quatre stations météorologiques (Kaolack, Foundiougne, Gossas et Fatick). Cela permet d'effectuer une comparaison entre les résultats de chacune des stations et la maille qui les contient.

Deuxièmement, les quatre stations contenues dans la maille possèdent très peu de données manquantes (Figure 23). En effet, elles ont toutes 0% de données manquantes à l'exception de la station de Foundiougne qui en possède 7%.

Troisièmement, les quatre stations possèdent 28 années communes avec WFDEI. Ces années communes vont permettre de réaliser des comparaisons des résultats avec les mailles sur une période relativement longue.

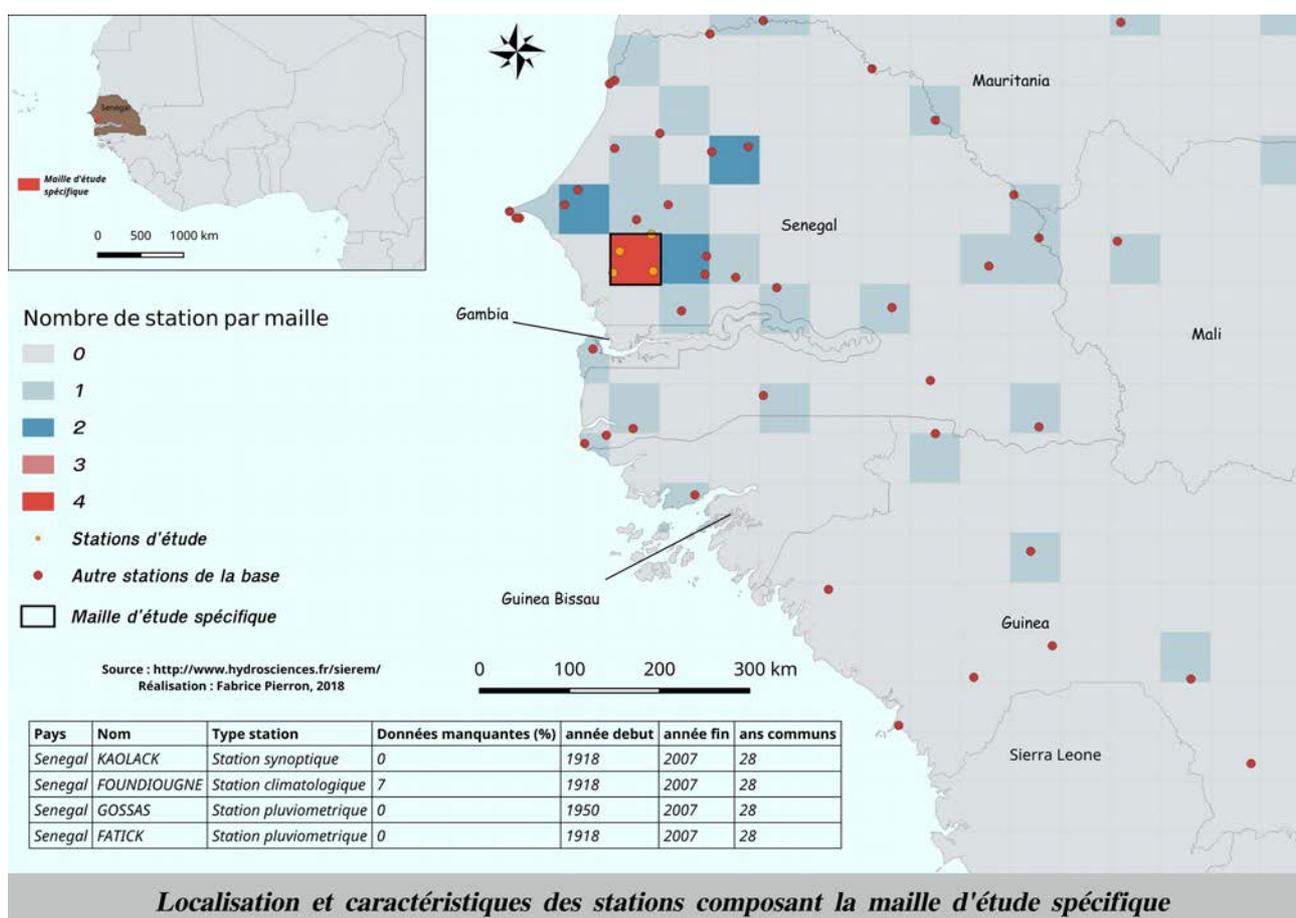


Figure 23 : Localisation et caractéristiques des stations composant la maille d'étude spécifique

Afin de mettre en avant les différents résultats obtenus pour les données fournies, nous allons nous focaliser sur trois indices climatiques bien spécifiques, pour la maille WFDEI et les quatre

stations localisées dans cette dernière. Pour effectuer, une représentation la plus lisible possible, nous avons choisi de nous concentrer sur des indices avec un pas de temps annuel. Cela évite d'avoir un surplus d'information, ce qui rendrait presque illisible la donnée. Ces trois indices sont les suivants :

- **somme annuelle de précipitations**
- **nombre de jours de précipitations supérieure à 1 mm par an**
- **nombre de jours de précipitations supérieure à 20 mm par an**

4.2) Étude de cas pour l'indice Somme Annuelle

Le premier cas que l'on a pu observer est caractérisé par une relative proximité des résultats entre les données issues des stations et WFDEI. La Figure 24, qui montre le résultat de WFDEI et des quatre stations moyennées pour les sommes annuelles de précipitations, indique des tendances proches. Le RRMSE entre WFDEI et les quatre stations n'est que de 23,42% (34 mm/ans). Cette proximité est également variable selon les années comme peut l'indiquer les chevauchements et les croisements entre les deux courbes. Ce premier cas montre ainsi qu'une relative proximité entre les sorties WFDEI et les données réelles peut exister.

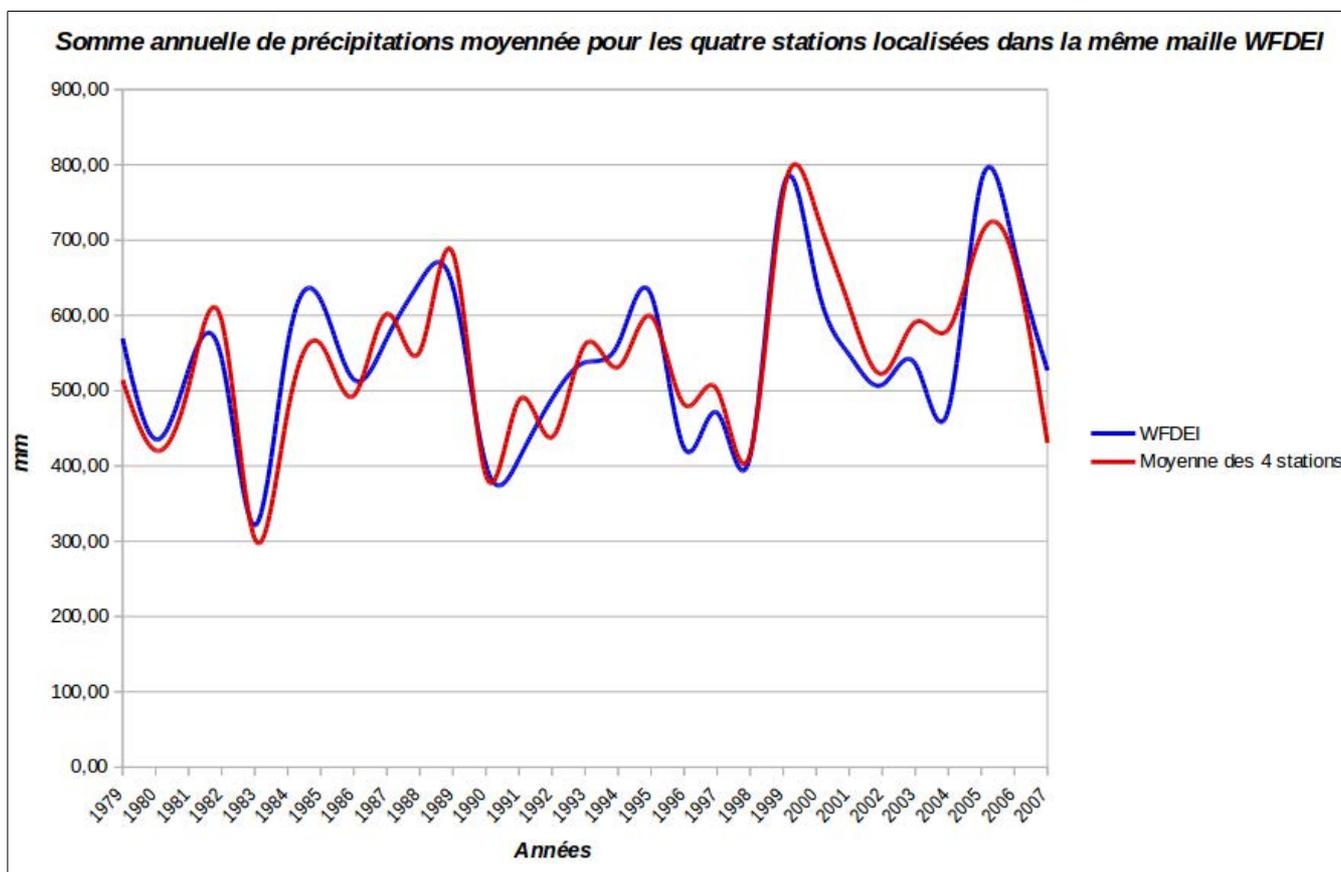


Figure 24 : Somme annuelle de précipitations moyennées pour quatre stations localisées dans une même maille WFDEI

Néanmoins, cette proximité peut être quelque peu atténuée lorsque l'on observe séparément les résultats pour chacune des stations (Figure 25). En effet, lorsque l'on s'intéresse aux quatre stations de manière individuelle, on constate que des variabilités peuvent exister. Les courbes de chacune des stations peuvent fortement varier d'une année sur l'autre et d'une station à l'autre. Ces variabilités

peuvent être mises en avant par les moyennes des sommes annuelles de précipitations pour les quatre stations et pour les différents RMSE et RRMSE (Figure 25).

Concernant, la moyenne on peut observer que cette dernière varie entre 488,8 mm (Gossas) et 604,77 mm (Foundiougne). Pour les RRMSE, on constate également des écarts importants entre les stations, avec une valeur minimale de RRMSE de 13,43 % pour la station de Kaolack et une valeur maximale de RRMSE de 33,29 % pour la station de Foundiougne. Cette variabilité tend à confirmer les différences que l'on avait précédemment pu observer.

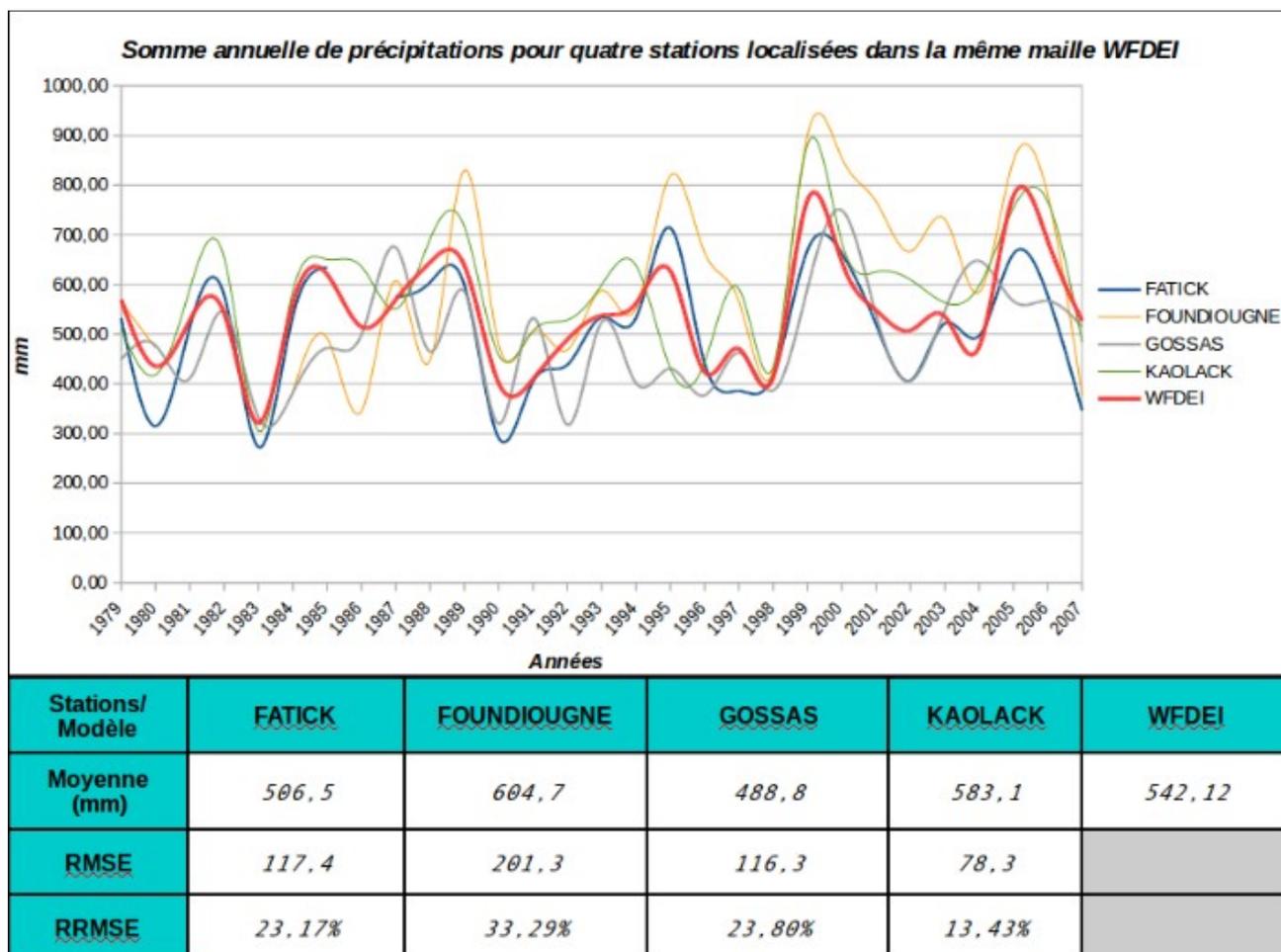


Figure 25 : Somme annuelle de précipitations pour les quatre stations localisées dans une même maille WFDEI

4.3) Étude de cas pour l'indice nombre de jours de précipitations supérieure à 1mm/ans

Le deuxième cas que l'on a pu observer dans le cadre de cette étude est caractérisé par une surestimation des données issues de WFDEI. La Figure 26, qui montre les mêmes données que précédemment pour l'indice « nombre de jours de précipitations supérieure à 1 mm », indique une différence notable entre les deux jeux de données. On peut en effet observer que les données de WFDEI surestiment nettement le nombre de jours de pluie par an. Le RRMSE entre WFDEI et les quatre stations est très élevé avec une valeur moyenne de 118,02 % (38 jours/ans). De plus, on peut observer que la courbe de WFDEI est systématiquement supérieure aux données des quatre stations. Cette absence de chevauchement ou de croisement indique que le modèle surestime le nombre de jours

de pluie par an pour l'ensemble des 28 années communes. Les valeurs moyennes des quatre stations et de WFDEI, pour cet indice, confirment cette tendance. Pour les quatre stations, on observe une moyenne maximale de 37,07 jours (Gossas), alors que la moyenne de WFDEI est de 72,52 jours. En moyenne, WFDEI a deux fois plus de jours de précipitations que les quatre stations.

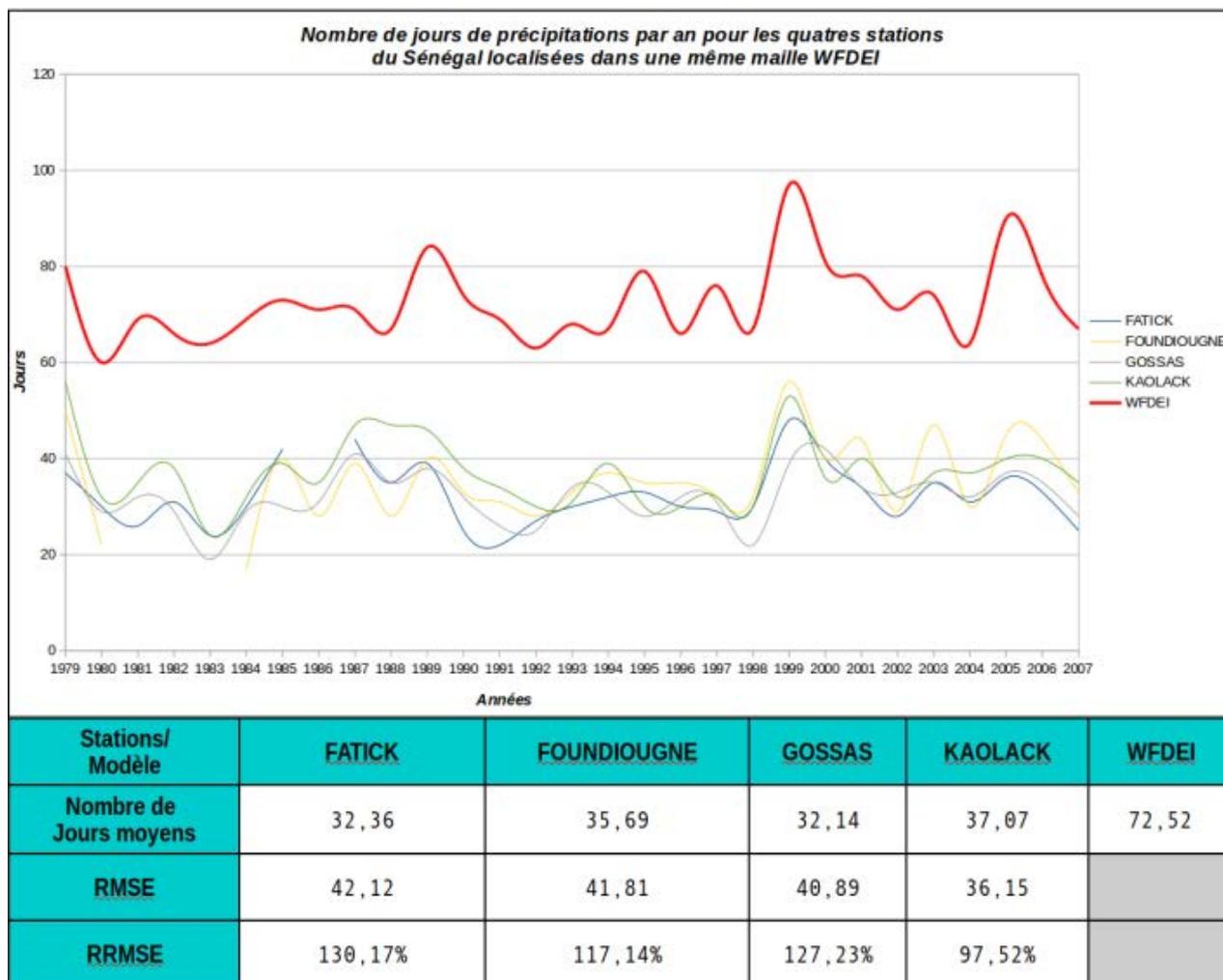


Figure 26 : Nombre de jours de précipitations par an moyennés pour quatre stations localisées dans une même maille WFDEI

4.4) Étude de cas pour l'indice nombre de jours de précipitations supérieure à 20 mm/ans

Le troisième cas que l'on a pu observer dans le cadre de la comparaison entre les stations et les mailles WFDEI est caractérisé par une sous-estimation des valeurs des modèles par rapport aux données réelles. La Figure 27 montre les mêmes types de données que pour les Figures 25 et 26, mais pour l'indice « Nombre de jours de précipitations supérieure à 20 mm ». Sur cette dernière, on peut constater que le modèle sous-estime le nombre de jours de précipitation par an supérieure à 20 mm. Le RRMSE entre WFDEI et les quatre stations est encore élevé avec une valeur de 75,77 % (6,21 jours/ans).

Les résultats obtenus pour cet indice sont diamétralement opposés à ceux que l'on a pu observer pour le nombre de jours de précipitations par an (Figure 26). En effet, pour cet indice, on peut observer

que les données des stations sont presque exclusivement supérieures aux données de WFDEI. L'absence de chevauchement ou de croisement entre les deux courbes confirme cette tendance (Figure 27). La valeur moyenne, pour cet indice, pour WFDEI, est de 2,90 jours/ans alors que la valeur minimale est de 8,29 jours/ans pour les stations (Fatick).

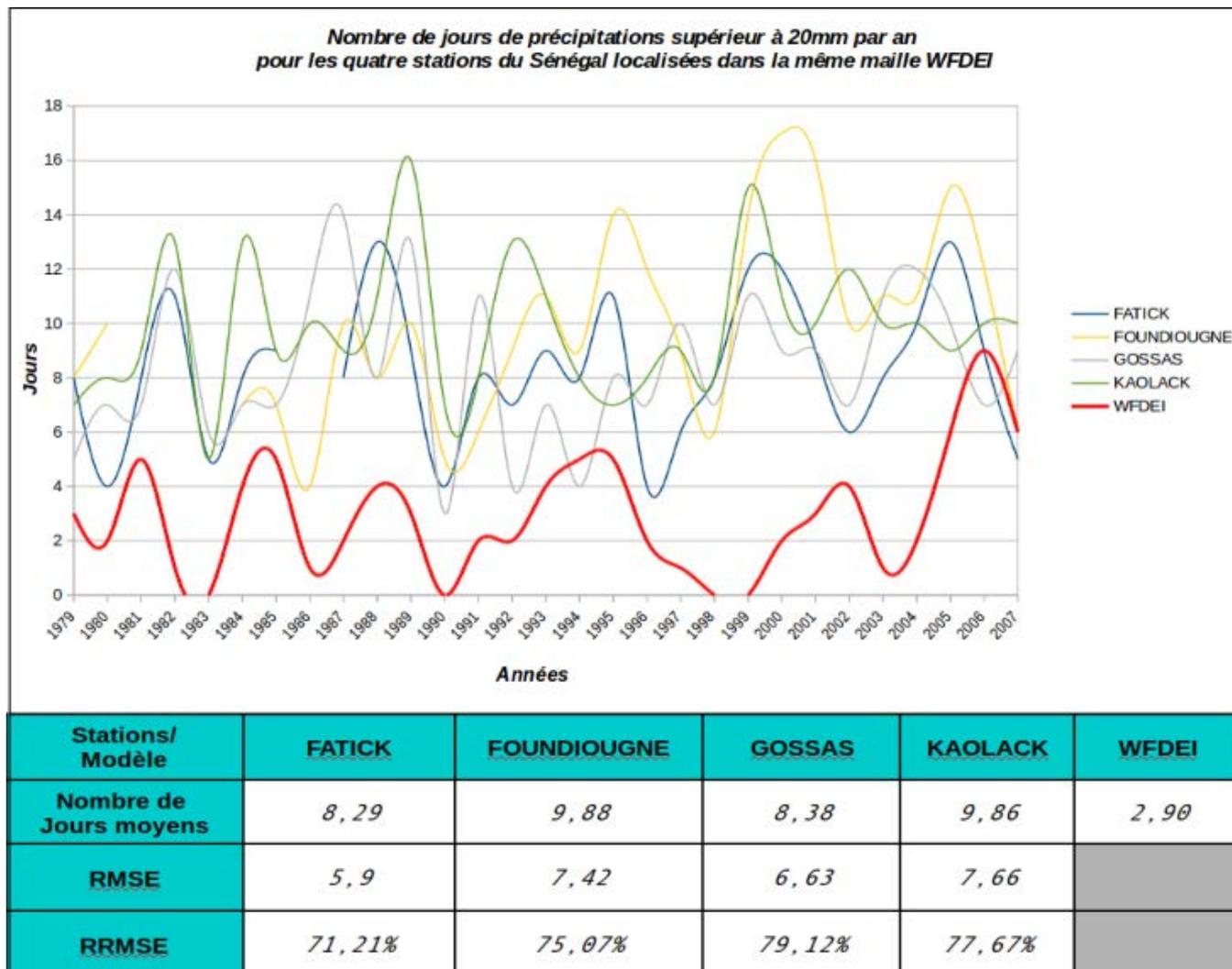


Figure 27 : Nombre de jours de précipitations supérieure à 20 mm par an moyenné pour quatre stations localisées dans une même maille WFDEI

VII Limites et perspectives

L'ensemble des traitements, des calculs et des résultats, nous ont offert la possibilité d'effectuer une analyse pour les deux échelles de données pour de nombreux indices climatiques. Cette analyse a notamment permis de mettre en avant des différences entre les résultats des données maillées et des données réelles. Des indicateurs statistiques, tels que le RMSE ou le RRMSE, ont été calculés afin de quantifier et qualifier les écarts de valeurs entre les deux jeux de données. Ces indicateurs ont notamment permis d'observer des écarts notables en fonction des indices et des zones climatiques.

Ainsi, nous avons notamment pu observer que certains indices climatiques pouvaient être proches pour les deux jeux de données. C'est notamment le cas de l'indice « **Somme annuelle** » ou « **Jours de mousson** ». Cette proximité semble relativement logique puisque les données de WFDEI sont issues d'une extrapolation basée sur des cumuls annuels de précipitations (Weedon et al, 2011). Néanmoins, on a également observé que certains indices montrés des écarts significatifs entre les deux jeux de données. Il semblerait en effet, que WFDEI ne parvienne pas à modéliser les valeurs extrêmes. Ainsi, les valeurs minimales et maximales ont tendance à être lissées. Ces tendances ont notamment pu être observées sur les indices « **Nombre de jours de précipitations supérieure à 1 mm par an** » et « **Nombre de jours de précipitations supérieure à 20 mm par an** » .

Les résultats que nous avons pu observer dans le cadre de ce stage, semblent également suivre ceux obtenus par Manzanos *et al* (2014). Ces derniers ont effectué une étude comparative, au Ghana, portant sur 14 stations et sur une série de modèles climatiques. Afin de quantifier ces écarts, Manzanos *et al* (2014) se sont également appuyés sur une série d'indicateurs climatiques. Ils ont alors mis en évidence des écarts significatifs entre les données issues des stations et les données issues des modèles, en fonction des indices étudiés.

Les travaux de Manzanos *et al* (2014) se sont aussi portés sur les variabilités en fonction des zones climatiques. Ces derniers ont montré que certaines zones climatiques avaient tendance à être mieux modéliser que d'autres. Ces résultats semblent confirmer les impressions que l'on a pu constater dans le cadre de ce projet. En effet, nous avons observé que les zones de faibles précipitations avaient tendance à être moins bien traitées par WFDEI.

Les travaux que nous avons effectués ont également offert la possibilité d'étudier les variabilités intra-maille. Ainsi, on a pu observer que la résolution spatiale des mailles est trop grande pour analyser l'ensemble des disparités climatiques des stations.

Enfin, nous avons constaté que la part de données manquantes pouvait influencer les résultats pour les calculs d'indices climatiques. En effet, dans le cadre de ce travail, nous avons observé un écart entre les stations avec peu ou pas de données manquantes ($\leq 5\%$) et les stations avec une part de données manquantes importantes.

Pour approfondir les résultats obtenus dans le cadre de ce stage, quelques améliorations pourraient apporter :

Tout d'abord, il semble important de rappeler qu'un seul modèle climatique a été analysé. Une comparaison des données issues d'autres modèles climatiques aurait été pertinente. Cette dernière aurait offert la possibilité d'effectuer une analyse plus aboutie des données climatiques. On aurait également pu mettre en avant les forces et les faiblesses de chacun des modèles en s'appuyant sur les différents indicateurs climatiques. Ces derniers nous auraient permis de quantifier et de qualifier les écarts de résultats entre les modèles climatiques.

Afin, d'analyser et d'approfondir du mieux possible les données des modèles, ils seraient pertinents de poursuivre l'analyse des indices climatiques, en intégrant d'autres stations. Ces dernières devraient permettre d'améliorer les analyses des erreurs entre les mailles et les stations. À court terme, cet ajout pourrait se centrer sur les stations de la base de données Hydrosociences/Sierem. À plus long terme, l'ajout de stations issues d'autres sources de données offrirait la possibilité d'améliorer les

études comparatives entre les deux échelles d'analyses. Pour réaliser cet approfondissement, il serait pertinent d'insérer l'ensemble des stations dans une même base de données. Cette dernière offrirait également la possibilité de stocker l'ensemble des différents jeux de données. L'ensemble des modèles climatiques étudiés, pourraient également être insérés dans cette dernière.

La question de la qualité des données fournies est également un enjeu central dans le cadre d'étude sur le climat. Ainsi, la gestion et le traitement des données manquantes ont fréquemment entraîné des problèmes dans l'analyse des résultats, comme nous avons pu le voir précédemment. À titre de comparaison, l'étude menée par Manzananas *et al* (2014) n'était basée que sur les stations avec une part de données manquantes inférieure ou égale à 2%. Afin de pallier, à cette concentration de données manquantes, il serait pertinent d'extrapoler les valeurs en se basant sur les données des années antérieures. Cette extrapolation offrirait la possibilité d'avoir une analyse plus complète pour l'ensemble des modèles climatiques.

Du point de vue de l'analyse statistique, il serait intéressant de calculer d'autres indicateurs d'écarts que le RMSE et le RRMSE. Ces derniers nous offrent une première approche de comparaison, néanmoins, ils ne sont pas systématiquement adaptés aux calculs de certains indices climatiques.

Nous avons notamment pu observer que le RMSE ne permettait pas de faire des comparaisons entre les indices climatiques avec des variables, des pas de temps ou des seuils différents. Le RRMSE, quant à lui, a montré ces limites pour le calcul des indices avec des faibles valeurs. (**nombre de jours de précipitations supérieure à 50 mm par an, nombre de jours de précipitations supérieure à 50 mm par mois, tas 40...**). Pour ces indices, les valeurs du RRMSE auront souvent tendance à être très élevées alors que les écarts réels ne seront pas nécessairement importants. Ainsi, il serait pertinent de calculer de nouveaux indicateurs d'écarts pour l'ensemble des indices climatiques. Ces derniers devraient permettre d'effectuer une analyse statistique pour l'ensemble des indices climatiques.

Enfin, des cartes des écarts entre le/les modèle(s) et les stations en fonction des zones et des indices climatiques pourraient être réalisées. Ces dernières permettraient de spatialiser les zones climatiques plus ou moins bien représentées par le/les différent(s) modèle(s). Cette cartographie pourrait aussi bien se faire à l'échelle des stations qu'à celle des mailles.

VIII Conclusion

Dans le cadre de ce travail, nous avons quantifié et qualifié l'erreur entre des données de stations météorologiques et des données issues d'une réanalyse climatique, WFDEI. Pour ce faire, nous avons sélectionné des indices climatiques. Ceux-ci ont été calculés, pour WFDEI et pour 190 stations météorologiques fournies par la base de données Hydrosiences/SIEREM. Ces indices climatiques permettent de quantifier les erreurs entre les deux jeux de données. Pour représenter l'ensemble des contextes climatiques, nous nous sommes appuyés sur des variables, des pas de temps et des seuils différents.

Afin, d'appréhender du mieux possible les écarts entre les données locales et les données maillées, nous avons effectué le calcul d'indices climatiques sur les stations avec une part de données manquantes < 30 %. Ce seuil nous a permis de faire un compromis entre la qualité et la quantité pour les données des stations fournies. Pour rappel, un nombre significatif de données météorologiques est nécessaire pour effectuer une analyse sur l'ensemble de la zone d'étude. Par ailleurs, cette sélection avait pour objectif de limiter l'influence de l'absence de données. En effectuant ce filtrage, nous avons pu calculer les indices climatiques pour les 190 stations restantes.

Les différentes cartes présentes dans ce rapport ont permis de spatialiser, de caractériser et de qualifier les données des stations et des mailles. Celles-ci ont été d'une précieuse aide pour analyser les disparités entre les deux échelles étudiées. Outre des cartes, des sorties graphiques ont été réalisées. Ces dernières permettent de comparer visuellement les différences entre la station sélectionnée et la maille la contenant. Afin, de quantifier cette disparité, nous avons choisi de calculer le RMSE pour chacune des comparaisons. Celui-ci offre la possibilité d'évaluer l'écart moyen entre les résultats de l'indice climatique choisi pour la station sélectionnée et pour la maille la plus proche.

Par la suite, nous avons effectué une analyse en nous focalisant sur les RMSE de l'ensemble des stations pour chacun des indices. Afin, d'effectuer une étude plus approfondie des écarts entre les stations et les modèles, nous avons choisi de répartir les RMSE en zones climatiques. Ces derniers nous ont permis de réaliser une première analyse, en montrant une première différence selon les zones. Néanmoins, en raison de variables, de seuils et de pas de temps différents, les RMSE se sont avérés difficilement comparables les uns par rapport aux autres.

Ainsi, nous avons choisi de calculer le RRMSE en nous appuyant sur l'échelle de Despovic *et al* (2016). Cette échelle nous a offert la possibilité d'avoir des valeurs permettant de qualifier les indices pour chaque stations. Le calcul du RRMSE nous a également permis d'effectuer une comparaison entre les différents indices climatiques calculés. En nous focalisant uniquement sur les résultats globaux des indices climatiques, nous avons pu observer une grande diversité de résultats selon les RRMSE. Certains indices ont montré des bonnes valeurs de RRMSE, alors qu'à l'inverse, d'autres indicateurs ont des valeurs très éloignés des stations. Cela signifie que ces indices sont mal représentés par WFDEI. Ces valeurs ont tendance à évoluer en fonction du seuil et du pas de temps de l'indice.

Afin, d'effectuer une comparaison des RRMSE de manière moins globale, nous avons fait le choix de répartir ces derniers par zones climatiques. Les résultats obtenus ont montré deux disparités. D'abord, on a constaté des écarts importants entre les différentes zones climatiques. Ils sembleraient que les zones avec des faibles précipitations moyennes annuelles soient mal modélisées par WFDEI. Ensuite, on a observé que la part de données manquantes, pouvait affecter les résultats des indices climatiques. Dans l'ensemble, les valeurs des RRMSE semblent plus faibles pour les stations avec peu de données manquantes, et ceux peu importe la zone climatique étudiée.

À moyen terme, ces indices doivent permettre d'améliorer les sorties des modèles. Cette amélioration doit d'offrir la possibilité aux acteurs de l'agriculture (organisations paysannes, filières, secteur privé agricole, banque de développement agricole, fournisseurs d'intrants, services agricoles et de météorologie) d'accéder à une donnée climatique fiable. De plus, pour améliorer les analyses, des nouveaux modèles et des nouvelles stations doivent être étudiés.

Bilan personnel

Le premier mois a été consacré à la prise en main des nouveaux outils. En outre, il a fallu appréhender le format NetCDF et le système d'exploitation Linux. Ces nouveautés ont nécessité un temps d'adaptation. Concernant, les données au format NetCDF, de nombreuses interrogations sur leur représentation graphiques et/ou cartographiques sont venues émailler le stage. Il s'agissait en outre de réfléchir à la manière de représenter les différents indices climatiques en fonction des pas de temps. Le questionnement du calcul des indices climatiques issues des données NetCDF a également été au centre de ce projet.

D'un point de vue thématique, de nombreuses connaissances ont été acquies. J'ai notamment pu apprendre des nouveaux éléments sur le climat en Afrique de l'Ouest. En outre, les relations entre les précipitations, les températures et leur rôle dans l'agriculture ont été appréhendés. L'importance de ces relations climat/agriculture me semblent être primordiale à l'avenir. Cette relation semble être d'autant plus importante, que les conditions climatiques futures restent relativement incertaines. La question de la relation entre l'accès aux ressources (eaux, nourritures..) en lien avec les changements climatiques me paraissent indispensables à comprendre. Ces différentes problématiques ont largement été abordées dans le cadre de ce stage.

L'un des autres éléments prépondérant dans le cadre de cette étude a été le développement de mes compétences en langage de programmation. J'ai notamment pu apprendre de nombreux nouveaux outils en Python grâce à ce stage. La prise en main de nouvelles bibliothèques et l'utilisation de nouveaux types de données sur Python ont été très enrichissantes. Ces compétences constituent à mon sens une vraie valeur ajoutée pour ma future carrière de géomaticien.

Dans le cadre de ce stage, j'ai également eu l'occasion d'approfondir mes compétences en cartographies et traitements des données spatialisées. En outre, j'ai eu l'opportunité d'appréhender de nouveaux outils de SIG.

Ce stage m'a également permis de travailler de manière autonome, en prenant des responsabilités. Ce fut notamment le cas pour le choix des indices climatiques ou la structuration des différents traitements sur Python. Cette autonomie m'a permis d'acquies de l'expérience d'un point de vue professionnel et humain.

Tout en ayant une autonomie importante, j'ai également pu développer des compétences dans le travail en équipe. La collaboration avec un deuxième stagiaire a été très enrichissante et très bénéfique pour mon propre travail.

Enfin, dans le cadre de ce stage, j'ai eu l'opportunité de rencontrer des professionnels de tous milieux (hydrologues, agronomes, gestionnaire de projet, modélisateurs, chercheurs, informaticiens...). Ces échanges m'ont permis d'acquies de nouvelles connaissances sur de nombreuses thématiques. Elles m'ont également offert la possibilité d'observer les méthodes de travail de ces différents corps de métiers.

Table des figures

Figure 1 : Cumul interannuel moyen de précipitations (1979-2013) sur l'ensemble de l'Afrique de l'Ouest.....	9
Figure 2 : Projection des évolutions de températures et de précipitations en fonction des scenarii climatiques.....	11
Figure 3 : Carte de localisation de la zone d'étude.....	12
Figure 4 : Diagramme de gantt.....	15
Figure 5 : Exemple de fichier NetCDF pour les données de précipitations WFDEI.....	17
Figure 6 : Inventaire des stations potentiellement disponibles dans la base de données Hydrosciences/Sierem.....	19
Figure 7 : Nombre de stations par zones administratives de niveau 2 pour chacun des pays de la zone d'étude.....	20
Figure 8 : Répartition des stations potentielles par zones climatiques.....	21
Figure 9 : Part de données manquantes pour les stations fournies par la base de données Hydrosciences/Sierem.....	22
Figure 10 : Nombre de stations par maille et nombre d'années en commun entre les stations et WFDEI pour les stations avec moins de 30% de données manquantes.....	24
Figure 11 : Exemple d'un dictionnaire pour la station de Dakar-Yoff au Sénégal.....	28
Figure 12 : Exemple d'appel de données stockées dans un dictionnaire.....	29
Figure 13 : Exemple d'utilisation de l'outil CDO pour le calcul de l'indice « Somme annuelle ».....	29
Figure 14 : Exemple d'utilisation de l'outil CDO avec deux opérateurs (indice R50 mm)....	30
Figure 15 : Structure des données de la base Hydrosciences/Sierem.....	31
Figure 16 : Calcul des indices climatiques sur Python.....	32
Figure 17 : Chaîne de traitement pour le calcul des indices climatiques pour les stations météorologiques.....	33
Figure 18 : Chaîne de traitement relative aux données maillées (NetCDF) sorties de WFDEI	34
Figure 19 : Chaîne de traitement pour l'automatisation des sorties graphiques.....	35
Figure 20 : Exemple de sorties graphiques pour plusieurs stations et plusieurs indices climatiques.....	37
Figure 21 : RRMSE moyen pour six indices climatiques.....	41
Figure 22 : RRMSE par zones de pluviométries moyennes annuelles avec et sans données manquantes pour six indices climatiques.....	44
Figure 23 : Localisation et caractéristiques des stations composant la maille d'étude spécifique.....	45
Figure 24 : Somme annuelle de précipitations moyennées pour quatre stations localisées dans une même maille WFDEI.....	46
Figure 25 : Somme annuelle de précipitations pour les quatre stations localisées dans une même maille WFDEI.....	47

Figure 26 : Nombre de jours de précipitations par an moyennées pour quatre stations localisées dans une même maille WFDEI.....	48
Figure 27 : Nombre de jours de précipitations supérieure à 20 mm par an moyenné pour quatre stations localisées dans une même maille WFDEI.....	49

Table des tableaux

Tableau 1 : Exemple de variables climatiques obtenues à l'aide de la méthodologie WFDEI	16
Tableau 2 : Variables au format NetCDF potentiellement exploitables pour le calcul des indices climatiques.....	18
Tableau 3 : Part de données manquantes par pays et par seuils pour les 243 stations fournies	23
Tableau 4 : Indices climatiques calculés dans le cadre de ce projet.....	25
Tableau 5 : Bibliothèques python utilisées pour les indices climatiques.....	27
Tableau 6 : Répartition des stations afin de calculer le RMSE moyen par zone climatique. .	38
Tableau 7 : RMSE sur 173 stations et pour 16 indices climatiques.....	39
Tableau 8 : RRMSE par zones climatiques pour six indices.....	42
Tableau 9 : RRMSE par zones climatiques pour les stations avec une part de données manquantes inférieure ou égale à 5 % pour six indices.....	42
Tableau 10 : Différence entre les RRMSE pour les stations avec une part de données manquantes inférieure ou égale à 5 % et les RRMSE pour l'ensemble des 173 stations.....	43

Table des matières

Résumé.....	3
Sommaire.....	4
Remerciements.....	5
I Introduction.....	6
II La structure d'accueil.....	7
1) Institut de Recherche pour le Développement (IRD).....	7
2) Maison de la Télédétection (MTD).....	7
3) Espace DEV.....	7
III Contexte d'étude.....	8
1) Le projet Agricora.....	8
2) Vulnérabilités et variabilités climatiques.....	8
2.1) Le contexte climatique de la zone d'étude.....	8
2.2) Vers un bouleversement climatique.....	10
2.3) Des modèles climatiques.....	11
3) Délimitation de la zone d'étude.....	12
IV Problématique objectifs et organisation du stage.....	13
1) Problématique.....	13
2) Objectifs.....	13
3) Gestion du stage.....	14
3.1) Organisation.....	14
3.2) Prévisionnel.....	14
V Données et méthodes.....	16
1) Les données disponibles.....	16
1.1) Les données fournies par les études climatiques.....	16
1.1.1) Le projet WATCH.....	16
1.1.2) Le format NetCDF.....	16
1.1.3) Description des données WFDEI.....	18
1.2) Données issues de stations météorologiques.....	18
1.2.1) Répartition spatiale de l'ensemble des stations potentiellement exploitables....	18
1.2.2) Répartition spatiale des stations potentiels par pays et par zones climatiques....	19
1.2.3) Filtrage des stations météorologiques.....	21
1.3) Données des stations et maille WFDEI.....	23
2) Outils et Méthodologies.....	25

2.1) Choix et calculs d'indices climatiques.....	25
2.2) Les outils utilisés.....	26
2.2.1) Python et ses librairies.....	27
2.2.2) CDO et NCVIEW pour les fichiers NetCDF.....	29
2.2.3) QGIS pour la cartographie.....	30
2.3) Chaîne de traitement pour les stations météorologiques.....	30
2.4) Chaîne de traitement pour WFDEI.....	34
2.5) Comparaison des données de stations et des données WFDEI.....	34
2.5.1) Automatisation des sorties graphiques.....	34
2.5.2) Calcul d'opérateurs statistiques.....	36
VI Résultats.....	37
1) Sorties graphiques.....	37
2) Calcul du RMSE.....	38
3) Étude comparative des indices climatiques à l'aide du RRMSE.....	40
3.1) Résultats du RRMSE pour six indices climatiques.....	40
3.2) Résultats du RRMSE par zones climatiques.....	41
3.3) Étude de l'influence des données manquantes.....	42
3.4) Synthèse des résultats des RRMSE.....	43
4) Étude de cas intra-maille.....	45
4.1) Choix de la maille d'étude spécifique.....	45
4.2) Étude de cas pour l'indice Somme Annuelle.....	46
4.3) Étude de cas pour l'indice nombre de jours de précipitations supérieure à 1mm/ans.....	47
4.4) Étude de cas pour l'indice nombre de jours de précipitations supérieure à 20 mm/ans.....	48
.....	48
VII Limites et perspectives.....	50
VIII Conclusion.....	52
Bilan personnel.....	53
Table des figures.....	54
Table des tableaux.....	55
Table des matières.....	56
Annexes.....	58
Bibliographie.....	65
Webographie.....	67