



COMITE PERMANENT INTER-ETATS DE LUTTE CONTRE LA SECHERESSE DANS LE SAHEL  
PERMANENT INTERSTATE COMMITTEE FOR DROUGHT CONTROL IN THE SAHEL  
COMITÉ PERMANENTE INTER-ESTADOS DE LUTA CONTRA A SECA NO SAHEL  
اللجنة الدائمة المشتركة لمحاربة التصحر في الساحل



**Centre Régional AGRHYMET**

## **DÉPARTEMENT FORMATION RECHERCHE**

RAPPORT DE STAGE DE 2<sup>ème</sup> ANNEE CYCLE INGENIEUR EN AGROMETEOROLOGIE

PROMOTION : 2016 – 2019

PRÉSENTÉ PAR : **MANDELA C. M. HOUNGNIBO**

---

### **MERGING DES DONNÉES D'OBSERVATIONS SATELLITAIRES ET IN-SITU POUR DES APPLICATIONS AGROCLIMATIQUES AU BENIN**

---

Période : 03 Avril-20 Juin 2018

Lieux de stage : MÉTÉO BENIN

Soutenu le.....devant le jury composé de :

PRÉSIDENT : .....

MEMBRE : .....

MAÎTRE DE STAGE : ING. HOUETO V. ARNAUD

## Dédicace

À

Niels Hendrik Noé Kayodé HOUNGNIBO

# Remerciements

Je rends tout d’abord grâce à Dieu Tout Puissant pour ses immenses bienfaits dans ma vie. Que son nom soit glorifié à jamais !

J’exprime ma gratitude à tous ceux qui ont œuvré directement ou indirectement à la réalisation de ce travail.

Je remercie très chaleureusement :

- ♥ **Kokou Marcellin NAKPON**, Directeur Général de METEO BENIN pour sa diligence et ses encouragements ;
- ♥ **Arnaud HOUETO** mon maître de stage pour sa disponibilité et ses pertinentes remarques sur ce travail ;
- ♥ **Emmanuel LAWIN** et **Abdou ALI** qui m’ont enseigné les bases de la géostatistique, élément très important pour mener à bien ce travail ;
- ♥ **Valérie SOUNOUKE, Robelec AVOUNYRANKOUN, Joseph BESSOU, Luc BIO YARA** pour leur soutien durant toute la période de stage.

J’exprime ma gratitude à tout le personnel du Département Formation Recherche du Centre Régional AGRHYMET et au personnel de MÉTÉO BENIN.

Toute ma reconnaissance au chef Département Formation Recherche **Pr. ATTA SANOUSSI** et à mon chef Filière **Dr. WAONGO Moussa**.

En dernier lieu, j’adresse mes sincères remerciements à mes camarades de classe et spécialement **Essoninam PASSIKE**.

A ma petite famille !

# Table des matières

<b>Avant-propos</b>	<b>viii</b>
<b>Résumé</b>	<b>ix</b>
<b>Abstract</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 État des connaissances</b>	<b>3</b>
<b>3 Matériel et Méthodes</b>	<b>6</b>
3.1 Zone d'étude . . . . .	6
3.2 Données . . . . .	7
3.2.1 Données Satellitaires . . . . .	7
3.2.2 Données climatiques in-situ . . . . .	8
3.3 Méthodes . . . . .	10
3.3.1 Contrôle de qualité des données in-situ . . . . .	10
3.3.2 Méthodologie pour la fusion des données . . . . .	15
3.3.3 Evaluation des données fusionnées . . . . .	17
3.4 Outils . . . . .	18
<b>4 Résultats et interprétations</b>	<b>19</b>
4.1 Contrôle de qualité des données . . . . .	19
4.2 Données mergées . . . . .	21
4.2.1 Réduction d'échelle des données de températures . . . . .	21
4.2.2 Correction de biais des données satellitaires brutes . . . . .	22
4.2.3 Données mergées de pluie . . . . .	24
4.2.4 Données mergées de températures . . . . .	25
4.3 Évaluation des données mergées . . . . .	26
<b>5 Discussions</b>	<b>29</b>
<b>6 Applications agrométéorologiques</b>	<b>30</b>
6.1 Résumé décadaire et mensuel de températures et de pluie . . . . .	30
6.2 Quelques paramètres agroclimatiques . . . . .	31
6.3 Séries temporelles de la pluviométrie dans une localité dépourvue de station . . .	32
<b>7 Conclusion</b>	<b>34</b>
<b>Références</b>	<b>35</b>

<b>Annexes</b>	<b>I</b>
Annexes A : Détails sur la détermination du facteur multiplicateur f . . . . .	I
Annexes B : Distribution théorique à laquelle nos séries de paramètres climatiques s'ajustent le mieux . . . . .	II
Annexes C : Extrait des résultats de contrôle de qualité . . . . .	IV
Annexes D : Extraits des résultats de correction de biais des données satellitaires brutes	V
Annexes E : Variogramme . . . . .	VI

## Liste des figures

1	Carte du réseau météorologique utilisé . . . . .	6
2	Régime pluviométrique et de températures des données in-situ et satellitaires (TAMSAT, JAR-55) sur la période 1983 – 2010 . . . . .	7
3	Évolution du nombre de stations relevant la pluie et la température au fil des années	8
4	Pourcentage de données manquantes de pluie quotidienne par mois sur la période 1981 – 2017 . . . . .	9
5	Pourcentage de données manquantes de températures par mois sur la période 1981 – 2017 . . . . .	10
6	Réseau pluviométrique opérationnel (points en noirs). Réseau global (points noirs et les croix). Les deux rectangles pour l'évaluation des produits mergés dans les zones pauvres en station (Z1) et dans les zones riches en station (Z2). . . . .	18
7	Températures maximales des mois de février (1981 – 2017) de Dassari- En rouge celle du 08 février 1998 . . . . .	19
8	Températures maximales des mois d'avril (1981 – 2017) à Cotonou- En rouge celle du 27 Avril 2008 . . . . .	20
9	Précipitation isolée à Dangbo (rouge) le 05/10/1999. . . . .	20
10	Sécheresse isolée à Dogbo (rouge) le 08/07/2000. . . . .	20
11	Forte déviation au dessus à Adjohoun (rouge) le 10/08/1987. . . . .	21
12	Forte déviation au dessus à Tanguiéta (rouge) le 30/08/1996. . . . .	21
13	Reggridding des données de températures du 01/05/1997. Observation (à gauche), Satellitaire brute de résolution 50 km (au milieu), Satellitaire brute de résolution 4 km (à droite). . . . .	22
14	Correction de biais des données satellitaires brutes de températures. Observation (à gauche), Satellitaire brute (au milieu), Satellitaire corrigé de biais (à droite). . .	23
15	Correction de biais des données satellitaires brutes de pluie (11/06/2017). Observation (à gauche), Satellitaire brute (au milieu), Satellitaire corrigé de biais (à droite). . . . .	23
16	Données mergées de pluie pour le 11/06/2017. Observation (en haut à gauche), satellitaire brute (en haut à droite), observation interpolée (en bas à gauche), mergée (en bas à droite). . . . .	24
17	Données mergées de températures pour le 01/05/1997. Observation (en haut et à gauche), satellitaire brute (en haut et à droite), observation interpolée (en bas et à gauche), mergée (en bas et à droite). . . . .	26
18	Diagramme de Taylor des séries de températures observées, satellitaires brutes et mergées . . . . .	27
19	Diagramme de Taylor des séries de pluies observées, satellitaires brutes et mergées	28
20	Cumul de la 3 <sup>eme</sup> décade de pluie du mois de juin 2010. Observation (à gauche), Mergé(à droite). . . . .	30

21	Moyenne mensuelle des températures maximales du mois de mai 1997. Observation (à gauche), Mergé(à droite) . . . . .	31
22	Moyenne mensuelle des températures minimales du mois de mai 1997. Observation (à gauche), Mergé(à droite) . . . . .	31
23	Evapotranspiration potentielle pour la journée du 03/06/2006. Observation (à gauche), Mergé (à droite).) . . . . .	31
24	Date de début de saison pour l'année 1997. Observation (à gauche), Mergé (à droite). . . . .	31
25	Régime décadaire de pluie à Gogounou et à Kandi (Période 1983 – 2015) . . . .	32
26	Varibilité interannuelle des pluies à Gogounou. . . . .	33

## Liste des tableaux

2	Pourcentage des valeurs aberrantes par mois sur l'ensemble des stations . . . . .	21
3	Statistiques d'évaluation des données fusionnées par validation croisée . . . . .	27

## Sigles-Acronymes-Notations et Abréviations

<b>2DVAR</b>	Assimilation des Données Variationnelles Bidimensionnelles.
<b>CDF</b>	Fonction de Distribution Cumulative.
<b>IRI</b>	International Research Institute for Climate and Society.
<b>JAR-55</b>	Japanese 55—year Reanalysis.
<b>MESA</b>	Monitoring For Environment and Security in Africa.
<b>MÉTÉO BENIN</b>	Agence Nationale de la Météorologie du Bénin.
<b>NOAA</b>	National Oceanic and Atmospheric Administration.
<b>OMM</b>	Organisation Météorologique Mondiale.
<b>Qmap</b>	Quantile Mapping.
<b>QQ</b>	Quantile-Quantile.
<b>RK</b>	Regression—Kriging.
<b>RMSE</b>	Erreur Quadratique Moyenne.
<b>SMT</b>	Système Mondial de Télécommunication.
<b>TAMSAT</b>	Tropical Applications of Meteorology using SATellite data and ground-based observations.
<b>TSO</b>	Température de Surface des Océans.



## **Avant-propos**

Le présent rapport rentre dans le cadre de notre stage de fin de deuxième année du cycle Ingénieur en agrométéorologie, que nous avons effectué à l'Agence Nationale de la Météorologie du Bénin (MÉTÉO BENIN). Au cours de ce stage, nous nous sommes impliqués dans les différentes activités de l'agence, ce qui nous a permis de nous imprégner des réelles difficultés que rencontrent les techniciens.

L'idée de cette étude vient de l'une des difficultés que rencontrent les services agrométéorologiques et climatologiques : la disponibilité de données météorologiques à une échelle spatiale assez fine lors de l'élaboration de produits agrométéorologiques et climatiques fiables. Cette étude se veut être une contribution proposant une alternative à ce problème de fluidité du réseau météorologique. Ainsi il a été combiné les données satellitaires estimées avec les données des stations pour tirer avantages des deux types d'observations. Les données satellitaires pour leurs bonnes résolutions spatiales et les données in-situ pour leur représentativité du climat local. Des propositions d'applications agrométéorologiques ont été faites pour montrer l'évidence de l'amélioration apportée par la combinaison des deux types de données.

Des difficultés n'ont pas manqué. Elles concernent particulièrement la disponibilité de ressources matérielles pour la validation croisée. Cette difficulté ne nous a pas permis d'évaluer les interpolations des données d'observations afin de faire une comparaison avec les statistiques d'évaluation des données mergées.

## Résumé

La gestion des risques climatiques nécessite la disponibilité de données météorologiques. La disponibilité de ces informations à travers toutes les contrées Béninoises reste un défi majeur pour l'Agence nationale de la météorologie du Bénin, du fait de la fluidité de son réseau d'observations météorologiques. Il est proposé dans cette étude, le merging des données in-situ et des données de réanalyses ou d'estimations satellitaires afin de bonifier les avantages des deux types de données pour pallier au problème de disponibilité de données. Les données in-situ ont été fusionnées avec les estimations satellitaires TAMSAT pour les précipitations et les données de réanalyses JAR-55 pour les températures. Pour ce faire, les données in-situ ont subi une procédure rigoureuse de contrôle de qualité. Ensuite, les estimations satellitaires et les données de réanalyses ont été corrigées de leurs biais par rapport à l'observation. Enfin, la fusion a été réalisée en utilisant la méthode du regression-kriging. Les données mergées ont été évaluées par validation croisée. L'analyse qualitative montre qu'il n'y a pas de différence significative dans les régions riches en postes pluviométriques, entre les données observées interpolées et les données mergées. Cependant dans les zones moins denses en postes pluviométriques les données mergées voire celles satellitaires brutes et satellitaires corrigées de biais montrent de meilleures performances que les interpolations de l'observation. Il est aussi constaté qu'un réseau optimal pourrait être défini pour atteindre la meilleure des performances possibles pour le merging. Les données mergées de températures reproduisent la plupart des structures spatiales de températures qui se trouvent sur l'observation et ressortent de plus les particularités géographiques liées à la température. Les produits mergés évalués par validation croisée, présentent, en général, de bonnes performances.

**Mots clés :** risques climatiques, merging, TAMSAT, JAR-55, validation croisée, regression-kriging, interpolées, Bénin.

## Abstract

Climate risk management requires the availability of meteorological data. The availability of this information throughout rural area of Benin remains a major challenge for Benin Meteorological Agency, due to the fluidity of its observation's network. This study proposes the merging of in-situ data and reanalysis or satellite estimation data in order to improve the advantages of both types of data to overcome the problem of data availability. In-situ data has been merged with satellite estimates TAMSAT for precipitation and reanalysis data JAR-55 for temperatures. To do this, the in-situ data has undergone a rigorous quality control procedure. And then, the satellite estimates and the reanalysis data were corrected for their biases with respect to the observations. Finally, the merging was achieved by regression-kriging. The merged data were evaluated by cross-validation. The qualitative analysis shows that there is no significant difference in rain gauge rich regions, between the gridded rain gauge and the combined satellite-gauge rainfall product. However, in areas with less rain gauge, the merged data, even satellite raw data and corrected satellite data, show better performance than the gridded rain gauge products. It is also noted that an optimal network could be defined to achieve the best possible performance for merging. The merged temperature data reproduce most of the spatial temperature structures that are found on the observation and also highlight the geographical particularities related to the temperature. Merged products evaluated by cross-validation, in general, show good skills.

**Key words:** climate risk, merging, TAMSAT, JAR-55, cross-validation, regression-kriging, gridded rain gauge, Benin.

# 1 Introduction

La disponibilité d'informations climatiques pertinentes pour la prise de décision, au niveau national et local, est essentielle pour évaluer les menaces et les opportunités climatiques et pour intégrer les questions climatiques dans les pratiques de développement. Ainsi, des séries historiques de paramètres climatiques avec une bonne couverture spatiale sont nécessaires dans l'élaboration de ces produits climatiques décisionnels (Dinku et al., 2014b). Cependant, la disponibilité de données climatiques fiables et à des échelles spatiales fines reste une contrainte majeure en Afrique pour la fourniture de produits climatiques de qualité (Hansen et al., 2006; Dinku et al., 2014a).

Au Bénin, le réseau météorologique national géré par l'Agence Nationale de la Météorologie du Bénin (MÉTÉO BENIN) présente une faible densité et ne respecte pas les normes de l'Organisation Météorologique Mondiale (OMM) pour la plupart des paramètres (Midingoyi, 2016). Les stations opérationnelles sont inégalement réparties et principalement situées dans les chefs lieu des communes. En outre, la transmission en temps réel et les lacunes dans les données climatiques historiques rendent difficiles les analyses climatiques.

Pour surmonter ces difficultés, les techniciens de MÉTÉO BENIN ont recours à des données satellitaires notamment les données estimées de pluie et de réanalyses disponibles sur la plateforme du projet Monitoring For Environment and Security in Africa (MESA) afin d'informer le public. La principale force des produits satellitaires est qu'ils offrent une bonne résolution spatiale. Cependant, ces produits ont leurs propres faiblesses, qui sont entre autres une courte série historique, des résolutions temporelles grossières, une inhomogénéité des séries chronologiques et une précision médiocre (Dinku et al., 2014b).

Au regard de ces observations il serait intéressant de combiner ces deux types de données (les données satellitaires et les données stations) en vue d'optimiser leurs avantages pour produire des séries historiques de paramètres climatiques de bonne qualité et de résolution spatiale fine. De nombreux produits développés par les grands centres tels que la National Oceanic and Atmospheric Administration (NOAA), l'International Research Institute for Climate and Society (IRI)... prennent déjà en compte des données au sol pour non seulement corriger les biais des données satellitaires mais aussi fusionner les observations au sol avec les estimations satellitaires. Toutefois, en Afrique, seules les observations synoptiques obtenues par le biais du Système Mondial de Télécommunication (SMT) sont généralement disponibles pour ces centres dans l'élaboration de ces produits (Dinku et al., 2014b). Pour le Bénin par exemple, uniquement six (6) stations sont accessibles via le SMT tandis que le réseau météorologique du Bénin comporte en plus des stations synoptiques, plus de quatre-vingt-dix (90) postes pluviométriques et plus d'une quinzaine de stations climatiques qui pourraient améliorer ce processus de combinaison. C'est fort de ce constat qu'a été réalisée cette étude intitulée : **merging des données d'observations satellitaires et in-situ pour des applications agroclimatiques.**

L'objectif global de cette étude est de produire des données climatiques historiques de température et de pluviométrie de bonne qualité et de résolution spatiale assez fine afin d'élaborer des informations agroclimatiques pertinentes pour la prise de décision. Plus spécifiquement, l'étude vise à :

- ✍️ contrôler la qualité des données d'observations au sol ;
- ✍️ corriger le biais des données satellitaires par rapport aux données in-situ ;
- ✍️ combiner les données satellitaires et les données au sol ;
- ✍️ évaluer les données mergées ;
- ✍️ proposer quelques applications agroclimatiques des données fusionnées.

L'essentiel de ce travail est résumé dans les lignes à suivre et est structuré en quatre (04) parties :

- ✍️ la première partie est consacrée à la revue de littérature sur le thème ;
- ✍️ la deuxième est dédiée aux méthodes et matériels utilisés dans le cadre de l'étude. Une importance particulière a été accordée au contrôle de qualité des données d'observations pour atteindre l'objectif de l'étude. Les méthodes utilisées pour la fusion des données ainsi que pour la validation des données fusionnées ont aussi été élucidées ;
- ✍️ les résultats et les discussions sont présentés respectivement dans la troisième et quatrième partie. Il est exposé dans un premier temps quelques résultats sur la qualité des données, ensuite sur la fusion des données et l'évaluation des données fusionnées et enfin des propositions d'applications agroclimatiques de ces données mergées ont été faites ;
- ✍️ finalement, il est fait une synthèse des principaux résultats qui découlent de ce travail avec quelques perspectives de recherche dans la dernière partie : la conclusion.

## 2 État des connaissances

Il est nécessaire avant d'entamer cette étude, de faire l'état de l'art sur la problématique de fusion des données d'observations satellitaires et in-situ. Une analyse critique sera faite sur les différentes méthodes utilisées afin de montrer leurs performances et leurs plausibles lacunes.

Dès l'avènement des observations satellitaires, bon nombre de chercheurs se sont attelés au développement d'algorithmes performants pour fournir des estimations satellitaires des principales variables climatiques et procéder à leurs validation par rapport aux observations classiques (L. Bernstein, 1982; Merchant et al., 2009; Petrenko et al., 2014). Du reste, les avancées scientifiques et technologiques ont permis l'amélioration au fil des années des algorithmes d'estimations. Par ailleurs, d'une façon concomitante, d'autres chercheurs ont pensé à l'amélioration de la qualité de ces produits estimés en les combinant avec les observations in-situ (Reynolds, 1988; Reynolds & Smith, 1994; Chao et al., 2009). La fusion des données satellitaires et in-situ est donc une question de recherches nouvelles qui passionnent bon nombre de chercheurs.

Plusieurs approches existent pour combiner différentes sources de données. L'étude de Reynolds (1988) sur la Température de Surface des Océans (TSO) a été l'un des travaux précurseurs dans le domaine de la combinaison des données climatiques. La méthodologie utilisée par l'auteur a consisté dans un premier temps sur les données d'observations historiques in-situ (bateaux et bouets météorologiques) à :

- contrôler leur qualité pour éliminer les valeurs aberrantes ;
- faire une interpolation spatiale ;
- appliquer un filtre spatial non linéaire basé sur la médiane développé par Tukey (1974) cité par Reynolds (1988) ;
- remplacer les valeurs médianes par les valeurs interpolées dans les régions à forte densité de mesures d'observations ;
- appliquer un lissage linéaire utilisant un filtrage binomial à deux dimensions.

Ensuite, sur les données satellitaires à :

- déterminer leurs biais par rapport à l'observation ;
- corriger ces biais par une méthode de régression ;
- appliquer un lissage linéaire utilisant un filtrage binomial à deux dimensions.

Enfin, la combinaison des deux types de données a été réalisée en utilisant, les observations in-situ pour définir les valeurs de température de référence dans les régions à forte densité de données in-situ et les observations satellitaires pour définir la forme du champ dans les régions avec peu ou pas de données in-situ. Ceci est fait en imposant que le champ de TSO satisfait à l'équation de Poisson (Oort & Rasmuson, 1971) cité par (Reynolds, 1988).

Cette méthode est une correction objective des biais satellitaires par rapport aux données in-situ. Toutefois les techniques utilisées dégradent la résolution spatiale du produit final à environ 6° de résolution (Reynolds & Smith, 1994).

Pour mieux préserver la résolution des données satellitaires Reynolds & Smith (1994) ont développé une technique d'analyse utilisant les données combinées selon la méthodologie de Reynolds (1988) décrite plus haut pour produire des données préliminaires satellitaires corrigées de faible résolution qui seront associées aux données in-situ et traitées par une méthode d'interpolation optimale sur une résolution de 1°. Cette technique conserve la correction des biais tout en améliorant la résolution spatiale des données fusionnées (Reynolds & Smith, 1994).

Good (2015), a utilisé cette méthode d'interpolation optimale pour combiner les données de températures de l'air estimées avec les données de température de l'air in-situ afin de fournir une série de données avec une représentation spatiale améliorée. Les données de température de l'air estimées ont été générées par un modèle de régression multilinéaire construit et validé qui utilise comme variables explicatives les données satellitaires de températures quotidiennes minimales et maximales, la fraction de végétation, la distance du point de grille à la côte, la latitude, l'élévation et la fraction urbaine.

Chao et al. (2009) ont indiqué qu'il serait souhaitable de formaliser mathématiquement l'algorithme d'interpolation optimale et d'améliorer son efficacité de calcul pour satisfaire aux exigences d'assimilation d'un grand nombre d'observations et de prise en compte des caractères inhomogènes et anisotropes des données. Pour pallier ces insuffisances, ces auteurs ont développé l'algorithme d'Assimilation des Données Variationnelles Bidimensionnelles (2DVAR) qui utilise un principe mathématique identique à la méthode d'optimisation optimale, mais sa mise en œuvre numérique est différente. La différence majeure est que le 2DVAR tire parti des méthodes avancées d'optimisation numérique à grande échelle.

L'étude de Dinku et al. (2014b) est un cas typique de combinaison de pluie estimée par satellite et de hauteur de pluie in-situ. Cette étude, réalisée sur l'Éthiopie à une échelle décadaire, a utilisé la méthodologie suivante :

- le contrôle de la cohérence temporelle et spatiale des données ainsi que des coordonnées des stations ;
- la calibration des paramètres de l'algorithme d'estimations des pluies satellitaires sur l'Éthiopie ;
- la fusion des deux types de données par deux méthodes : l'ajustement de biais et la méthode du regression-kriging (RK) (Hengl et al., 2007).

D'autres études en l'occurrence celle de Doan et al. (2014) et de Manz et al. (2016) ont utilisé pour la fusion des données de précipitations des méthodes d'interpolation allant des déterministes simples (la pondération inverse de distance) aux statistiques et géostatistiques complexes (la régression linéaire simple, krigéage, co-krigéage et le krigéage avec dérive externe...).

D'une façon générale, les méthodologies adoptées par la plupart des auteurs consistent au contrôle de qualité des données in-situ et à l'utilisation d'approches géostatistiques du plus simple au plus complexe pour la combinaison des deux types de données. Toutefois la pertinence des approches géostatistiques dépendra de la densité des réseaux in-situ et de la résolution des données satellitaires. En effet, lorsque les données sont suffisamment denses, la plupart des techniques d'interpolation donnent des résultats presque identiques. Des problèmes surviennent lorsque les données deviennent clairsemées et irrégulièrement distribuées ([Smith et al., 1998](#)). Pour une étude assez approfondie, Il est légitime de diversifier les méthodes géostatistiques et de voir dans la validation celle qui s'approprie le mieux à la réalité observée. Toutefois, vu le temps imparti pour ce travail, seule la correction de biais et la Regression-Kriging seront utilisées pour la fusion des deux types de données.



### 3 Matériel et Méthodes

#### 3.1 Zone d'étude

Le Bénin est situé en Afrique de l'Ouest et est limité au Nord par le fleuve Niger qui le sépare de la République du Niger ; au Nord-Ouest par le Burkina Faso, à l'Ouest par le Togo, à l'Est par le Nigeria et au Sud par l'Océan-Atlantique (Figure 1). Le pays couvre une superficie de 112622 km<sup>2</sup> ; il s'étend sur 670 km, du fleuve Niger à la côte atlantique qui est longue de 121 km. Le Bénin offre un relief peu accidenté. Le point culminant du pays (658 m) se situe sur la chaîne de l'Atacora au nord-ouest du pays. Le réseau hydrographique du pays comprend 3048 km de cours d'eau et 333 km<sup>2</sup> de plans d'eau (lacs et lagunes) (PNUD, 2009).

De par sa position géographique, le Bénin est sous l'influence des interactions de la mousson et de l'harmattan qui régissent son climat. Ainsi le sud du pays connaît un climat tropical humide (un régime pluviométrique bimodal) et le Nord un climat tropical sec (une unique saison pluvieuse). Au regard des données prises à MÉTÉO BENIN et en tenant compte de la période de référence 1981 – 2010, la pluviométrie moyenne annuelle varie de 700 mm à l'extrême Nord à plus de 1300 mm au Sud. La moyenne annuelle de la température de l'air oscille autour de 27°C, le maximum absolu pouvant dépasser 45°C au Nord. Les températures minimales moyennes varient entre 21,6°C (Nord) à 23,6°C (Sud) tandis que les températures maximales moyennes fluctuent entre 33,8°C (Nord) à 32,2°C (Sud). Il convient de souligner que le Nord-ouest bénéficie d'un climat Atacorien caractérisé par des températures plus fraîches à cause de l'altitude et des orages plus fréquents.

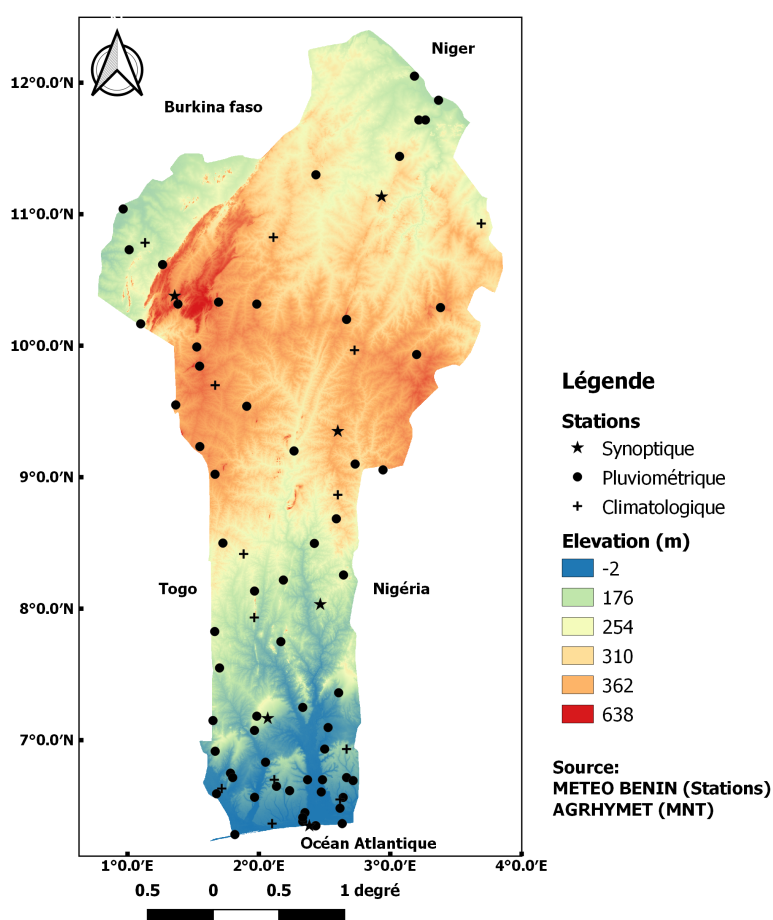


FIGURE 1 – Carte du réseau météorologique utilisé

## 3.2 Données

### 3.2.1 Données Satellitaires

Plusieurs sources de données satellitaires sont disponibles. Cependant, il est constaté que bon nombre d'entre elles n'ont pas une série historique assez conséquente et/ou ne disposent pas d'une résolution spatiale et temporelle assez fine pour des analyses agroclimatiques. D'autres même s'ils présentent tous ces atouts ne reproduisent pas au mieux la climatologie zonale du pays. C'est pourquoi pour ce travail, il a été choisi d'utiliser les données satellitaires de pluie estimées Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT) ([Tarnavsky et al., 2014](#)) pour la fusion des données de pluie et le Japanese 55-year Reanalysis (JAR-55) ([Ebata et al., 2011](#)) pour la fusion des données de températures. En effet, les données TAMSAT possèdent une assez bonne résolution spatiale (moins de 4 km) et temporelle (quotidienne). Avec une série historique conséquente (1983 – 2017), elles sont aussi disponibles en temps réel pour le suivi agroclimatologique. Les données JAR-55 ont une résolution spatiale d'environ 50 km avec une fréquence journalière et présentent une bonne série historique (1951 – 2015). En outre, ces deux sources de données satellitaires reproduisent au mieux la climatologie en Afrique de l'ouest. Les travaux de [Arnaud et al. \(1996\)](#); [Novella \(2010\)](#); [Lorente-Plazas et al. \(2014\)](#); [Birkel \(2015\)](#) en disent long.

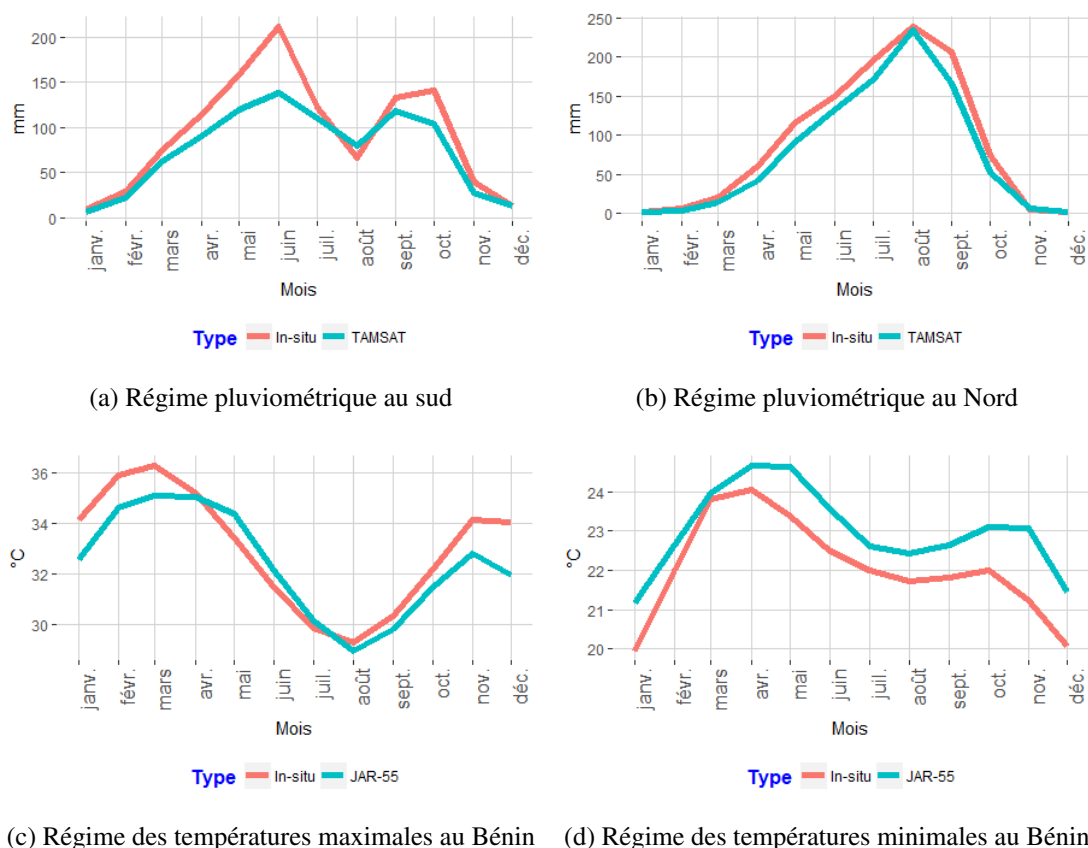


FIGURE 2 – Régime pluviométrique et de températures des données in-situ et satellitaires (TAMSAT, JAR-55) sur la période 1983 – 2010

Nous nous sommes exercés à voir spécifiquement la représentativité de ces données sur le Bénin (figure 2). Les régimes pluviométriques moyens sur la période 1981 – 2010 ont été déterminés pour le sud ( $[1.5^{\circ}\text{-}2.8^{\circ}\text{E}; 6.2^{\circ}\text{-}7.5^{\circ}\text{N}]$ ) et pour le nord du pays ( $[0.8^{\circ}\text{-}3.8^{\circ}\text{E}; 9^{\circ}\text{-}12.8^{\circ}\text{N}]$ ). Pour les températures, les régimes ont été réalisés pour l'ensemble du pays ( $[0.8^{\circ}\text{-}3.8^{\circ}\text{E}; 6.2^{\circ}\text{-}12.8^{\circ}\text{N}]$ ). Les régimes pluviométriques bimodales au sud et unimodales au nord observés au Bénin sont presque bien reproduits par les estimations satellitaires TAMSAT (figure 2a, 2b). Il en est de même pour les données de réanalyses de températures JAR-55 (figure 2c, 2d), malgré qu'elles surestiment. Une surestimation remarquée par Lorente-Plazas et al. (2014).

### 3.2.2 Données climatiques in-situ

Les données exploitées dans cette étude proviennent de la base de données de MÉTÉO BENIN. Le réseau météorologique national est constitué de stations synoptiques, climatiques et des postes pluviométriques. A l'exception des postes pluviométriques où il n'est mesuré que la pluie, sur les autres types de stations, il est mesuré la plupart des paramètres climatiques dont la pluie et la température qui font l'objet de cette étude. Il n'est retenu pour cette étude que les stations du réseau présentant des données historiques conséquentes de pluie et de températures. Ainsi le réseau météorologique utilisé est constitué de 83 postes pluviométriques, 20 stations synoptiques et climatologiques mesurant la température (figure 1). Il est constaté une inégale répartition spatiale des stations avec une concentration relativement forte au sud qu'au nord du pays et une insuffisance des stations dans les zones de relief (figure 1).

Une variation du nombre de stations relevant les variables d'intérêt au fil des années est observée (figure 3). On note une évolution relativement régulière du nombre de poste pluviométrique, montrant l'effort de densification du réseau pluviométrique.

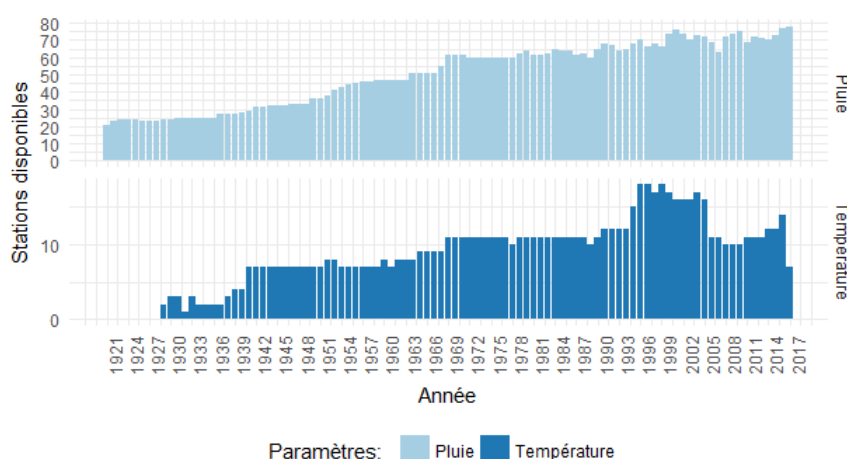
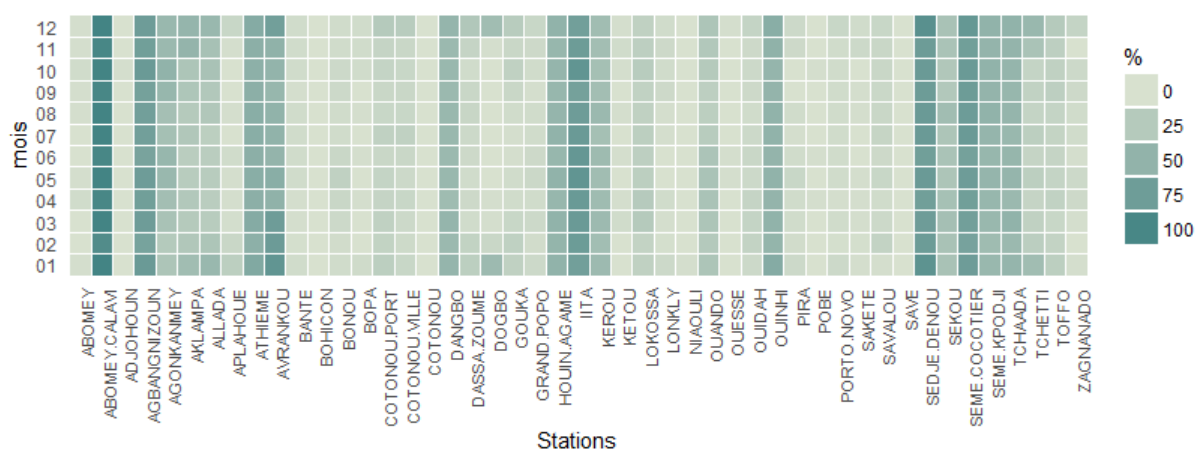


FIGURE 3 – Évolution du nombre de stations relevant la pluie et la température au fil des années

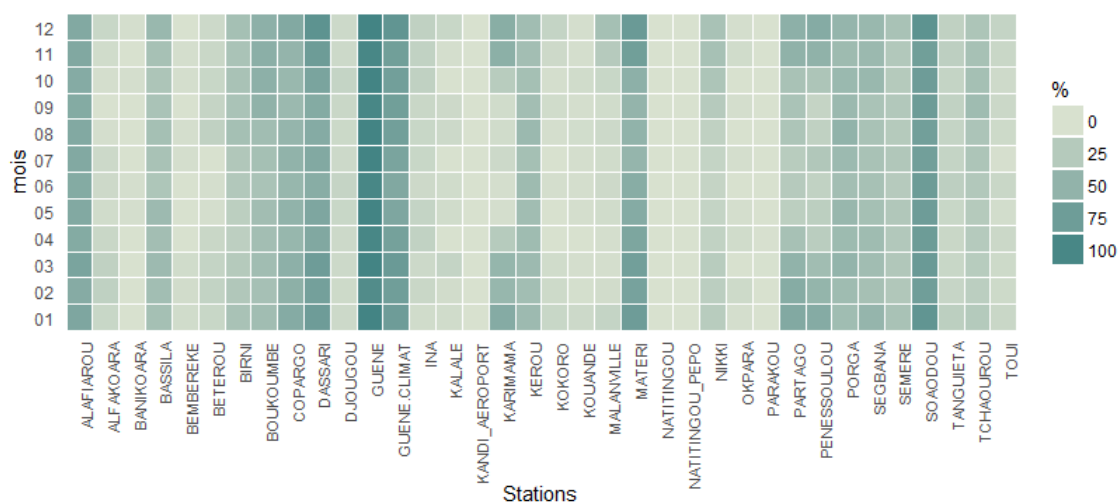
Au début des années soixante-et-dix (Années de sécheresses dans la sous région), plus d'une soixantaine de postes pluviométrique était déjà disponibles pour mesurer la pluie. Pour ce qui concerne les températures, une dizaine de stations étaient déjà disponibles autour des années

soixante-et-dix avec un pic s’observant autour des années quatre vingt-et-dix. Au regard de cette évolution dans la disponibilité des stations d’observations, et vu la série temporelle des données satellitaires (TAMSAT, JAR-55) disponibles, les périodes 1983 – 2017 pour la pluviométrie et 1983 – 2015 pour les températures ont été retenues pour ce travail.

Les données manquantes sur les périodes d’études 1983-2017 pour la pluie et 1983-2015 pour les températures varient en fonction du mois par station (figure 4).



(a) Données manquantes de pluie pour les stations du sud



(b) Données manquantes de pluie pour les stations du nord

FIGURE 4 – Pourcentage de données manquantes de pluie quotidienne par mois sur la période 1981 – 2017

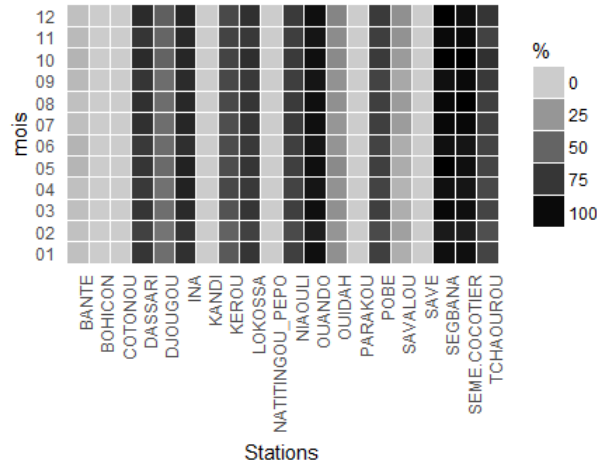


FIGURE 5 – Pourcentage de données manquantes de températures par mois sur la période 1981 – 2017

Sur les 83 postes pluviométriques, sept stations (Abomey-Calavi, Agbagnizoun, IITA, Guéné, Sédjédénou, Sèmè-Cocotier, Soaodou) présentent de forts taux de données manquantes (supérieurs à 75%). Pour l'ensemble des stations, des taux supérieurs à 50% se retrouvent régulièrement au mois de janvier, février novembre et décembre (Saison sèche) (figure 4a, 4b).

Pour ce qui concerne les données de températures, cinq stations (Dassari, Ina, Lokossa, Ouando, Ségbana, Sèmè-Cocotier) présentent des taux extrêmes (supérieurs à 95%) de données manquantes (figure 5).

### 3.3 Méthodes

#### 3.3.1 Contrôle de qualité des données in-situ

Dans l'atteinte des objectifs de cette étude, il est impératif de contrôler la qualité des données afin de détecter et d'identifier les erreurs issues du processus de mesure, de manipulation, de formatage, de transmission et d'archivage des données.

Pour l'analyse des valeurs aberrantes, il a été procédé à la standardisation des différents paramètres. Une méthode familière utilisée à cette fin est le score standardisé ou scores Z (Schulz, 2002). La formule générale pour convertir une variable X, au score standardisé (score Z) est la suivante :

$$Z(X) = \frac{X - \mu_X}{\sigma_X} \quad (1)$$

où  $\mu_X$  est la moyenne et  $\sigma_X$  est l'écart-type de la série de la variable. Le score standardisé permet de savoir à combien d'écart-type une observation se situe de la moyenne. Dans le cas spécifique où X suit une distribution normale, Z(X) a une moyenne de 0 et une variance de 1. Ce score est d'une aide précieuse dans la détermination des valeurs aberrantes.

Une observation  $X_i$  est une valeur aberrante si la valeur absolue du score Z est supérieure à un seuil donné f'(connu sous le nom de facteur multiplicateur) (Seo, 2006).

$$\frac{|X_i - \mu_X|}{\sigma_X} \succ f' \quad (2)$$

Puisque, la moyenne est très sensible aux valeurs extrêmes, [Eischeid et al. \(1995\)](#) ont choisi de standardiser les observations en soustrayant la médiane de l'observation et ont divisé cette différence par l'intervalle interquartile. L'équation (2) devient :

$$\frac{|X_i^j - Q_{50}^j|}{IQR^j} \succ f \quad (3)$$

où  $X_i^j$  la valeur du paramètre climatique pour le jour  $i$  ( $i \in [1, 31]$ ) du mois  $j$  ( $j \in [1, 12]$ ),  $Q_{50}^j$  est la médiane de la série journalière des mois  $j$  sur la période 1981-2010 et  $IQR^j$  l'intervalle interquartile de la série journalière des mois  $j$  sur la période 1981-2010.

En fait, la médiane est considérée comme une mesure robuste par rapport aux valeurs aberrantes, et aux biais dans les distributions asymétriques.

Une valeur typique de  $f$  proposée pour identifier les outliers est 3 (Velleman & Hoaglin, 1981) cité par ([Eischeid et al., 1995](#)).

Dans le cadre de ce travail, le facteur multiplicatif  $f$  a été déterminé en utilisant un test ayant comme hypothèse nulle ( $H_0$ ) la non existence de valeurs aberrantes dans les données. Afin de rejeter l'hypothèse nulle, un degré de risque  $\alpha$  est donné en termes de valeurs critiques et/ou de niveau de confiance. Les détails sur la détermination de  $f$  sont présentés en **annexe A**.  $f$  est ainsi donnée par l'équation :

$$f = \frac{F^{-1}(1 - \alpha)}{2(F^{-1}(0.75) - F^{-1}(0.25))} \quad (4)$$

où  $F^{-1}$  est la Fonction de Distribution Cumulative inverse (ou fonction quantile) pour la série de notre variable climatique, supposée distribuée normalement dans toute la démonstration. Le niveau de confiance choisi pour tout le processus de contrôle est de 95%.

L'utilisation de l'équation (4) a nécessité la détermination des distributions théoriques auxquelles nos séries journalières de pluie et de températures des mois  $j$  ( $j \in [1, 12]$ ) sur la période 1981 – 2010, s'ajustent le mieux. Ainsi pour chaque mois la série journalière sur la période 1981 – 2010 est constituée. Ensuite, nous utilisons le diagramme de Cullen & Frey (1999) cité par [Delignette-Muller & Dutang \(2015\)](#) pour voir les distributions qui s'approchent le mieux de cette série en terme d'asymétrie (Skewness) et d'aplatissement (Kurtosis). La distribution la plus proche étant identifiée, il est enfin recherché ses paramètres par la méthode des moments.

Des courbes de la Fonction de Distribution Cumulative (CDF), Quantile-Quantile (QQ) de la distribution théorique et de la distribution empirique sont réalisées pour explorer la qualité de l'ajustement.

De l'analyse des différents résultats issus du processus, il ressort que la loi gamma offre un meilleur ajustement à nos séries de précipitations et la loi normale à nos séries de températures. Des extraits de résultats obtenus pour les séries de pluie et les séries de températures des mois de

juillet sont présentées en **annexe B**.

#### **a** Contrôle de la qualité des données de précipitation

##### — Cohérence temporelle :

La loi Gamma est l'une des distributions qui offre un ajustement relativement meilleur pour les séries journalières de précipitations des mois sur la période 1981 – 2010 ; elle est bien connue pour sa distribution asymétrique. Afin d'appliquer le principe du score Z décrit ci-dessus, nous devons transformer la variable (précipitation) pour obtenir une variable distribuée selon la loi normale (Schulz, 2002; Krishnamoorthy et al., 2008).

Wilson & Hilferty, (1931) cité par Krishnamoorthy et al. (2008) ont trouvé un moyen de transformer une variable  $\chi^2$  au score Z afin que leurs p-value soient très proches. La première étape de la transformation consiste à transformer la statistique du  $\chi^2$  en une variable plus normalement distribuée. Ils ont montré que la racine  $p^{eme}$  de la variable chi-carré divisée par ses  $n$  degrés de liberté est presque distribuée normalement et que :

si

$$X = \left( \frac{\chi_n^2}{n} \right)^{\frac{1}{p}} \quad (5)$$

alors

$$\sigma_x^2 \approx \left( \frac{1}{p^2} \right) \left( \frac{2}{n} \right) \quad (6)$$

et

$$\mu_x \approx 1 - \sigma_x^2 \quad (7)$$

Wilson & Hilferty, (1931) choisissent  $p = 3$  pour leur transformation. En remplaçant, les équations (5, 6, 7) dans (1), la transformation complète en termes d'une variable chi-carré, Y, de n degrés de liberté est :

$$W(Y) = \frac{\left( \frac{Y}{n} \right)^{1/3} - \left( 1 - \left( \frac{1}{9} \right) \left( \frac{2}{n} \right) \right)}{\sqrt{\left( \frac{1}{9} \right) \left( \frac{2}{n} \right)}} \quad (8)$$

l'équation (8) est la forme de base d'une transformation de normalisation. les p-values de W(Y) sont très proches des p-values des variables normalement distribuées. Si Z est normalement distribué, alors  $P(Z < W(Y)) \approx P(Y < y)$ . Ainsi, W(Y) s'approxime à la statistique de Student et f peut être choisi comme une fonction de  $(1 - \alpha)^{eme}$  quantile de la distribution de Student avec  $(n - 1)$  degrés de liberté.

##### — Cohérence spatiale :

La cohérence spatiale des données de précipitation a été vérifiée en comparant les valeurs des stations voisines à la même date. Les seuils sont fixés pour déterminer si une valeur d'observation présente une trop grande déviation par rapport aux valeurs des stations voisines. Le principe du score Z décrit ci-dessus est également appliqué pour détecter les valeurs fortement déviées. Scherrer et al. (2010), ont fourni 3 cas d'in vraisemblance pour

la vérification spatiale des hauteurs quotiennes de pluie : les précipitations isolées, les sécheresses isolées, les trop grandes déviations.

**Précipitation isolée :** Une valeur d'observation quotidienne est invraisemblable et appelée précipitation isolée si cette valeur est très grande par rapport à celles des stations voisines. Cette valeur vérifie les conditions suivantes : lorsque la valeur maximale de hauteur de pluie des stations voisines prises dans un rayon de 30 Km de la station cible, est  $< 1$  mm et la hauteur de pluie de la station cible  $> 10$  mm alors cette observation est qualifiée de précipitation isolée.

**Sécheresse isolée :** Une valeur d'observation quotidienne est appelée sécheresse isolée si cette valeur vérifie les conditions suivantes : lorsque la valeur maximale de hauteur de pluie des stations voisines prises dans un rayon de 30 Km de la station cible, est  $\geq 5$  mm, la hauteur de la station cible  $< 1$  mm et le premier quartile des hauteurs de pluie des stations voisines  $\geq 10$  mm

**Les trop grandes déviations :** Pour contrôler la déviation d'une observation des valeurs des stations voisines, l'approche du score Z a été utilisée.

Une observation est très déviée au-dessus et est invraisemblable si :

$$\frac{P_{obs} - Q_{50}}{IQR} > 2.f \text{ et } P_{obs} > 10mm \quad (9)$$

où  $P_{obs}$  : hauteur de la station cible,  $Q_{50}$  : La médiane des stations voisines,  $IQR$  : L'intervalle interquartile des stations voisines,  $f$  : facteur multiplicateur pris égal à 3 (Velleman & Hoaglin, 1981) cité par [Eischeid et al. \(1995\)](#).

Une observation est très déviée en-dessous et est invraisemblable si :

$$\frac{P_{obs} - Q_{50}}{IQR} < -f \text{ et } P_{obs} \leq 1mm \quad (10)$$

Les différents seuils fixés tiennent compte de l'analyse sur la variation spatiale et temporelle des données décadaires de pluie présentée par [Midingoyi \(2016\)](#) dans son étude intitulée : "Approche géostatistique d'optimisation du réseau pluviométrique du Bénin pour un meilleur suivi de la campagne agricole".

## **b Contrôle de la qualité des données de températures**

### **— Cohérence temporelle :**

L'équation (3) est utilisée pour la détection des outliers des données de températures. Les données de températures suivant une distribution approximativement normale, la fonction de distribution cumulative de la loi normale est utilisée dans l'équation (4) pour déterminer le facteur multiplicatif. Les valeurs aberrantes sont détectées pour chaque station et pour chaque jour.

### **— Cohérence spatiale :**



Un test de régression spatiale est utilisé pour vérifier la cohérence spatiale des données (Hubbard & You, 2005; Hubbard et al., 2005). C'est une approche de contrôle de qualité qui vérifie si la variable cible se situe dans l'intervalle de confiance formé par les données des stations environnantes sur une fenêtre de longueur  $n = 30$  (exemple : pour le 1<sup>er</sup> janvier 1981, la fenêtre est constituée des 30 valeurs de 1<sup>er</sup> janvier de 1981 à 2010). Toutes les stations à une certaine distance (100 Km) de la station d'intérêt sont sélectionnées, et une régression linéaire est effectuée pour chaque station couplée à la station d'intérêt et centrée sur la donnée d'intérêt (températures). Pour chaque station environnante, des estimations basées sur le modèle de régression sont formées :

$$x_i = a_i + b_i y_i \quad (11)$$

où  $y_i$  est la série de température de longueur  $n$  de la  $i^{eme}$  station voisine,  $x_i$  est la série des estimations à la station d'intérêt,  $a_i$  et  $b_i$  sont les coefficients de régression entre la station cible et la  $i^{eme}$  station voisine.

Une estimation pondérée ( $x'_j$ ) est obtenue pour le jour  $j$  en utilisant l'erreur-type d'estimation ( $s$ ), également appelée Erreur Quadratique Moyenne (RMSE) dans le processus de pondération :

$$x'_j = \frac{\sum_{i=1}^N \frac{x_{ij}^2}{s_i^2}}{\sum_{i=1}^N \frac{1}{s_i^2}} \quad (12)$$

où  $x_{ij}$  est la valeur estimée pour le jour  $j$  avec l'équation (11) de régression linéaire,  $N$  est le nombre de stations avoisinantes utilisé ( $N \geq 2$ ) dans l'estimation.

L'erreur-type pondérée de l'estimation ( $s'$ ) est calculée :

$$\frac{1}{s'^2} = \frac{\sum_{i=1}^N \frac{1}{s_i^2}}{N} \quad (13)$$

Ainsi les intervalles de confiance peuvent être déterminés sur la base de  $s'$  et la valeur  $x$  de la station cible pour le jour  $j$  peut être testée pour déterminer si elle se situe ou non dans les intervalles de confiance :

$$\frac{|x - x'|}{s'} \succ f \quad (14)$$

le facteur  $f$  est obtenu en utilisant la méthode décrite ci-dessus (équation 4)

#### — Vraisemblabilité des données de températures :

Il est contrôlé ici la cohérence entre les données de températures minimales et maximales d'un même jour, les jours successifs de même températures et les amplitudes thermiques suspectes.

Il est important de remarquer qu'avant toutes ces procédures de contrôle il a été procédé à une

vérification des coordonnées des stations en détectant les stations ayant les mêmes coordonnées et les stations ayant des coordonnées situées en dehors des limites du pays.

### 3.3.2 Méthodologie pour la fusion des données

#### **a** Fusion des données de précipitations

La fusion des données de précipitations a été réalisée en deux étapes : la correction de biais des données de pluies estimées par satellite et la fusion proprement dite des données corrigées de biais avec l'observation.

La méthode Quantile Mapping (Qmap) a été utilisée pour la correction de biais des données de pluies estimées par satellite. Cette technique est l'une des méthodes statistiques de correction de biais les plus utilisés (Ines & Hansen, 2006; Piani et al., 2010a; Gudmundsson et al., 2012; Ajaaj et al., 2015). La méthode Qmap se présente comme suit :

- La détermination des fonctions de distribution cumulative (*CDF* ou *F*) des séries journalières de chaque mois de l'année, des pluies estimées ( $F_e$ ) et des pluies observées ( $F_o$ ) sur la période climatologique 1983-2010. Il faut préciser que les séries des pluies estimées sont prises aux points de grilles proches des coordonnées géographiques des stations in-situ. En fait les coordonnées géographiques sont majorées et minorées de  $0.04^\circ$  et une moyenne spatiale de toutes les grilles qui s'y retrouvent est faite pour le jour ;
- les Fonctions de distribution cumulative empiriques des deux séries de pluie (estimées et observées) sont ajustées à une distribution théorique. Plusieurs études utilisent comme distribution théorique une mixture des distributions de Bernoulli et Gamma, dans laquelle la loi de Bernoulli est utilisée pour modéliser la probabilité d'occurrence des précipitations et la loi Gamma utilisée pour modéliser les quantités de précipitation (Gudmundsson et al., 2012; Cannon, 2008). Les paramètres des distributions des deux lois sont estimés par la méthode de maximum de vraisemblance. Ainsi  $F_e$  devient  $\hat{F}_e$  : fonction de distribution cumulative ajustée des pluies estimées et  $F_o$  devient  $\hat{F}_o$  : fonction de distribution cumulative ajustée des pluies observées ;
- les paramètres des distributions sont interpolés par la méthode de pondération inverse des distances (Shepard, 1968) cité par (Contreras & Ferri, 2016) aux grilles des données TAM-SAT (4 km de résolution)
- l'utilisation d'une fonction de transfert (Piani et al., 2010a; Gudmundsson et al., 2012; Ajaaj et al., 2015) pour la correction de biais des pluies estimées selon l'équation suivante :

$$\hat{P}_{i,corr} = \hat{F}_o^{-1}(\hat{F}_e(\hat{P}_i))$$

où  $\hat{P}_{i,corr}$  est la hauteur de pluie corrigée de biais pour le jour  $i$ ,  $\hat{P}_i$  est la hauteur de pluie estimée par satellite,  $\hat{F}_o^{-1}$  est l'inverse de la fonction de distribution cumulative ajustée (fonction quantile) des séries journalières de chaque mois de l'année, des pluies observées (station) sur la période climatologique 1983 – 2010 et  $\hat{F}_e$  la fonction de distribution cumulative

ajustée des séries journalières de chaque mois de l'année, des pluies satellitaires estimées sur la période climatologique 1983 – 2010.

La fusion des données satellitaires corrigées de biais et des données in-situ a été réalisée par la Regression-Kriging (RK). La RK estime la valeur d'une variable en un endroit donné en combinant deux approches : un modèle de régression et le krigéage ordinaire. Le modèle linéaire généralisé est utilisé pour ajuster les variations explicatives afin de prédire la tendance spatiale des données et le krigéage ordinaire (ajustant un variogramme) est utilisé pour interpoler les résidus du modèle de régression c'est à dire les variations non expliquées (Hengl et al., 2007; Dinku et al., 2014b).

Soit  $z(s_1), z(s_2), ..., z(s_n)$ , les relevés des paramètres climatiques dans des localités, avec  $(s_i) = (x_i, y_i)$  et  $x_i, y_i$  les coordonnées géographiques,  $n$  le nombre d'observations. La valeur estimée  $\hat{z}(s_0)$  par RK à une localité dépourvu de station est (Odeh et al., 1994) cité par (Hengl et al., 2004) :

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0)$$

où  $\hat{m}(s_0)$  le terme de la tendance spatiale ajustée en utilisant une analyse de régression linéaire généralisée et  $\hat{e}(s_0)$  le terme des valeurs résiduelles de la régression interpolée en utilisant un krigéage ordinaire. Il vient donc :

$$\hat{z}(s_0) = \sum_{k=0}^p \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^n w_i(s_0) \cdot e(s_i)$$

où  $\hat{\beta}_k$  sont les coefficients du modèle de régression déterminés en utilisant la méthode des moindres carrés généralisés,  $q_k(s_0)$  est la  $k^{eme}$  variable explicative ou auxiliaire au point (Les données satellitaires estimées de pluie sont utilisées comme variables auxiliaires)  $s_0$ ,  $p$  le nombre de prédicteurs,  $w_i(s_0)$  sont les poids déterminés de sorte que la variance de l'erreur de d'interpolation des résidus soit minimisée et  $e(s_i)$  les valeurs résiduelles de la régression.

Le merging (fusion) des données de précipitations a été réalisé(e) sur deux ensembles de données séparément. Le premier est constitué de l'ensemble des postes pluviométriques du réseau utilisé et le second qui est nommé opérationnel est constitué des 55 postes pluviométriques qui transmettent régulièrement leurs données à une fréquence décadaire voire journalière (figure 6, en points noirs).

## **b Fusion des données de températures**

En ce qui concerne la fusion des données de températures, il a été procédé comme suit :

- le Reggridding des données de réanalyses de températures JAR-55 de 50 à 4 km en utilisant la méthode d'interpolation bilinéaire (<https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/regridding-overview>) aux grilles

des données de pluies satellitaires estimées (TAMSAT). Le Regridding est le processus d'interpolation d'une grille de données de résolution à une autre grille de résolution supérieure ou inférieure. Le choix a été porté sur l'interpolation bilinéaire du fait du caractère rectiligne des grilles sources et de destination de nos données ;

- l'utilisation de la méthode quantile mapping pour la correction des biais des données de réanalyses de températures JAR-55 de résolution 4 km. Il faut noter que les Fonctions de distribution cumulative empiriques des deux séries de températures (estimées et observées) sont ajustées à la distribution normale (Piani et al., 2010b) ;
- la fusion des données satellitaires corrigées de biais et des données in-situ a été réalisée par du Regression-Kriging. Elle a été réalisée avec l'ensemble des stations relevant les données de température.

### 3.3.3 Evaluation des données fusionnées

Les données mergées sont évaluées en utilisant le principe de validation croisée spécifiquement le Leave-One-Out Cross Validation (LOOCV) (Dinku et al., 2014b; Manz et al., 2016). Elle consiste à éliminer à tour de rôle chaque station choisie aléatoirement, puis à merger les données satellitaires (en point de grille) avec les données in-situ restantes afin d'estimer la valeur des paramètres à l'emplacement de la station retirée. Les valeurs estimées ( $\tilde{X}_i$ ) et réelles ( $X_i$ ) aux emplacements des stations omises sont comparées. Dinku et al. (2014b) ont retenu des statistiques pour l'évaluation des données mergées , il s'agit :

- du coefficient de corrélation :  $r(X, \tilde{X})$  ;
- de l'erreur absolue moyenne :  $MAE = \frac{\sum_1^N |X_i - \tilde{X}_i|}{N}$  J. Willmott & Matsuura (2005) ont montré les avantages que présente ce critère par rapport au RMSE dans l'évaluation de la performance d'un modèle ;
- du coefficient d'efficacité de Nash-Sutcliffe :  $NSE = 1 - \left[ \frac{\sum_1^N (X_i - \tilde{X}_i)^2}{\sum_1^N (X_i - \bar{X})^2} \right]$  Nash & Sutcliffe, (1970) cité par Legates & McCabe (1999) ont défini le coefficient d'efficacité qui varie entre  $-\infty$  et 1. Les plus fortes valeurs indiquent une fusion convenable des données ;
- du biais ou du résidu moyen :  $e = \frac{\sum_1^N X_i - \tilde{X}_i}{N}$ .

Le processus de validation croisée a été réalisée pour l'ensemble des stations pour les données de températures. Pour ce qui concerne les données de précipitations mergées, la validation a été réalisée non seulement pour l'ensemble des stations (réseau global, points noirs et les croix) mais aussi pour le réseau opérationnel (figure 6, en points noirs), une zone fluide en poste pluviométrique (Z1) et une zone riche en postes pluviométriques (Z2) (figure 6).

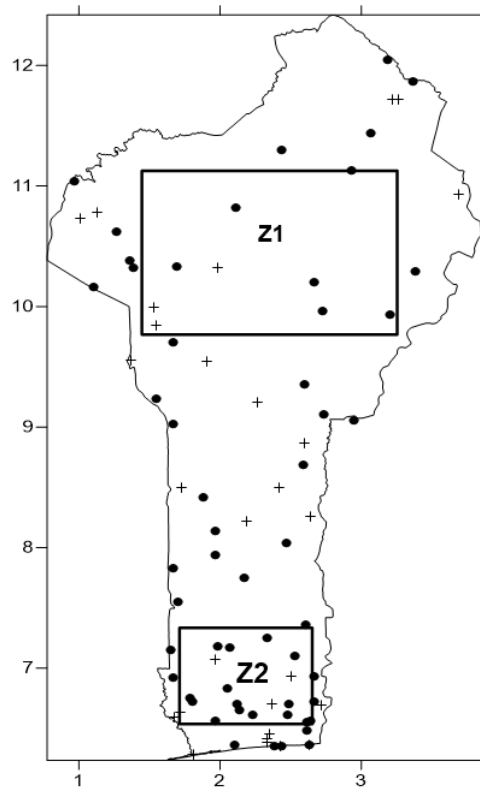


FIGURE 6 – Réseau pluviométrique opérationnel (points en noirs). Réseau global (points noirs et les croix). Les deux rectangles pour l'évaluation des produits mergés dans les zones pauvres en station (Z1) et dans les zones riches en station (Z2).

Des diagrammes de Taylor ([Taylor, 2001](#)) ont été aussi réalisés pour apprécier la qualité des données mergées. Ce diagramme très utilisé dans le domaine de la modélisation, fournit un cadre graphique permettant de comparer une suite de variables provenant de divers modèles ou réanalyses aux données de référence. Il est produit sur ce diagramme un résumé statistique concis de la manière dont les modèles correspondent aux données de référence en ce qui concerne leurs corrélations, leurs erreurs quadratiques moyennes (RMSE) et le ratio de leurs écart-types. Dans le cas de cette étude les modèles correspondent aux données mergées et satellitaires et la référence aux données observées. Les séries de précipitations et de températures ont été standardisées avant la réalisation du diagramme.

### 3.4 Outils

Dans le cadre de ce travail plusieurs outils ont été utilisés. Au nombre de ces outils on note :

- des scripts écrits sous le langage R (<https://www.r-project.org/>) pour le traitement des données et la réalisation des différents graphiques et cartes ;
- le package `gstat` (<http://www.gstat.org/>) du langage R (<https://www.r-project.org/>) est implémenté dans l'outil Climate Data Tools (CDT) (<https://github.com/rijaf/CDT>) pour le merging et l'évaluation des produits mergés ;
- l'éditeur  $\text{\LaTeX}$  pour la compilation de ce document.

## 4 Résultats et interprétations

### 4.1 Contrôle de qualité des données

Un extrait des résultats issus du contrôle de qualité des données est présenté en **annexe C**. Les valeurs aberrantes détectées ont été vérifiées en consultant les archives au niveau de MÉTÉO BENIN. Celles-ci sont remplacées par les valeurs des archives lorsqu'elles ne sont pas conformes et considérées comme manquantes lorsqu'elles sont conformes aux archives mais paraissent inexplicables climatologiquement. C'est le cas par exemple des données de températures maximales de la station de Dassari pour le 08 février 1998 (figure 7) où une valeur de 49°C a été reportée et conforme à l'archive. Cette valeur a été simplement considérée comme une valeur manquante étant donné qu'on n'a pas pu l'expliquer d'un point de vue climatologique. Pour le cas de la station de Cotonou (figure 8) la forte valeur de température maximale du 27 Avril 2008 (37.5°C) relevée dans la base de données n'est pas conforme à celle retrouvée dans les archives (27.5°C). Une erreur liée probablement à l'enregistrement des données dans la base de données qui a été corrigée.

Le contrôle de cohérence spatial a révélé tous les cas possibles (précipitations isolées, sécheresse isolées, forte déviation au dessus et en dessous) de valeurs aberrantes. Il est exposé ici quelques cas. Le 05 octobre 1999 il a été relevé à Dangbo une hauteur de pluie de 59.6 mm (figure 9) sans une goutte de pluie significative dans le voisinage immédiat : c'est une précipitation isolée. A Dogbo (figure 10) le 08 juillet 2000 l'effet contraire a été observé : c'est une sécheresse isolée. Deux exemples de valeurs fortement déviées au dessus à Adjohoun (figure 11) le 10 Août 1987 et à Tanguiéta (figure 12) le 30 Août 1996 ont été présentés.

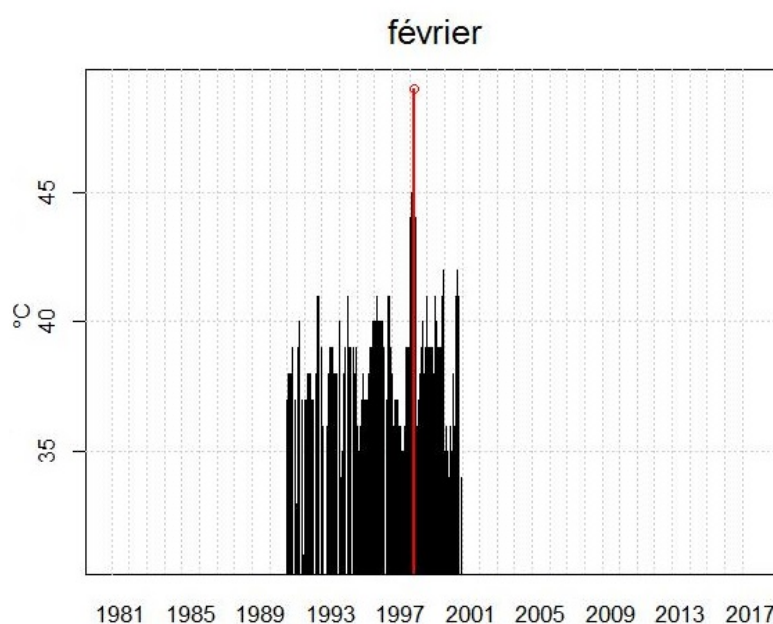


FIGURE 7 – Températures maximales des mois de février (1981 – 2017) de Dassari- En rouge celle du 08 février 1998

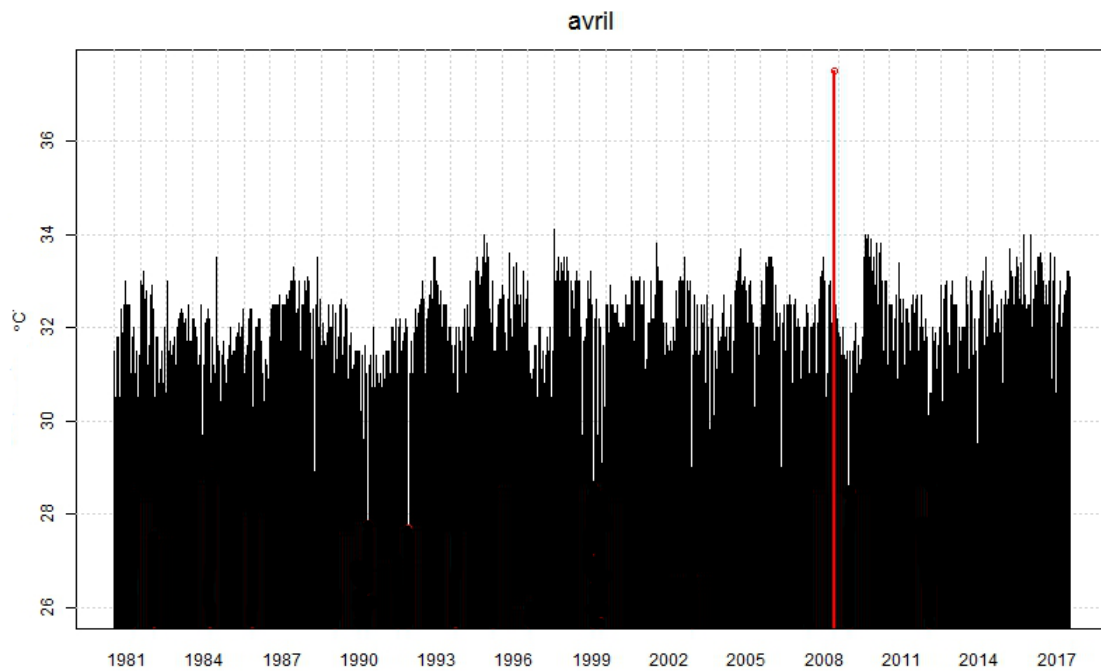


FIGURE 8 – Températures maximales des mois d'avril (1981 – 2017) à Cotonou- En rouge celle du 27 Avril 2008

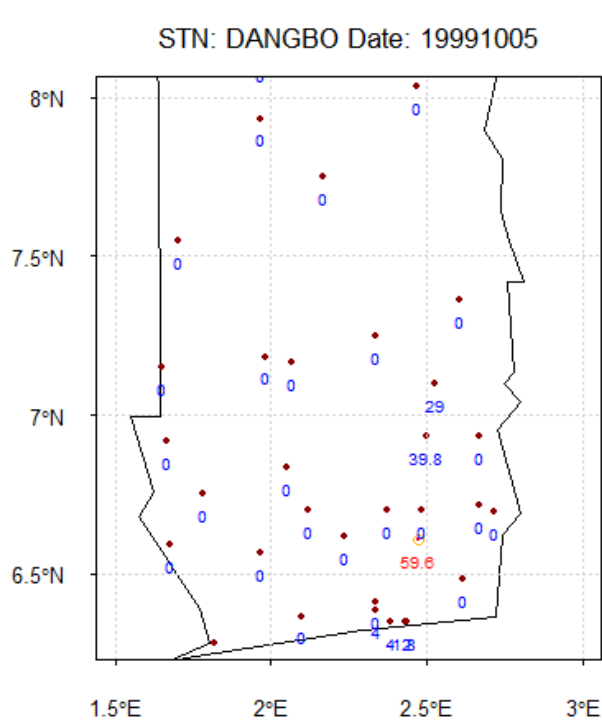


FIGURE 9 – Précipitation isolée à Dangbo (rouge) le 05/10/1999.

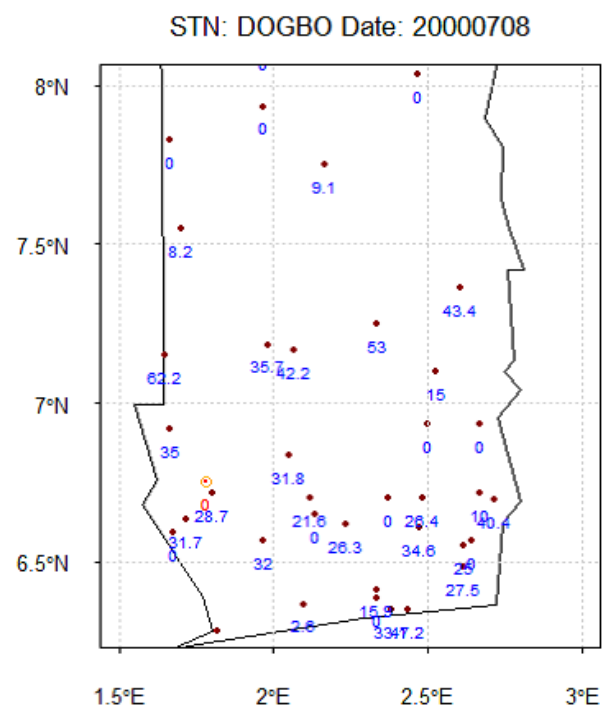


FIGURE 10 – Sécheresse isolée à Dogbo (rouge) le 08/07/2000.

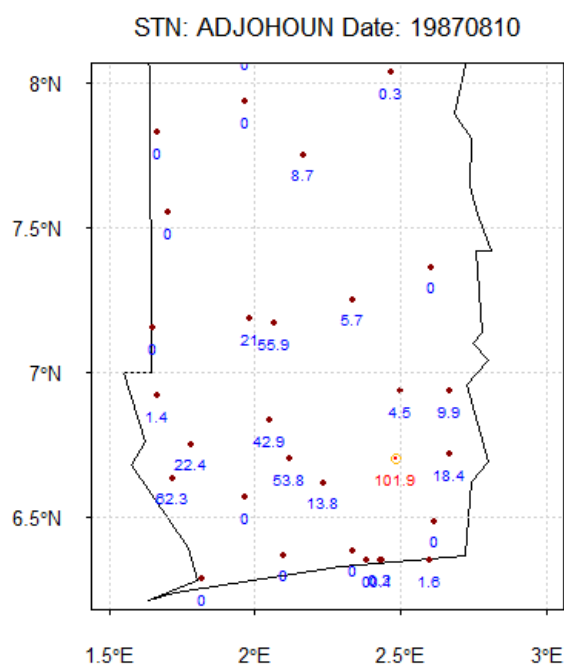


FIGURE 11 – Forte déviation au dessus à Ad-johoun (rouge) le 10/08/1987.

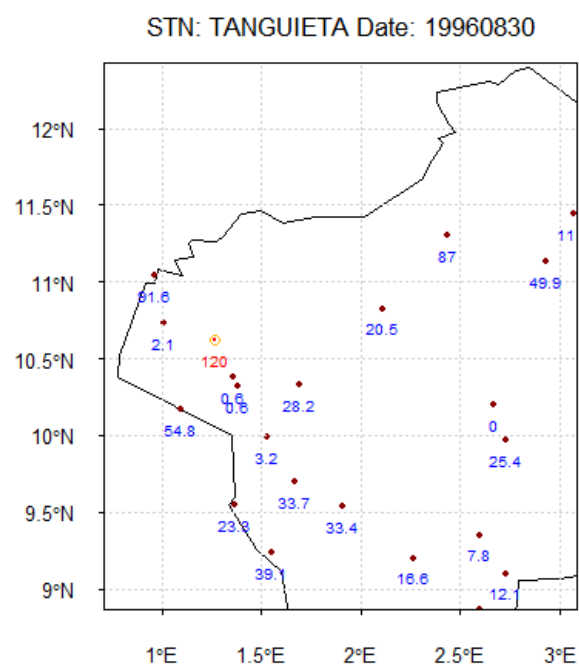


FIGURE 12 – Forte déviation au dessus à Tan-guiéta (rouge) le 30/08/1996.

Le pourcentage de données aberrantes par mois (rapport entre le nombre de données aberrantes et le nombre de données non manquantes du mois) sur l'ensemble des stations est présenté dans le Tableau 2. On note de fort pourcentage pour les mois pluvieux en ce qui concerne la pluie. Les valeurs aberrantes ont été vérifiées et analysées afin d'être conservées ou rejetées. Toute chose qui a abouti à l'amélioration de la qualité de nos séries climatiques journalières.

Tableau 2 – Pourcentage des valeurs aberrantes par mois sur l'ensemble des stations

Paramètres	Mois											
	Jan	Fév	Mar	Avr	Mai	Juin	Juil	Août	Sept	Oct	Nov	Déc
RR	0.02	0.01	0.02	0.04	0.08	0.15	0.14	0.18	0.11	0.11	0.04	0.4
TMIN	0.08	0.21	0.16	0.03	0.04	0.19	0.13	0.14	0.07	0.32	0.165	0.13
TMAX	0.41	0.33	0.27	0.37	0.41	0.23	0.21	0.05	0.41	0.18	0.38	0.34

## 4.2 Données mergées

### 4.2.1 Réduction d'échelle des données de températures

D'une façon générale, il a été constaté que le reggridding ne dénature pas la structure des champs de températures, la variation spatiale est relativement conservée. Les structures sont évidemment plus lissées que dans la faible résolution spatiale (50 km). L'exemple présenté pour la journée du 01/05/1997 (figure 13) révèle qualitativement très peu de différence entre les champs avant (50 km de résolution au milieu) et après (4km de résolution à droite) l'interpolation, que l'on soit en zones de montagnes (Nord-Ouest du pays notamment confère figure 1) ou en zones côtières.



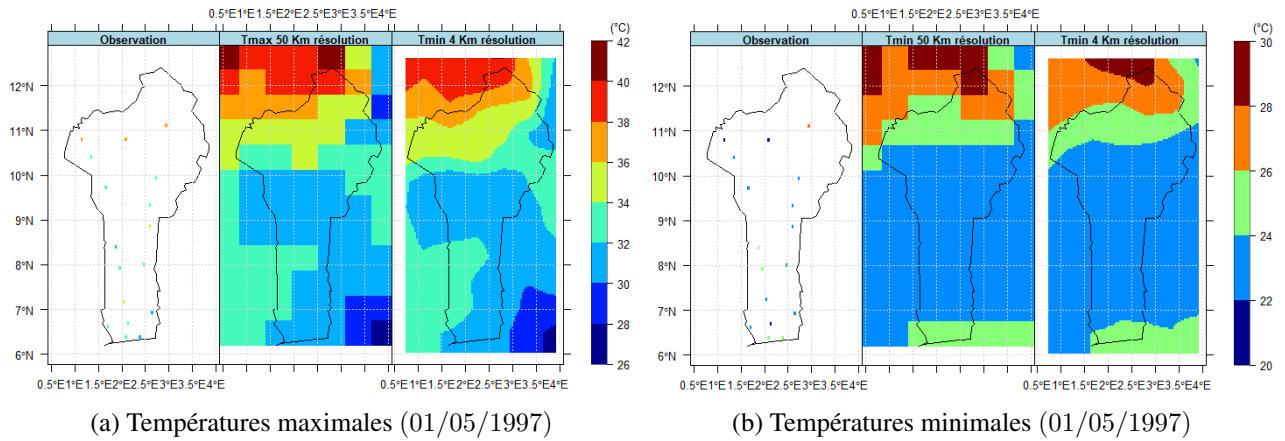


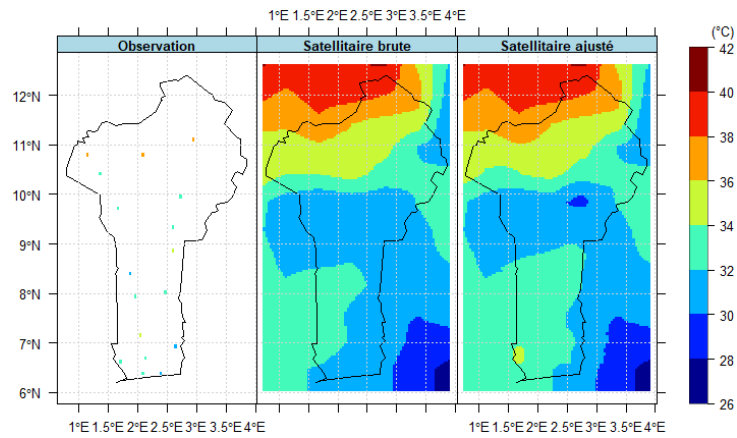
FIGURE 13 – Reggridding des données de températures du 01/05/1997. Observation (à gauche), Satellitaire brute de résolution 50 km (au milieu), Satellitaire brute de résolution 4 km (à droite).

#### 4.2.2 Correction de biais des données satellitaires brutes

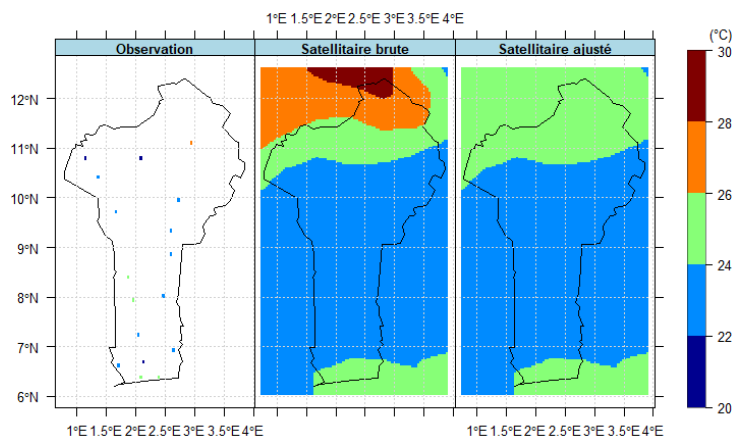
Il est présenté des exemples de données d'estimations satellitaires de pluie et de réanalyses de températures, corrigées de biais (figure 14 et 15) pour les journées du 10/05/1997 pour les températures et 11/06/2017 pour la pluie. Les données corrigées de biais sont comparées aux données satellitaires estimées ou de réanalyses brutes et aux données observées. Les améliorations sont évidentes. Certaines structures observées et non prises en compte dans les champs de données satellitaires brutes sont mis en relief dans les champs de données corrigées. C'est le cas par exemple des températures maximales où il est constaté une extension de la structure du champ comprise entre 32 et 34°C au sud avec un coin chaud à l'ouest autour de 7°N de latitude reproduisant assez bien l'observation (figure 14a). Pour ce qui concerne les températures minimales, il est noté une assez bonne reproduction de l'observation aussi bien pour les données satellitaires brutes que pour leurs corrections de la côte à la latitude de 11° N. Néanmoins, la correction apportée autour de la latitude 11°N n'est pas tout à fait conforme à l'observation (figure 14b).

Il est constaté une sous-estimation des hauteurs de pluie aussi bien par les données satellitaires brutes que par leurs corrections. Toutefois les structures observées sont beaucoup plus mises en évidence sur le champ de données corrigées. C'est le cas au Sud, en dessous de 07°N de latitude, où les deux structures à l'est et à l'ouest sont plus mises en relief sur le champ de données ajustées. C'est aussi le cas au centre-est autour de 08°N de latitude et au nord entre 09°N et 10°N où l'observation est nettement plus représentée par les données satellitaires ajustées que celles satellitaires brutes. Ceci montre que la sous-estimation des données ajustées est moindre que celle des données brutes (figure 15).

Les courbes des fonctions cumulatives de distributions empiriques des données observées, satellitaires brutes et satellitaires ajustées pour les séries des mois de mai, juin, juillet et Août sont présentés en **annexe D**. Ces courbes permettent d'avoir une appréciation de la qualité des corrections apportées en terme de distribution des différentes séries.



(a) Températures maximales (01/05/1997)



(b) Températures minimales (01/05/1997)

FIGURE 14 – Correction de biais des données satellitaires brutes de températures. Observation (à gauche), Satellitaire brute (au milieu), Satellitaire corrigé de biais (à droite).

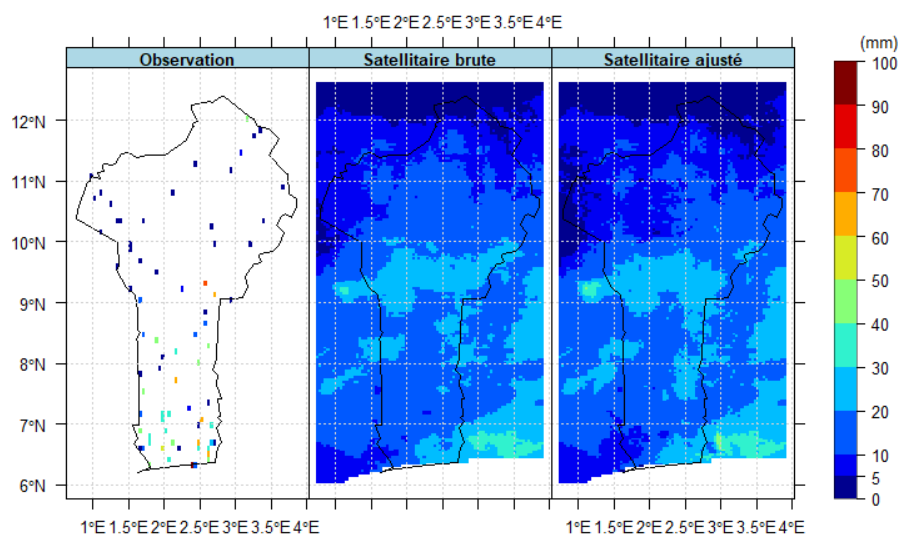


FIGURE 15 – Correction de biais des données satellitaires brutes de pluie (11/06/2017). Observation (à gauche), Satellitaire brute (au milieu), Satellitaire corrigé de biais (à droite).

### 4.2.3 Données mergées de pluie

Des exemples de produits mergés de pluie sont présentés pour la journée du 11/06/2017 aussi bien pour le réseau opérationnel (figure 16a) que pour le réseau global (figure 16b). Pour chaque figure (16a et 16b), nous avons respectivement, en haut et à gauche les données observées, en haut et à droite les données satellitaires brutes estimées de pluie, en bas et à gauche une interpolation des données observées par krigéage ordinaire aux grilles des données TAMSAT et enfin en bas et à droite les données fusionnées de pluie (satellitaire corrigé de biais et station). Les variogrammes ayant servi d'ajustement pour l'interpolation se trouvent en **annexe E**.

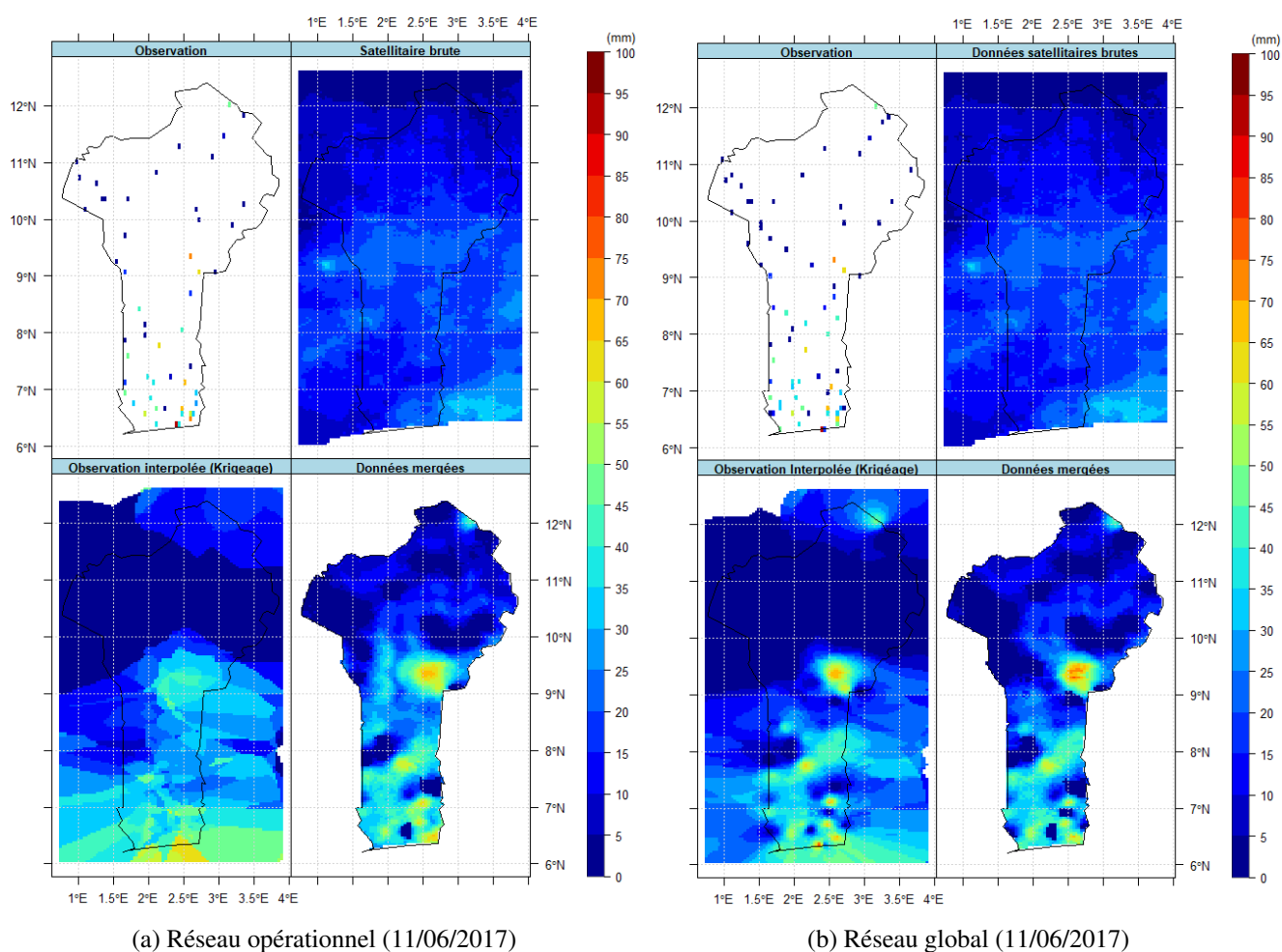


FIGURE 16 – Données mergées de pluie pour le 11/06/2017. Observation (en haut à gauche), satellitaire brute (en haut à droite), observation interpolée (en bas à gauche), mergée (en bas à droite).

La densité du réseau reste la principale faiblesse de l'observation qui est pourtant notre référence. Ceci est plus perceptible dans les régions du nord du pays.

Sur le produit satellitaire brut, on retrouve les principales structures spatiales de précipitations observées. Toutefois, ces données satellitaires sous-estiment les quantités observées sur la majeure partie du pays.

Les interpolations de l'observation, montrent les mêmes structures spatiales de pluie que les données in-situ mais avec un certain lissage. L'autre problème avec ces interpolations, se situe à l'extrême nord dans la zone dépourvue de station au dessus de 11.5°N. La forte valeur enregistrée à la station de Karimama (3.18°E, 12.5°N) a influencé sensiblement les estimations de l'interpolation dans toute la zone. L'interpolation de l'observation de la figure 16a n'a pas bien reproduit la structure spatiale de pluie observée dans la boîte comprise entre 1.5°E - 2°E de longitude et 8°N - 9°N de latitude, qui est assez bien représenté par le produit satellitaire brut et par l'interpolation de l'observation de la figure 16b. Ceci est probablement dû au nombre de stations utilisé pour l'interpolation (55 pour la figure 16a contre 83 pour la figure 16b) et plus précisément au nombre de stations dans la boîte pour les deux cas (2 contre 3). Le lissage est moins important au niveau de l'interpolation de la figure 16b, mais le problème à l'extrême nord du pays persiste.

Les données mergées représentent bien l'observation dans les deux cas. Toutefois Leur qualité est meilleure avec le réseau global qu'avec le réseau opérationnel. Il est aussi constaté que l'interpolation de l'observation avec le réseau global (figure 16b) affiche des champs similaires aux données mergées dans les zones du sud denses en postes pluviométriques. Au nord, moins densifié une situation contraire s'observe pour les données interpolées et ceci est plus accentué pour le réseau opérationnel.

#### **4.2.4 Données mergées de températures**

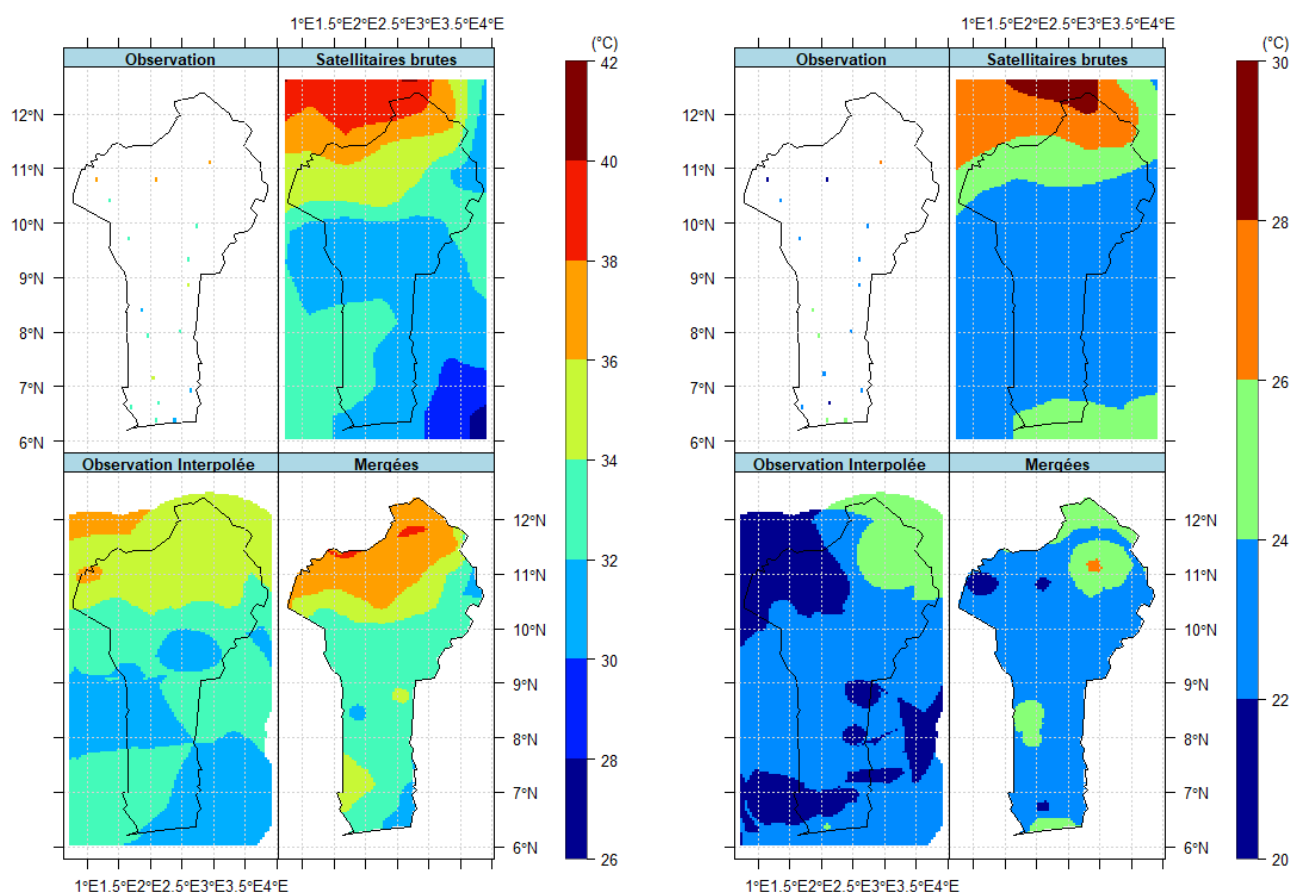
Il est présenté des exemples de produits mergés de températures maximales (figure 17a) et minimales (figure 17b) pour la journée du 01/05/1997. Pour chaque figure nous avons respectivement, en haut et à gauche les données observées, en haut et à droite les données de réanalyses brutes de températures, en bas et à gauche une interpolation des données observées par la pondération inverse des distances aux grilles des données TAMSAT et enfin en bas et à droite les données fusionnées de températures.

Les données de réanalyses de températures maximales (figure 17a) sous-estiment les valeurs de températures maximales observées, bien qu'elles représentent bien les champs de températures qui se trouvent dans l'observation. Pour ce qui concerne les températures minimales ((figure 17b) on note une sur-estimation générale avec une très bonne représentation (même champ de températures) de l'observation.

Les interpolations de l'observation pour les deux figures (figure 17a,17b) affichent un profond lissage dénaturant certaines valeurs remarquables sur l'observation. C'est l'exemple de la bande du littoral pour les températures minimales où les fortes valeurs (entre 24°C et 26°C) disparaissent dans le champ interpolé. Le lissage est évidemment dû au nombre réduit de station

pour l'interpolation.

Les produits mergés de températures maximales et minimales affichent toutes les structures spatiales de températures présentes dans l'observation avec des valeurs presque égales aux valeurs observées. Les particularités de chaque région sont mises en évidence. On peut remarquer la bonne représentation des faibles températures minimales sur les zones de relief au nord ouest.



(a) Températures maximales (01/05/1997)

(b) Températures minimales (01/05/1997)

FIGURE 17 – Données mergées de températures pour le 01/05/1997. Observation (en haut et à gauche), satellitaire brute (en haut et à droite), observation interpolée (en bas et à gauche), mergée (en bas et à droite).

### 4.3 Évaluation des données mergées

Les produits mergés montrent en général de bonnes performances. La corrélation avec l'observation est plus forte pour les températures maximales mergées (0.926) que pour les températures minimales mergées (0.911). Toutefois le coefficient d'efficacité des températures minimales est supérieur à celui des températures maximales (Tableau 3). En comparaison avec les données de réanalyses brutes (figure 18a,18b), on observe sur le diagramme de Taylor que les séries mergées

sont plus proches de l'observation en termes de corrélation (forte pour les produits mergés) et d'erreur quadratique moyenne (faible pour les produits mergés).

Les données mergées de pluie avec le réseau global montre une bonne corrélation, une bonne efficacité et moins d'erreurs que celles mergées avec le réseau opérationnel (Tableau 3). Les données mergées pour les deux réseaux présentent des performances meilleurs que leurs paires d'observations satellitaires brutes.

Pour la région pauvre en postes pluviométriques (Z1), ils se trouvent que les données satellitaires brutes montrent des performances presque équivalente aux produits mergés alors que sur la zone riche en postes pluviométriques (Z2) la performance des données mergées est nettement meilleure.

Tableau 3 – Statistiques d'évaluation des données fusionnées par validation croisée

	Précipitations				Températures	
	Zone Z1	Zone Z2	Réseau opérationnel	Réseau global	Maximales	Minimales
$r(X, \tilde{X})$	0.419	0.621	0.787	0.892	0.926	0.911
$NSE$	0.223	0.547	0.631	0.730	0.845	0.867
$MAE$	3.59	3.01	2.89	2.08	1.76	1.98
$e$	0.863	0.855	0.889	0.874	0.494	0.521

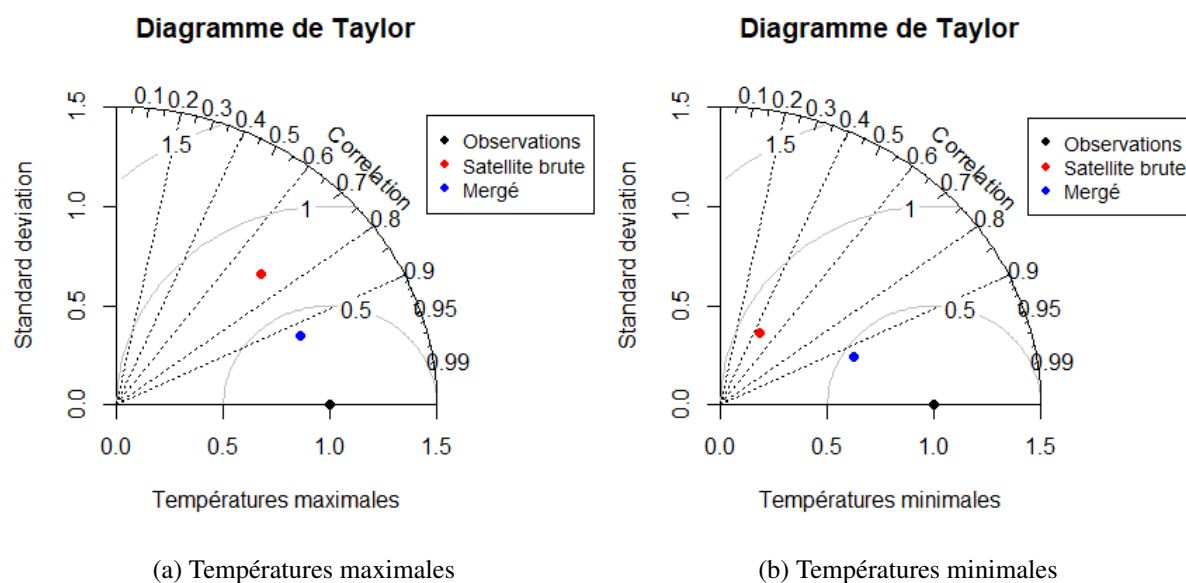


FIGURE 18 – Diagramme de Taylor des séries de températures observées, satellitaires brutes et mergées

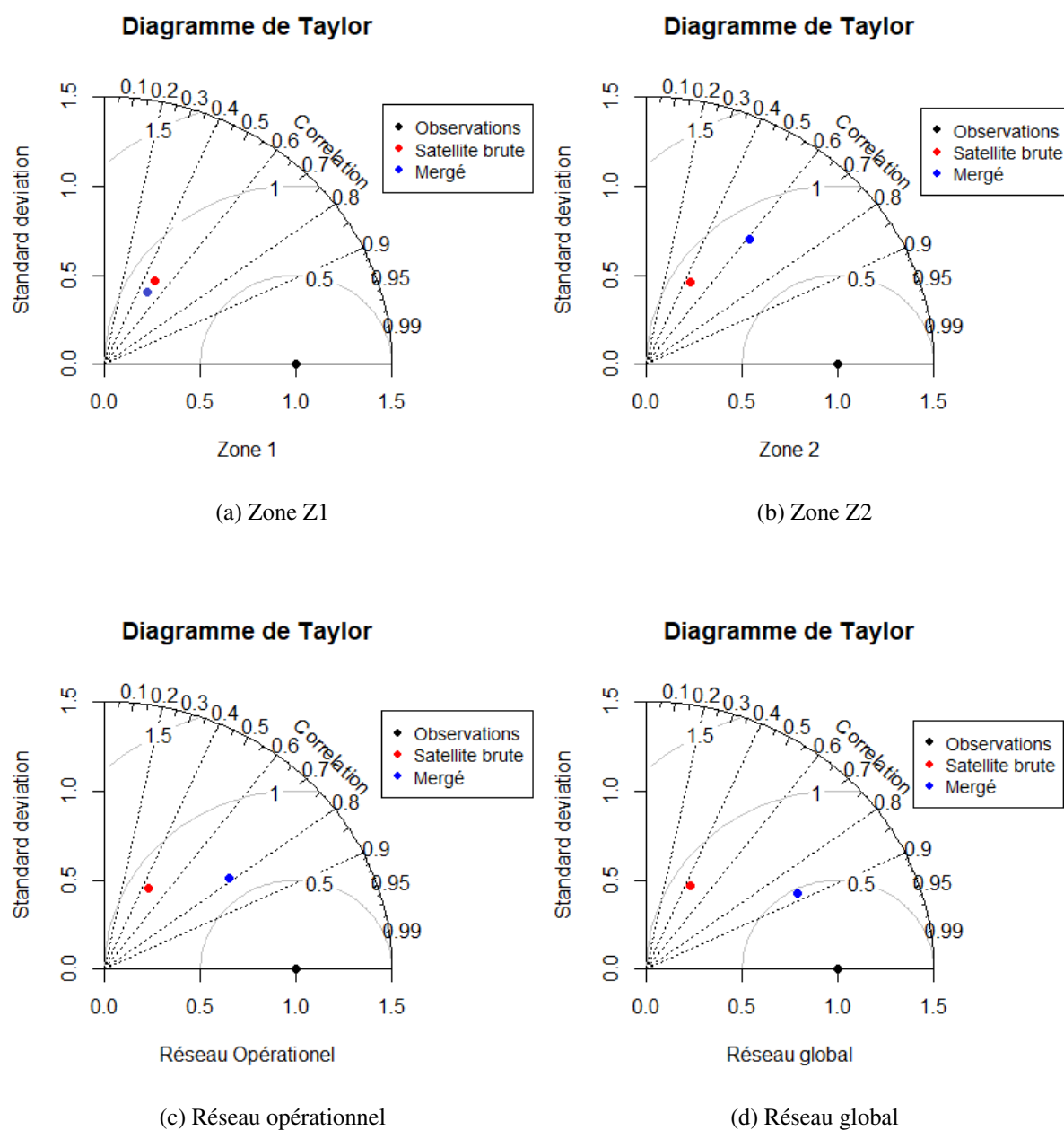


FIGURE 19 – Diagramme de Taylor des séries de pluies observées, satellitaires brutes et mergées



## 5 Discussions

Les données corrigées de biais présentent des performances assez meilleures pour les températures que pour les précipitations. Les champs de températures corrigées ressemblent presque parfaitement à l'observation. Ceci est probablement dû à la distribution (Variation spatiale) continue de cette variable. De plus, les données de précipitations corrigées de biais sous-estiment toujours l'observation mais dans une moindre mesure que celles satellitaires brutes. (Dinku et al., 2014b) par contre trouve de meilleures performances pour les données TAMSAT corrigées de biais, dans son étude sur l'Ethiopie. Toutefois, il est important de souligner qu'il a travaillé à une échelle décadaire avec une méthode de correction multiplicative de biais et des algorithmes TAMSAT parfaitement calibrés sur son domaine d'étude.

Les produits mergés pour les précipitations résolvent à un certain degré aussi bien la fluidité du réseau d'observation que la sous-estimation de l'observation par les données satellitaires brutes. Bien qu'une évaluation statistique n'ait été réalisée pour les interpolations de l'observation, une analyse qualitative simple permet de remarquer que ces dernières sont plus représentatives de l'observation que les données satellitaires brutes à l'exception de la zone Z1 (moins riche en postes pluviométriques). Les données satellitaires brutes représentent mieux le champ d'observation dans cette zone que les interpolations de l'observation, quoiqu'elles sous-estiment en quantité l'observation. Les produits fusionnés corrigent sensiblement cette lacune en termes d'écart ( $\approx 0.4$ ) à l'observation. Les produits fusionnés et interpolés sont presque similaires sur les zones riches en postes pluviométriques du pays. L'avantage des produits mergés réside dans les zones du pays dépourvues de stations, ce résultat est conforme à celui de Dinku et al. (2014b). L'augmentation du nombre de stations de 55 (réseau opérationnel) à 83 (réseau global) a amélioré aussi bien les produits mergés que ceux interpolés. La structure spatiale du produit mergé avec 55 postes (figure 16a) est proche de celle du produit mergé avec 83 postes (figure 16b), toutes deux proches de l'observation. C'est une preuve qu'un réseau optimal pourrait être défini pour le processus de merging. Un réseau qui ne tiendrait pas seulement compte du nombre de stations mais aussi se pencherait sur la bonne répartition spatiale des stations. L'étude de Midingoyi (2016) devient intéressante dans ce cadre, en ce sens qu'un réseau optimum a été proposé par ce dernier pour le suivi agrométéorologique décadaire, même si le but primordial du réseau qu'il propose est la minimisation de la variance d'erreur d'interpolation des pluies décadales.

Les produits mergés pour les températures montrent de très bonne performance, malgré le nombre réduit de stations utilisées pour le processus. Ceci pourrait être lié à la distribution spatiale continue de la variable. De plus certaines particularités d'ordre géographique sont mises en évidence sur ces produits mergés de températures, il s'agit de : l'effet côtier sur les températures, les températures minimales sur les reliefs dans les dépressions et vallées.



## 6 Applications agrométéorologiques

D'une façon générale, les données mergées permettent la fourniture de l'information climatique à une résolution spatiale réduite. La fluidité des réseaux d'observations se trouve corrigée un temps soit peu. Il est présenté ici quelques analyses agrométéorologiques en comparaison aux données in-situ.

### 6.1 Résumé décadaire et mensuel de températures et de pluie

Dans l'exemple du cumul décadaire (figure 20), une bonne partie de l'information n'est pas re-tracée par les données observées surtout dans le nord du pays à cause de la densité du réseau. Cependant, le cumul mergé donne l'information sur la précipitation décadaire sur tout le pays, ce qui est très intéressant dans la production des informations agrométéorologiques décadaires.

Les figures 21 et 22 mettent en exergue, une fois de plus l'effet bonifiant des produits mergés. Toutes les particularités géographiques (bordure de mer, zone de relief...) en terme de températures sont ressorties sur les produits mergés. Un zonage climatique en fonction de la température apparaît nettement.

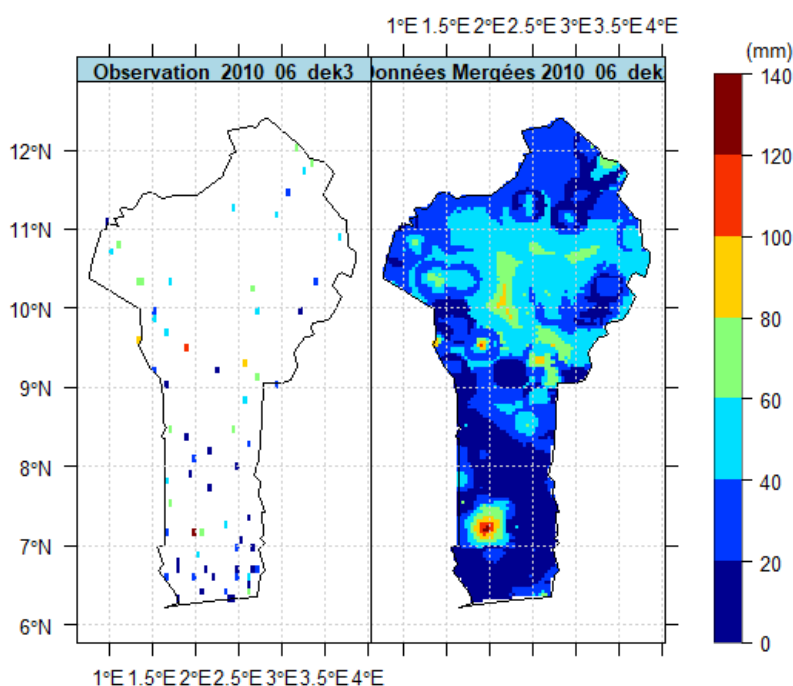


FIGURE 20 – Cumul de la 3<sup>ème</sup> décennie de pluie du mois de juin 2010. Observation (à gauche), Mergé(à droite).

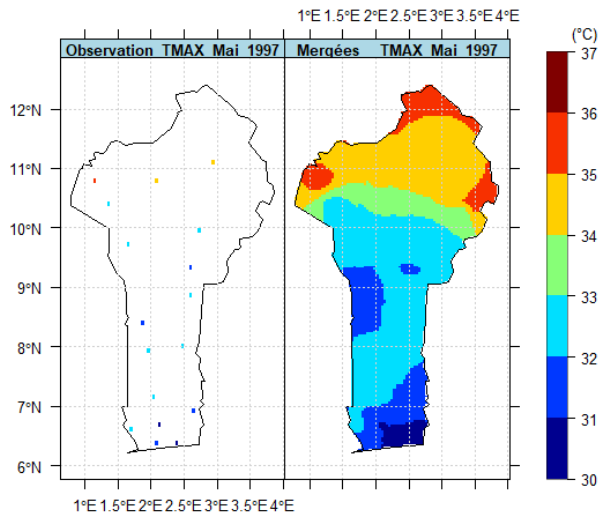


FIGURE 21 – Moyenne mensuelle des températures maximales du mois de mai 1997. Observation (à gauche), Mergé (à droite)

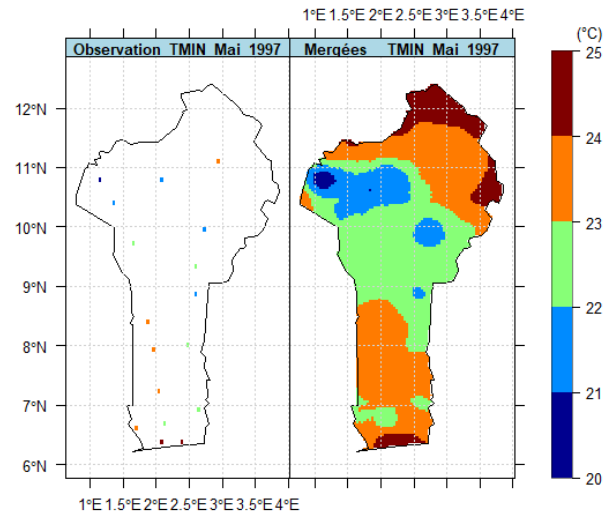


FIGURE 22 – Moyenne mensuelle des températures minimales du mois de mai 1997. Observation (à gauche), Mergé (à droite)

## 6.2 Quelques paramètres agroclimatiques

Des paramètres agroclimatiques ont été déterminés avec les produits mergés (figure 23, 24). Il s'agit notamment de l'évapotranspiration potentielle (ETP) calculée avec la formule de Hargreaves et al., 1985 cité par (Waongo, 2017) et de la date de début de la grande saison pluvieuse au sud et de l'unique saison pluvieuse au nord du pays pour l'année 1997. L'ETP calculée avec le produit mergé reflète parfaitement l'évolution spatiale en termes de demande en eau potentielle de l'atmosphère sur le pays (figure 23).

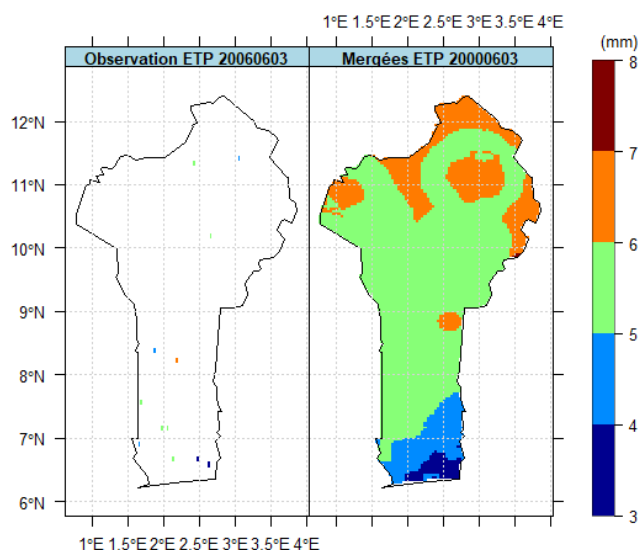


FIGURE 23 – Evapotranspiration potentielle pour la journée du 03/06/2006. Observation (à gauche), Mergé (à droite).

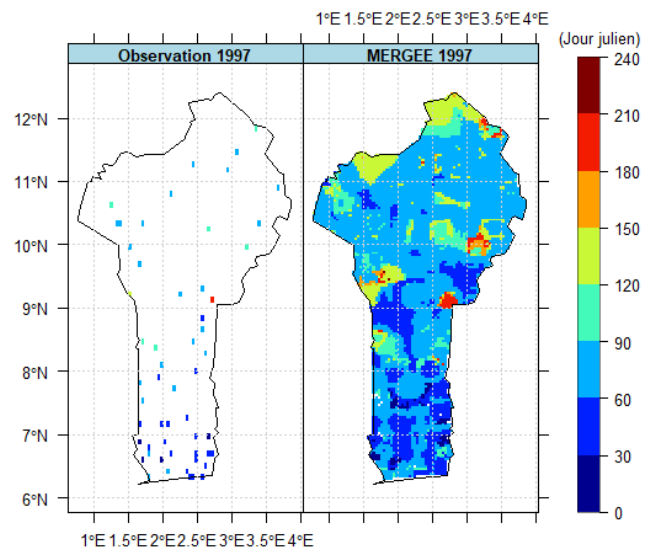


FIGURE 24 – Date de début de saison pour l'année 1997. Observation (à gauche), Mergé (à droite).

Pour le démarrage des saisons (figure 24), les critères utilisés sont adaptés de Sivakumar (1988) et

se présentent comme suit : pour la zone climatique à régime bimodale (guinéenne), le démarrage de la grande saison des pluies correspond à la date à partir du 1<sup>er</sup> Février recueillant une hauteur d'eau d'au moins vingt (20) mm sur 01 à 03 jours consécutifs, sans qu'il n'y ait de séquences sèches de plus de dix (10) jours dans les trente (30) jours qui suivent. Pour ce qui concerne la zone climatique à régime monomodal, la saison des pluies démarre à partir du 15 MARS, lorsqu'on enregistre au moins 20 mm de pluie en 01 à 03 jours consécutifs et ceci sans épisodes secs de plus de 10 jours dans les 30 jours qui suivent. De l'observation de la figure 24, il ressort des dates comprises entre le 1<sup>er</sup> Février (précoce) et le 1<sup>er</sup> Avril (tardive) au sud. Au nord les dates les plus tardives se situent autour de mi-juillet. Ainsi, en plus d'une parfaite représentation de l'observation, l'information est disponible pour toutes les contrées à une résolution de 4 km.

### 6.3 Séries temporelles de la pluviométrie dans une localité dépourvue de station

Les données mergées ont été utilisées pour des analyses climatiques (figure 25 et 26) dans la localité de Gogounou (2.6°E, 10.9°N) où aucune station d'observation météorologique n'est disponible. La station de référence (synoptique) la plus proche se situe à 100 km plus bas en latitude : la station de Kandi (2.93°E, 11.13°N). Les distributions des cumuls décennaux de pluie sur la période 1983-2015 des deux stations sont comparées. De l'observation de la figure 25, il ressort une parfaite ressemblance entre les deux séries en termes de distributions, et comme on devrait s'y attendre les cumuls décennaux de pluie à Gogounou sont inférieurs à ceux de Kandi. Nous avons vérifié si les données mergées reproduisent les faits climatiques majeurs en l'occurrence la sécheresse des années quatre-vingts observées sur presque toutes les stations du nord (figure 26).

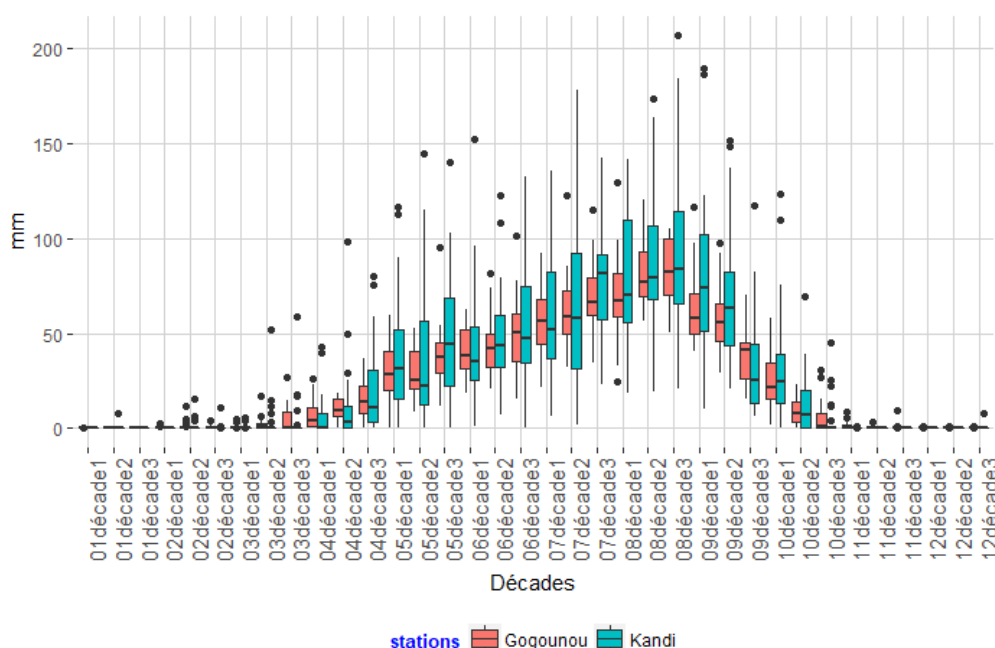


FIGURE 25 – Régime décennal de pluie à Gogounou et à Kandi (Période 1983 – 2015)

En effet, de l'observation de la figure 26 deux grandes tendances sont distinguées : une période d'années sèches persistante de 1983 – 1993 et une période d'années humides plus moins persistante de 1993 – 2015.

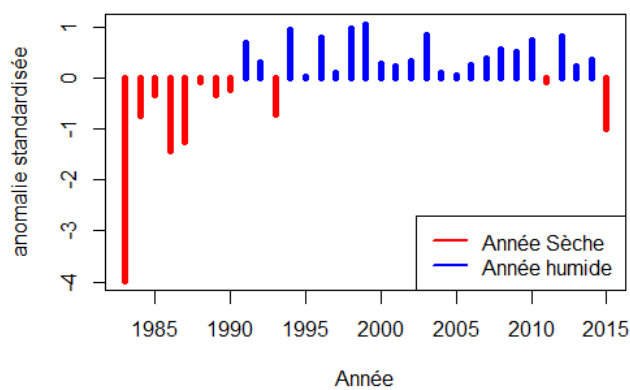


FIGURE 26 – Variabilité interannuelle des pluies à Gogounou.

## 7 Conclusion

Le but primordial de ce travail est de contribuer à la résolution du problème de disponibilité de données dans des zones d'intérêt pour le décideur. En effet, la couverture du réseau météorologique reste fluide pour la fourniture d'informations dans toutes les régions. L'approche proposée pour venir à bout de ce problème est la combinaison des données in-situ et des données satellitaires. Les données satellitaires présentant une bonne couverture spatiale ont été corrigées de leurs biais par rapport à l'observation et fusionnées aux données d'observations ; elles aussi purifiées de toutes valeurs aberrantes par des méthodes rigoureuses. La méthodologie utilisée pour merger les deux types de données est la Regression-Kriging.

Il ressort du processus, que les données mergées présentent de bonne performance pour les températures. En ce qui concerne les données de précipitations les performances sont meilleures dans les zones densifiées en postes pluviométriques. Dans les zones fluides, les données satellitaires brutes présentent des performances équivalentes aux produits mergés. Du réseau opérationnel au global, on note une certaine amélioration qualitative des champs de précipitations. Toutefois en terme de performance la différence n'est pas significative. Les interpolations de l'observation pour le réseau présente des lissages et il serait intéressant de produire à chaque carte élaborée par interpolation de produire celle de la variance d'erreur.

Ce travail présente quelques limites qu'ils importent d'évoquer. La proportion de données manquantes est une des premières limites pour le merging. En ce sens qu'elle réduit la série de données pour la correction de biais. La seconde limite est l'utilisation d'une seule méthode pour la correction de biais. La dernière et non la moindre est la non évaluation des interpolations de l'observation pour avoir une bonne comparaison avec les produits mergés.

En tenant compte des quelques limites citées plus haut, nous pouvons proposer en termes de perspectives pour cet travail :

- ✍ la calibration des algorithmes d'estimation de pluie TAMSAT sur le Bénin en utilisant tout le réseau pluviométrique ;
- ✍ l'utilisation de plusieurs méthodes de correction de biais enfin de choisir celles qui reproduit le mieux les observations ;
- ✍ la diversification des approches géostatistiques de fusion pour comparaison ;
- ✍ la fusion à d'autres échelles temporelles (pentadaires, décadaires, mensuelles, annuelles) afin de voir l'échelle qui produit de bonnes performances.

## Références

- Ajaaj, A., Mishra, A., & Khan, A. (2015). Comparison of bias correction techniques for gpcc rainfall data in semi-arid climate. *Stochastic Environmental Research and Risk Assessment*, 30.
- Arnaud, Y., Taupin, J.-D., & Laurent, H. (1996). Validation de deux méthodes d'estimation des pluies par satellite (EPSAT-Lannion et TAMSAT-Reading) à l'aide du réseau dense de pluviographes de l'expérience EPSAT-Niger. In *Problèmes de validation des méthodes d'estimation des précipitations par satellite en Afrique intertropicale*, Colloques et Séminaires (pp. 163–168). ORSTOM.
- Birkel, Sean D et Mayewski, P. (2015). *Analyse du climat historique et projeté du Mali, Sahel ouest-africain*. Technical report, CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).
- Cannon, A. J. (2008). Probabilistic multisite precipitation downscaling by an expanded bernoulli–gamma density network. *Journal of Hydrometeorology*, 9(6), 1284–1300.
- Chao, Y., Li, Z., Farrara, J. D., & Hung, P. (2009). Blending sea surface temperatures from multiple satellites and in situ observations for coastal oceans. *Journal of Atmospheric and Oceanic Technology*, 26(7), 1415–1426.
- Contreras, L. & Ferri, C. (2016). Wind-sensitive interpolation of urban air pollution forecasts. *Procedia Computer Science*, 80, 313 – 323. International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA.
- Delignette-Muller, M. & Dutang, C. (2015). fitdistrplus : An r package for fitting distributions. *Journal of Statistical Software*, 64, 1–34.
- Dinku, T., Block, P., Sharoff, J., Hailemariam, K., Osgood, D., del Corral, J., Cousin, R., & Thomson, M. C. (2014a). Bridging critical gaps in climate services and applications in africa. *Earth Perspectives*, 1(1), 15.
- Dinku, T., Hailemariam, K., Maidment, R., Tarnavsky, E., & Connor, S. (2014b). Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over ethiopia. *International Journal of Climatology*, 34(7), 2489–2504.
- Doan, C. D., Liu, J., Liong, S.-Y., Kim, D. E., Sanders, R., & Fewtrell, T. (2014). Combining satellite and gauge precipitation data with co-kriging method for jakarta region. In *Proceedings of the 11th International Conference on Hydroinformatics* New York, NY, USA : CUNY Academic Works.
- Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., Kamahori, H., Kobayashi, C., Endo, H., Soma, M., Oikawa, Y., & Ishimizu, T. (2011). The japanese 55-year reanalysis : An interim report. *SOLA*, 7, 149–152.
- Eischeid, J. K., Bruce Baker, C., Karl, T. R., & Diaz, H. F. (1995). The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology*, 34(12), 2787–2795.

- Faniriantsoa, R. & Dinku, T. (2017). Climate Data Tools (Version 4.5). <https://github.com/rijaf/CDT>. Consulté en mai 2018.
- Good, E. (2015). Daily minimum and maximum surface air temperatures from geostationary satellite data. *Journal of Geophysical Research : Atmospheres*, 120(6), 2306–2324.
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical note : Downscaling RCM precipitation to the station scale using statistical transformations — a comparison of methods. *Hydrology and Earth System Sciences*, 16(9), 3383–3390.
- Hansen, J., Hellmuth, M., Thomson, M., & Williams, J. (2006). *A gap analysis for the implementation of the Global Climate Observing System Programme in Africa*. Technical report, IRI Technical Report No. TR/06/1. International Research Institute for Climate and Society, Palisades, NY.
- Hengl, T., Heuvelink, G., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression - kriging. *Geoderma*, 120(1-2-2017), 75–93.
- Hengl, T., Heuvelink, G. B., & Rossiter, D. G. (2007). About regression-kriging : From equations to case studies. *Computers & Geosciences*, 33(10), 1301 – 1315. Spatial Analysis.
- Hubbard, K. G., Goddard, S., Sorensen, W. D., Wells, N., & Osugi, T. T. (2005). Performance of quality assurance procedures for an applied climate information system. *Journal of Atmospheric and Oceanic Technology*, 22(1), 105–112.
- Hubbard, K. G. & You, J. (2005). Sensitivity analysis of quality assurance using the spatial regression approach—a case study of the maximum/minimum air temperature. *Journal of Atmospheric and Oceanic Technology*, 22(10), 1520–1530.
- Ines, A. V. & Hansen, J. W. (2006). Bias correction of daily gcm rainfall for crop simulation studies. *Agricultural and Forest Meteorology*, 138(1), 44 – 53.
- J. Willmott, C. & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79.
- Krishnamoorthy, K., Mathew, T., & Mukherjee, S. (2008). Normal-based methods for a gamma distribution. *Technometrics*, 50(1), 69–78.
- L. Bernstein, R. (1982). Sea surface temperature estimation using the noaa 6 satellite advanced very high resolution radiometer. *Journal of Geophysical Research*, 87, 9455–9465.
- Legates, D. R. & McCabe, G. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35, 233–241.
- Lorente-Plazas, R., García-Díez, M., Jimenez-Guerrero, P., Fernández, J., & Montavez, J. P. (2014). On the added value and sensitivity of wrf to driving conditions over cordex-africa domain. In *EGU General Assembly Conference Abstracts*, volume 16 of *EGU General Assembly Conference Abstracts* (pp. 10889).

- Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L. A., & Rodríguez-Sánchez, J.-P. (2016). High-resolution satellite-gauge merged precipitation climatologies of the tropical andes. *Journal of Geophysical Research : Atmospheres*, 121(3), 1190–1207.
- Merchant, C. J., Harris, A. R., Maturi, E., Embury, O., MacCallum, S. N., Mittaz, J., & Old, C. P. (2009). Sea surface temperature estimation from the geostationary operational environmental satellite-12 (goes-12). *Journal of Atmospheric and Oceanic Technology*, 26(3), 570–581.
- Midingoyi, A. C. (2016). Approche géostatistique d’optimisation du réseau pluviométrique du Bénin pour un meilleur suivi de la campagne agricole. Master’s thesis, Chaire Internationale en Physique Mathématique et Applications : Cotonou, 75p.
- National Center for Atmospheric Research Staff (2014). The climate data guide : Regridding overview. <https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/regridding-overview>. Consulté en mai 2017.
- Novella, Nick et Thiaw, W. (2010). Validation des produits de pluie dérivés de satellites sur le sahel. *Wyle Information Systems / CPC / NOAA*, (pp. 1–9).
- Pebesma, E. (2004). Multivariable geostatistics in s : the gstat package. <http://www.gstat.org/>. Consulté en mai 2018.
- Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J., & Dash, P. (2014). Evaluation and selection of sst regression algorithms for jpss viirs. *Journal of Geophysical Research : Atmospheres*, 119(8), 4580–4599.
- Piani, C., Haerter, J. O., & Coppola, E. (2010a). Statistical bias correction for daily precipitation in regional climate models over europe. *Theoretical and Applied Climatology*, 99(1), 187–192.
- Piani, C., Weeden, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., & Haerter, J. O. (2010b). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology*, 395(3), 199 – 215.
- PNUD (2009). *Quatrième Rapport National du Bénin sur la Convention des Nations Unies sur la diversité biologique*. Technical report, Programme des Nations Unies pour le Développement.
- R Project for Statistical Computing (2018). R. <https://www.r-project.org/>. Consulté en mai 2018.
- Reynolds, R. W. (1988). A real-time global sea surface temperature analysis. *Journal of Climate*, 1(1), 75–87.
- Reynolds, R. W. & Smith, T. M. (1994). Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate*, 7(6), 929–948.
- Scherrer, S. C., Geijtenbeek, D. V., Naguel, C., Croci-Maspoli, M., & Appenzeller, C. (2010). Operational quality control of daily precipitation using spatio-climatological plausibility testing. In *European Conference on Applied Climatology Zurich*, Switzerland.



- Schulz, M. (2002). The standardization of mean-squares. *Rasch Measurement Transactions*, 16(2), 878–879.
- Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets*. PhD thesis, University of Pittsburgh.
- Sivakumar, M. (1988). Predicting rainy season potential from the onset of rains in southern sahelian and sudanian climatic zones of west africa. *Agricultural and Forest Meteorology*, 42(4), 295 – 305.
- Smith, T. M., Livezey, R. E., & Shen, S. S. (1998). An improved method for analyzing sparse and irregularly distributed sst data on a regular grid : The tropical pacific ocean. *Journal of Climate*, 11(7), 1717–1729.
- Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R. P., Stringer, M., Chadwick, R., & Kayitakire, F. (2014). Extension of the tamsat satellite-based rainfall monitoring over africa and from 1983 to present. *Journal of Applied Meteorology and Climatology*, 53(12), 2805–2822.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research : Atmospheres*, 106(D7), 7183–7192.
- Waongo, M. (2017). *Cours d'évapotranspiration potentielle pour les étudiants de 1ère année du Département Formation et Recherche Centre Régionale AGRHYMET*. Fascicule de cours, FA, Niamey, 112 p.

## Annexes

### Annexes A : Détails sur la détermination du facteur multiplicateur f

le facteur multiplicatif  $f$  a été déterminé en utilisant un test ayant comme hypothèse nulle ( $H_0$ ) la non existence de valeurs aberrantes dans les données. Afin de rejeter l'hypothèse nulle, un degré de risque  $\alpha$  est donné en termes de valeurs critiques et/ou de niveau de confiance. En effet, les scores Z sont des mesures d'écart-type et les p-values des probabilités, les deux sont des statistiques associées avec la distribution normale standard.

Cette distribution relie les écarts-types aux probabilités et permet d'attribuer une signification et une confiance aux scores Z et aux p-values (Figure 1).

Dans le cas d'espèce, en fixant le risque  $\alpha$  que nous convenons négligeable  $\alpha = 0.05$ . Il en résulte un certain intervalle  $[-f, f]$  dans lequel Z (notre variable décision) a une probabilité  $(1-\alpha)$  de tomber si l'hypothèse nulle est exacte et, par conséquent, hors duquel Z a une probabilité  $\alpha$  petite de tomber. Autrement dit : Si  $H_0$  est vraie,

$$\mathcal{P}(|Z| < f) = 1 - \alpha \quad (1)$$

(1) peut aussi s'écrire  $\mathcal{P}(-f < Z < f) = 1 - \alpha \Rightarrow$

$$\mathcal{P}\left(-f < \frac{(X_i - Q_{50})}{IQR} < f\right) = 1 - \alpha \Rightarrow$$

$$\Rightarrow \mathcal{P}(-f * IQR + Q_{50} < X_i < f * IQR + Q_{50}) = 1 - \alpha \Rightarrow$$

$$\Rightarrow F(f * IQR + Q_{50}) - F(-f * IQR + Q_{50}) = 1 - \alpha$$

où  $F$  est la fonction cumulative de distribution (fonction de répartition) de la série de notre variable climatique. En introduisant l'inverse de  $F$  dans cette dernière ligne on a :

$$F^{-1}(F(f * IQR + Q_{50}) - F(-f * IQR + Q_{50})) = F^{-1}(1 - \alpha) \Rightarrow$$

$$\Rightarrow (f * IQR + Q_{50}) - (-f * IQR + Q_{50}) = F^{-1}(1 - \alpha) \Rightarrow 2 * f * IQR = F^{-1}(1 - \alpha)$$

$$\text{or } IQR = Q_{75} - Q_{25} = F^{-1}(0.75) - F^{-1}(0.25)$$

On déduit donc f :

$$f = \frac{F^{-1}(1 - \alpha)}{2(F^{-1}(0.75) - F^{-1}(0.25))} \quad (2)$$

où  $F^{-1}$  est la Fonction de Distribution Cumulative inverse (ou fonction quantile) pour la série de notre variable climatique, supposée distribuée normalement dans toute la démonstration. Le niveau de confiance choisi pour tout le processus de contrôle est de 95%.

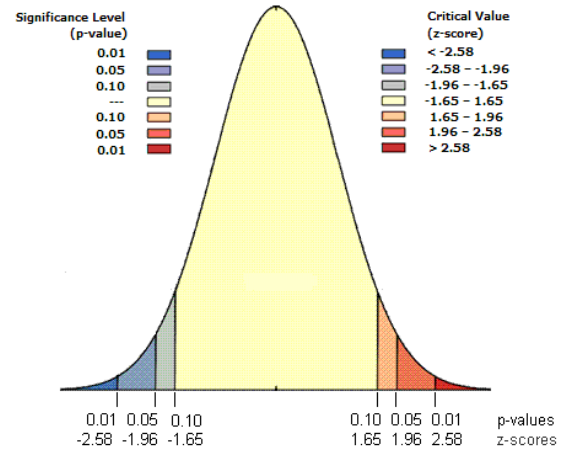
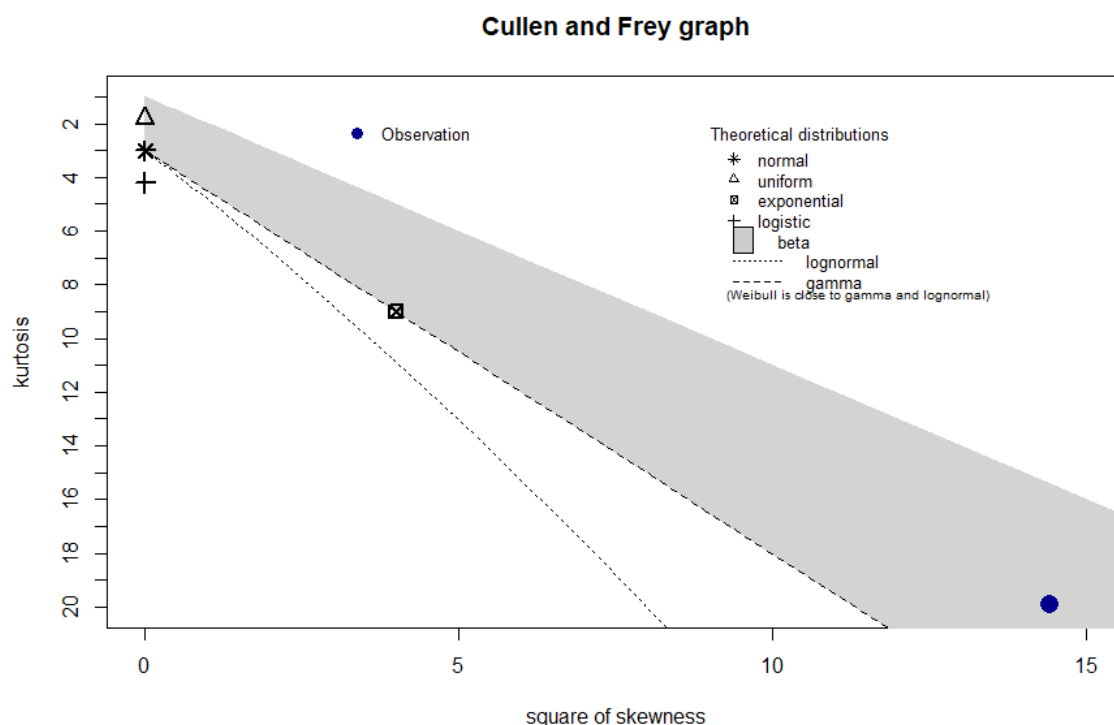
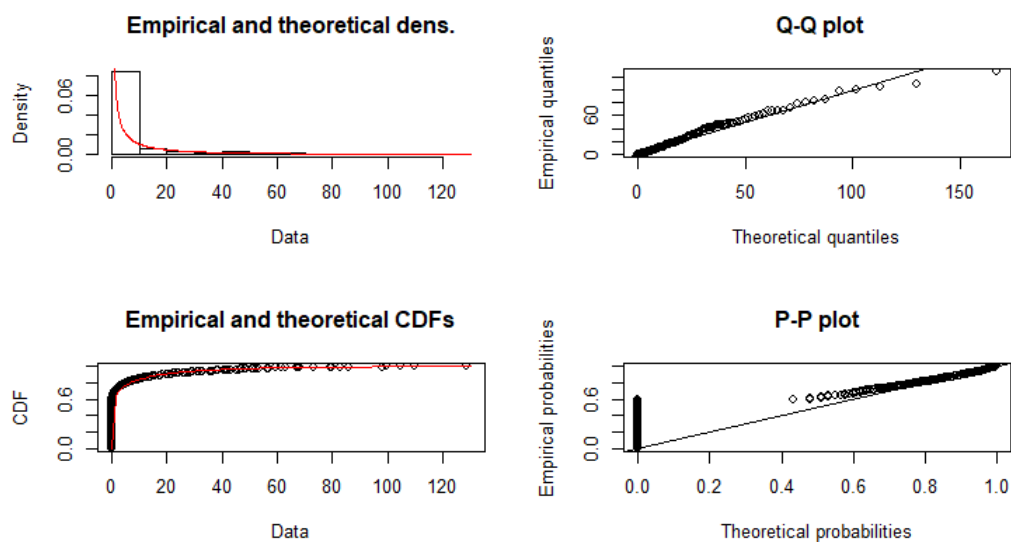


FIGURE 1 – Fonction de densité de la loi normale

## Annexes B : Distribution théorique à laquelle nos séries de paramètres climatiques s'ajustent le mieux

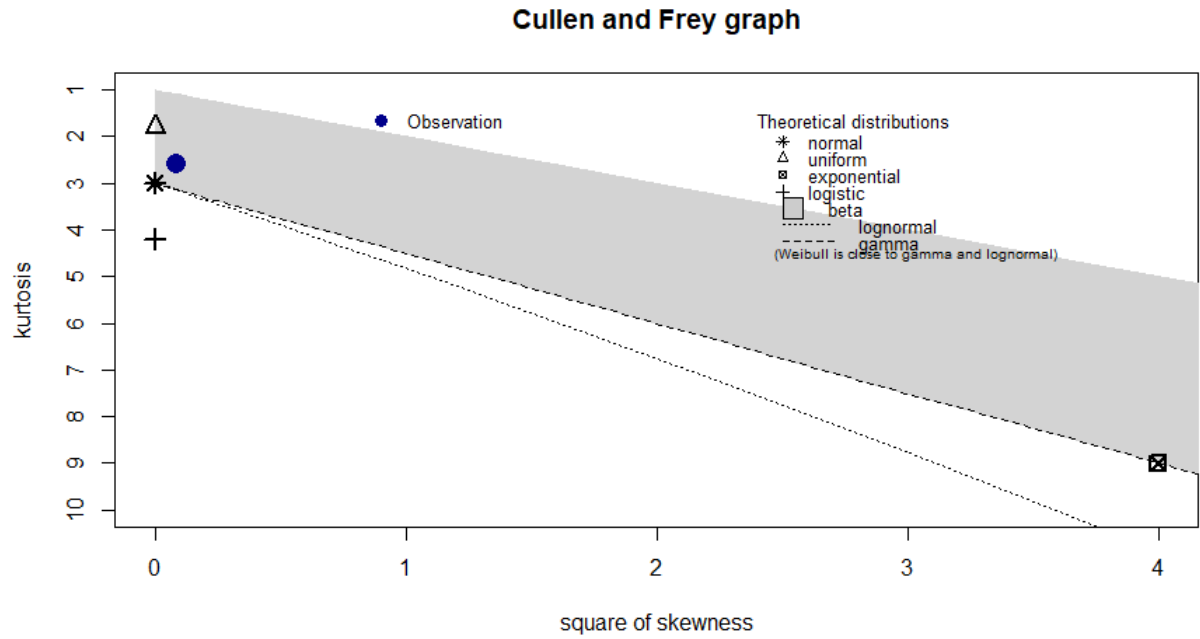


(a) Diagramme de cullen & Frey

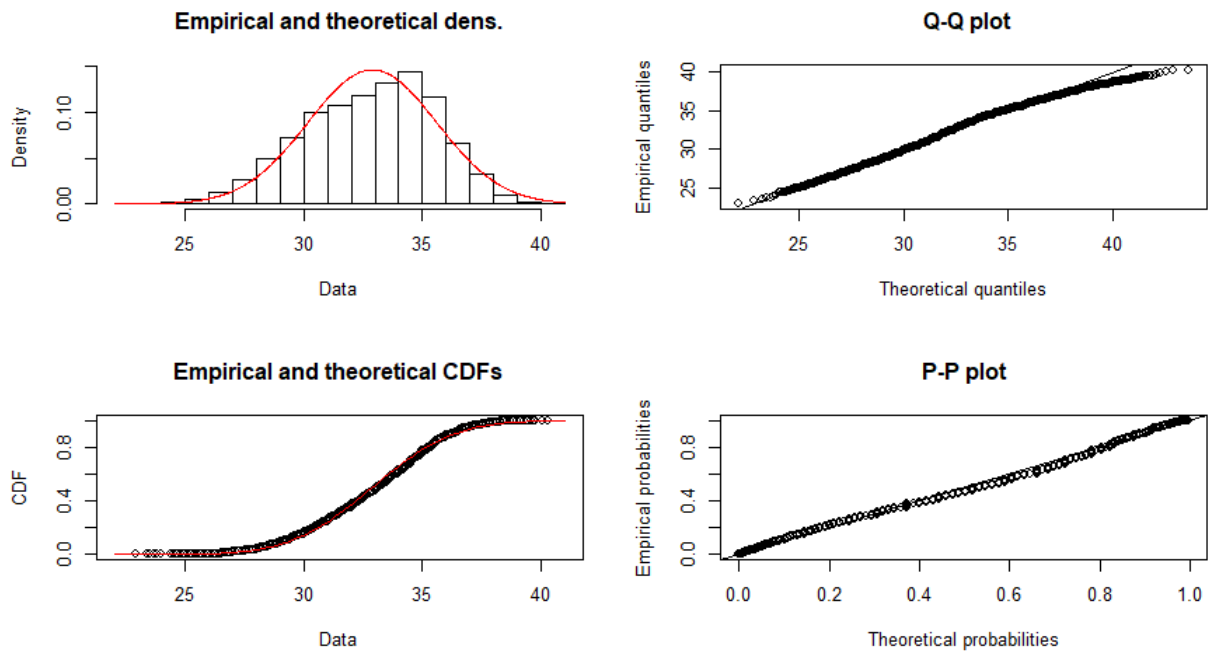


(b) Courbes de la fonction de distribution cumulative et quantile-quantile (rouge distribution théorique)

FIGURE 2 – Ajustement des séries journalières de pluies des mois de juillet sur la période 1981-2010



(a) Diagramme de cullen & Frey



(b) Courbes de la fonction de distribution cumulative et quantile-quantile (rouge distribution théorique)

FIGURE 3 – Ajustement des séries journalières de températures des mois de juillet sur la période 1981-2010

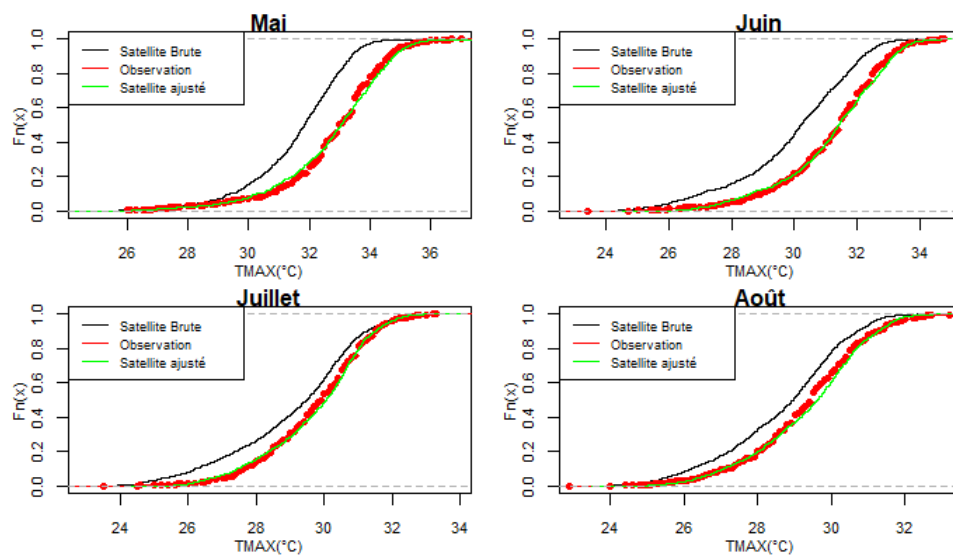
## Annexes C : Extrait des résultats de contrôle de qualité

Tableau 1 – Extrait de données aberrantes

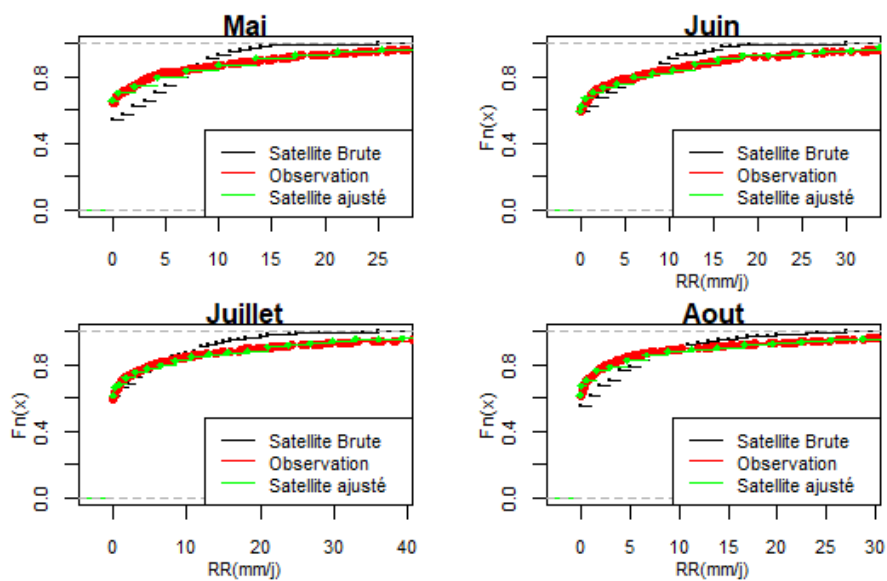
Stations	Paramètres	Dates	Valeurs observées	Observations
Abomey	RR	08071990	138 mm	VA
Abomey	RR	25062010	150 mm	VA
Adjohoun	RR	14071991	111 mm	VA
Adjohoun	RR	01102002	48 mm	TLDD
Bopa	RR	22091987	192 mm	VA
Bopa	RR	17011988	63 mm	TLDD
Dangbo	RR	27062012	144 mm	VA et TLDD
Kandi Aéroport	RR	21081989	90 mm	VA
Kounde	RR	18062013	95 mm	VA
Malanville	RR	06071982	132 mm	VA
Bantè	TMAX	12011996	40 °C	VA
Bantè	TMAX	17062013	39 °C	VA
Bantè	TMAX	03081983	23 °C	VA
Dassari	TMAX	08021998	49°C	VA
Ouidah	TMAX	03071981	36 °C	VA
Tchaourou	TMAX	05112014	23 °C	VA
Dassari	TMIN	25011991	27 °C	VA
Ouidah	TMIN	22012008	13 °C	VA

RR : Précipitations, TMAX : Températures maximales, TMIN : Températures minimales, VA : Valeurs aberrantes, TLDD : Trop large déviation au-dessus.

## Annexes D : Extraits des résultats de correction de biais des données satellitaires brutes



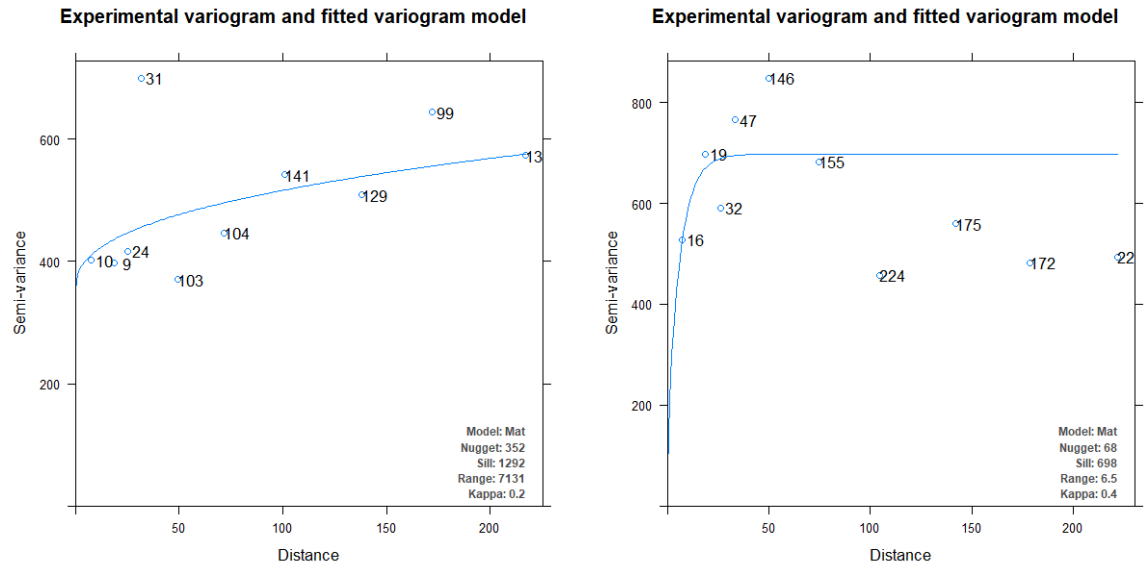
(a) Températures maximales



(b) Précipitations

FIGURE 4 – Courbes cumulatives de distributions des données satellitaires brutes, ajustées et des données in-situ

# Annexes E : Variogramme



(a) Variogrammes du réseau opérationnel

(b) Variogrammes du réseau global

FIGURE 5 – Variogrammes utilisés pour l’interpolation des données d’observations