Données utilisées - Matériels traitement de données

III.1 Données de précipitations [38]

Les données de précipitations utilisées dans ce présent travail ont une résolution temporelle respective de 1 mois et s'étalent sur une période allant de 1985 à 2015. Ces données sont issues de la base de données de la Direction Générale de la Météorologie malagasy. Les séries temporelles de précipitations sont produites en combinant des données in situ de stations de Météo malagasy et des précipitations estimées par données satellitaires, après contrôle de qualité. Les données pluviométriques des stations de Beloha, Bekily, Tsihombe et d'Ambovombe ont été utilisées.

III.2 Données TSM de AVHRR Pathfinder

Depuis les années 1980, et grâce aux satellites, il est possible de mesurer indirectement la TSM. Les ensembles de données TSM AVHRR Pathfinder sont utilisés dans cette étude. Cet ensemble de données contient des moyennes mensuelles et annuelles de la température de la surface de la mer allant de 1985 à 2015. Dotées d'une résolution spatiale de 4km, ces données sont en format GeoTIFF. Elles sont affectées de la projection GCS_WGS_1984.

III.3 Les indices climatiques

III.3.1 SIOD [39]

Les valeurs des indices du dipôle subtropical de l'océan Indien (SIOD) sont disponibles sur le site web du JAMSTEC (Japan Agency for Marine-Earth Science and Technology). Ces indices sont calculés à partir de la différence entre les anomalies mensuelles de la TSM de la région (55°E à 65°E ; 37°S à 27°S) et de la région (90°E à 100°E ; 28°S à 18°S) dans le sud de l'océan Indien.

III.3.2 Oceanic Nino Index (ONI) [40]

L'indice océanique Nino (ONI) est l'un des principaux indices utilisés pour surveiller l'oscillation australe El Niño (ENSO). L'ONI est calculé en faisant la moyenne des anomalies de température de la surface de la mer dans une zone de l'océan Pacifique équatorial centre-est, appelée région Nino-3.4 (5°S à 5°N, 170°W à 120°W). De plus, une moyenne sur trois mois (moyenne mobile) est calculée afin de mieux isoler la variabilité étroitement liée au phénomène ENSO.

Dans ce cas, les événements sont définis comme 5 valeurs consécutives d'anomalies égales ou au-dessus de $+0.5^{\circ}\text{C}$ pour El Niño, inférieur ou égal à $-0,5^{\circ}\text{C}$ pour La Niña. Entre $-0,5^{\circ}\text{C}$ et $+0,5^{\circ}\text{C}$ pour les événements neutres. L'ONI est disponible sur le site de la NOAA.

III.4 Analyse de données

R est un langage de programmation d'analyse statistique, ayant pour but de pouvoir organiser et traiter des jeux de données afin d'y appliquer des tests statistiques et de représenter ces données graphiquement. [41]

Il permet de manipuler avec flexibilité différents types de données tels que les fichiers spatiaux et les séries temporelles grâce à ses nombreux packages, disponibles gratuitement sur internet. Des scripts R sont élaborés afin de pouvoir réaliser les différents traitements et tests statistiques sur chaque point de grille.

III.5 Outils de visualisations

L'utilisation d'un logiciel SIG est nécessaire pour faciliter la visualisation des résultats. Ainsi, les analyses thématiques et opérations de spatialisations sont faites sous ArcGIS version 10.3.1.

CHAPITRE 4 : METHODOLOGIES

IV.1 Statistiques descriptives [28]

Les traitements que nous avons effectués sur les données de la TSM portent sur des opérations simples (moyenne, écart-type) pour établir les caractères climatologiques du TSM au sud de Madagascar. La climatologie des ensembles de données de la TSM est calculée pour chaque point de grille en faisant la moyenne des données de la TSM au cours de 1985-2015.

En effet, la moyenne d'une série est la valeur centrale d'une série des valeurs données. Elle s'obtient en faisant la somme de tous les éléments de cette série, divisée par le nombre d'éléments composant cette série.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Avec x_i : une série composée de n données

$\bar{x}$: la moyenne de la série

n : effectifs, nombres d'éléments composant la série

L'écart-type (σ_x) représente la dispersion d'une valeur prise dans une série par rapport à la moyenne de cette série. Et lorsque l'écart-type est grand, cela veut dire qu'il y a une forte dispersion de la série par rapport à la moyenne. Et sinon cela implique le contraire.

$$\sigma_x = \sqrt{\left[\frac{1}{n} \left(\sum_{i=1}^n x_i^2 \right) - ((\bar{x})^2) \right]}$$

Diagramme de Hovmöller

La coupe longitude-temps ou latitude-temps est une représentation graphique des conditions météorologiques observées sur une bande de temps selon les longitudes ou latitudes considérées. Elle donne donc une vue dans l'évolution temporelle d'un paramètre étudié. En d'autres termes, c'est une des représentations graphiques des séries temporelles. On a appliqué cette méthode sur la série de la TSM.

IV.3 Anomalies

L'anomalie (a_i) représente ce qui s'écarte de la normale. Elle s'obtient en faisant la différence de la valeur d'un élément dans une série considérée (x_i) avec la moyenne de cet échantillon ($\bar{x}$). L'anomalie sert à montrer l'écart des éléments d'un échantillon par rapport à la moyenne et est calculée par la formule suivante :

$$a_i = x_i - \bar{x}$$

IV.4 Indice de Précipitation Standardisée (IPS) [29]

Développé en 1993, l'Indice de Précipitation Standardisée encore appelé « IPS » permet de caractériser les périodes humides et les périodes sèches. C'est un indice à la fois puissant, souple d'utilisation et simple à calculer, ayant pour unique donnée nécessaire à son calcul les précipitations. Il peut être employé à différentes échelles de temps allant de 1 à 72 mois. Pour le calculer, il est préférable d'avoir en possession des recueils de données mensuels des précipitations allant au moins sur 20 à 30 ans, mais de préférence sur 50 à 60 ans.

Les valeurs positives de l'IPS indiquent des précipitations supérieures à la médiane et les valeurs négatives, des précipitations inférieures à la médiane.

Sa formulation pour le mois i à l'année j est :

$$IPS_{ij} = \frac{RR_j(i) - RR_N(i)}{\sigma_i}$$

Avec :

IPS_{ij} : IPS pour le mois i à l'année j

$RR_j(i)$: précipitation (en mm) du mois i à l'année j

$RR_N(i)$: Précipitation normale (en mm) du mois i à l'année j

σ_i : Ecart-type des précipitations du mois i

Compte tenu de la climatologie de la précipitation sur chaque district (Figures 6,8,10,12) exposé dans la partie I, section I.6, la plupart des précipitations dans la région Androy se produisent durant l'Asara. Les quantités moyennes supérieures à 40mm se produisent de décembre à mars. On a pu constater que le maximum se présente chaque mois de janvier. La connaissance des divers facteurs contribuant à la modulation de ce paramètre au cours de cette saison s'avère alors d'une importance capitale. Le déficit ou l'excès du taux de précipitation aura des conséquences majeures sur la population et les organismes vivants. Ainsi la surveillance de cette période permettra donc de se préparer à toutes formes d'anomalies de précipitations pouvant se produire. Ainsi, les différentes analyses entreprises sur chaque district se portent en plein cœur de l'Asara, en décembre à janvier (DJF) pour l'IPS-3 et de décembre à mars (DJFM) pour l'IPS-4.

IV.5 Relation du TSM avec ENSO et SIOD

Pour deux grandeurs statistiques X et Y quantitatifs observés. L'analyse de corrélations détermine s'il existe une relation entre les variables X et Y. Elle permet aussi de caractériser la forme de la liaison (la relation) entre ces deux grandeurs (positive ou négative, linéaire ou non linéaire, monotone ou non monotone) et de tester ensuite si la liaison est statistiquement significative afin de quantifier l'intensité de cette liaison. **[30]**

Le coefficient de corrélation linéaire de Pearson a été calculé localement pour étudier la relation entre ENSO par rapport à la TSM et l'indice de SIOD avec la TSM.

Le coefficient de corrélation linéaire est donné par la formule suivante :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

La valeur de ce coefficient varie entre -1 et +1. Elle est proche de 0 pour des variables dont la variation est indépendante l'une de l'autre, positive (respectivement négative) pour les valeurs qui varient dans le même sens (dans le sens contraire). **[31]**

Le coefficient de corrélation entre ENSO et TSM (respectivement SIOD et TSM) est calculé pour chaque mois. Pour séries de données de la TSM, l'opération de centrage et réduction a été effectuée. À noter que la valeur négative de la TSM correspond à une TSM froide et la valeur positive de la TSM marque la signature d'une TSM chaude.

IV.6 Relation entre TSM et précipitation

Deux méthodes ont été utilisées afin de déterminer la relation entre précipitations et TSM. Elles ont été appliquées sur la série de précipitation normalisée IPS-3 (respectivement IPS-4) et les données de la TSM qui sont déjà centrées et réduites. Il s'agit d'une méthode d'analyse de corrélation et d'analyse composite sur la période de 1985- 2015.

IV.6.1 Corrélation entre TSM avec IPS-3 et IPS-4

Le coefficient de corrélation de Spearman est utilisé pour mesurer la liaison entre TSM et IPS-3 (respectivement TSM et IPS-4). Ce coefficient mesure l'intensité de la liaison entre deux caractères quantitatifs.

Il est défini comme suit, pour deux séries de variables X et Y :

$$r = \frac{\sum_{i=1}^n R_{X_i} R_{Y_i} - N}{\sqrt{(\sum_{i=1}^n R_{X_i}^2 - N)(\sum_{i=1}^n R_{Y_i}^2 - N)}} \quad \text{où } N = n \left(\frac{n+1}{2} \right)^2$$

Avec :

R_{X_i} : Rang de la valeur X_i au sein de X

R_{Y_i} : Rang de la valeur Y_i au sein de Y

n : nombre de valeur

Le coefficient de corrélation de Spearman se base sur le rang des valeurs dans les séries. Il a été utilisé puisqu'il est adapté pour le cas où les variables ne suivent pas la loi de distribution normale (cas des séries de précipitations).

La valeur du coefficient varie entre -1 et +1, plus r est proche de -1 ou de +1, plus la corrélation est forte et plus les deux variables sont liées. Si r est positif, les deux variables évoluent dans le même sens, mais si r est négatif, les deux variables évoluent en sens inverse. Les coefficients de corrélations obtenues sont ensuite testés avec un seuil de risque $\alpha = 0,05$ selon l'hypothèse nulle (H_0) que le coefficient de Spearman est égal à zéro. [30]

Cependant, la relation statistique entre TSM et IPS-3 (respectivement TSM et IPS-4) est faite comme suit :

- Pour chaque district, IPS au cours de la saison dite « Asara » sont utilisés, c'est-à-dire précipitation de décembre à février (DJF) pour IPS-3 et précipitation de décembre à mars (DJFM) pour IPS-4.
- Coefficient de corrélation entre TSM et IPS-3 (respectivement TSM et IPS-4) est calculé et le résultat obtenu est ensuite testé.

- Le coefficient de corrélation est d'abord calculé sans décalage temporel, c'est-à-dire, on calcule le coefficient de corrélation entre TSM et IPS-3 (respectivement TSM et IPS-4) au cours de la période de décembre à février (DJF) (respectivement de décembre à mars (DJFM)). Alors on va noter comme suit : [lag-0, (TSM : DJF, IPS-3 : DJF)] (respectivement [lag-0, (TSM : DJFM, IPS-4 : DJFM)])
- Pour des objectifs opérationnels en matière de prévision saisonnière des précipitations à l'échelle des districts de la région Androy, des corrélations avec des décalages négatifs de la TSM ont été faites par rapport aux précipitations. En terme plus simple, on fixe la valeur de l'IPS-3 de DJF (respectivement IPS-4 de DJFM) et la valeur de la TSM varie selon le décalage opéré au cours des différents calculs.

Ainsi ces décalages temporels se font comme suit pour chaque cas :

Cas IPS-3 :

Décalage de 1 mois, pour la TSM de septembre à novembre : [lag-1, (TSM : SON, IPS-3 : DJF)]
 Décalage de 2 mois, pour la TSM d'août à octobre : [lag-2, (TSM : ASO, IPS-3 : DJF)]
 Décalage de 3 mois, pour la TSM de juillet à septembre : [lag-3, (TSM : JAS, IPS-3 : DJF)]
 Décalage de 4 mois, pour la TSM de juin à août : [lag-4, (TSM : JJA, IPS-3 : DJF)]
 Décalage de 5 mois, pour la TSM de mai à juillet : [lag-5, (TSM : MJJ, IPS-3 : DJF)]
 Décalage de 6 mois, pour la TSM d'avril à juin : [lag-6, (TSM : AMJ, IPS-3 : DJF)]
 Décalage de 7 mois, pour la TSM de mars à mai : [lag-7, (TSM : MAM, IPS-3 : DJF)]
 Décalage de 8 mois, pour la TSM de février à avril : [lag-8, (TSM : FMA, IPS-3 : DJF)]
 Décalage de 9 mois, pour la TSM de janvier à mars : [lag-9, (TSM : JFM, IPS-3 : DJF)]

Cas IPS-4 :

Décalage de 1 mois, pour la TSM d'août à novembre : [lag-1, (TSM : ASON, IPS-4 : DJFM)]
 Décalage de 2 mois, pour la TSM de juillet à octobre : [lag-2, (TSM : JASO, IPS-4 : DJFM)]
 Décalage de 3 mois, pour la TSM de juin à septembre : [lag-3, (TSM : JJAS, IPS-4 : DJFM)]
 Décalage de 4 mois, pour la TSM de mai à août : [lag-4, (TSM : MJJA, IPS-4 : DJFM)]
 Décalage de 5 mois, pour la TSM d'avril à juillet : [lag-5, (TSM : AMJJ, IPS-4 : DJFM)]
 Décalage de 6 mois, pour la TSM de mars à juin : [lag-6, (TSM : MAMJ, IPS-4 : DJFM)]
 Décalage de 7 mois, pour la TSM de février à mai : [lag-7, (TSM : FMAM, IPS-4 : DJFM)]
 Décalage de 8 mois, pour la TSM de janvier à avril : [lag-8, (TSM : JFMA, IPS-4 : DJFM)]

IV.6.2 Analyse composite [31]

La méthode des composites est utilisée pour voir si un signal préalablement identifié dans un premier paramètre z apparaît dans un second paramètre v . Deux échantillons (les composites) contenant les valeurs prises par v quand z connaît respectivement des anomalies négatives et positives marquées ont été construites. On conclut ensuite à l'existence statistique du signal dans v en testant la différence des moyennes des deux échantillons grâce au test en t de Student. Il est recommandé de faire des analyses avec différents seuils d'anomalies pour z , afin de s'assurer que les composites représentent bien l'état pris par v pour tout signal dans z . De même que pour les téléconnexions décalées, il est possible de considérer z et v à deux pas de temps différents et de travailler de façon partielle en supprimant ou gardant dans les échantillons les observations comportant un signal particulier (Montecinos et al., 2000 ; Camberlin et Philippon, 2002 entre autres).

L'un des avantages de la méthode des composites par rapport, notamment, à la méthode des corrélations linéaires est qu'aucune hypothèse n'est faite sur le type de relation entre les deux paramètres z et v étudiés : cette relation peut tout aussi bien être linéaire que non linéaire. Par contre, on suppose que les anomalies de z sont toujours associées ou dues aux mêmes types d'anomalies dans v . D'autre part, le composite, fondé sur la comparaison des moyennes, est plus robuste, moins sensible aux valeurs extrêmes.

Soit deux paramètres A et B . La valeur du signal de A est notée a . Alors, d'après le principe de l'analyse composite, on va former un échantillon Θ de l'indice a pour avoir la composite B_Θ qui est donnée par la formule empirique :

$$B_\Theta = E(B_t \mid a_t \in \Theta)$$

Mais en pratique, l'espérance est remplacée par une moyenne des valeurs de B , correspondant à $a \in \Theta$ pour obtenir le composite. L'équation devient alors :

$$\widehat{B}_\Theta = \frac{1}{k} \sum_{j=1}^k B_{t_j}$$

Avec : $t_1, \dots, t_k$ pour $a_{t_j} \in \Theta$

Mis en œuvre de l'analyse :

Ce sont les valeurs de l'IPS-3 (respectivement IPS-4) de décembre-janvier-février (respectivement de décembre-janvier-février-mars) qui seront considérées afin de choisir les années prises en comptes pour l'élaboration des composites.

Nous allons distinguer la différence entre TSM au cours des évènements déficitaires et excédentaires sur chaque district. Une condition a été mise en place afin d'exclure les années avec des précipitations normales. Lorsque l'IPS-3 ou l'IPS-4 est supérieur à 1, on liste cette année dans la classe excédentaire et quand l'IPS-3 ou l'IPS-4 est inférieur à -1, on met cette année dans la classe des périodes déficitaires.

Les différences des moyennes des variables (TSM) en fonction des évènements sont testées par le test de Student pour évaluer la signification statistique des résultats obtenus.

On procède de la même manière comme on l'a fait avec l'analyse de corrélation, c'est-à-dire qu'on a recours à des opérations de décalage.