

Variabilité pluviométrique à Haïti: constitution de la base de données mensuelles et résultats préliminaires

Vincent Moron
Aix-Marseille Université & CEREGE, UM 34 CNRS

15 Novembre 2012

1. Concaténation des 3 bases de données pluviométriques

Un premier travail a consisté à vérifier individuellement chaque base de données. Ensuite, les données ont été intercomparées et éventuellement complétées, notamment grâce à la base de Chemonics. Toutefois, la base "CNIGS-DB" a été considérée comme la plus fiable et la plus complète *a priori* car tous les mois n'y sont pas systématiquement renseignés alors que de nombreux doutes existent à propos de la base "Chemonics" avec des zéros en grand nombre sur des mois où cela semble peu réaliste. *A priori*, il n'existe pas de moyens objectifs de discerner le vrai du faux dans ce cas, à part réaliser une inter-comparaison avec des stations proches mais cela demande un gros travail analytique afin de sélectionner d'abord les stations les plus fiables (mais sur quels critères ?), puis estimer le degré de vraisemblance pour chaque unité temporelle/station. Enfin, la base de données CIAT contient à mon avis les pluies TRMM à la résolution de 0.25° (ce sont des données satellites multi-source calibrée avec des stations pluviométriques. Leur qualité est variable dans la zone tropicale. Cf Huffmann et al., 200 ?).

1.1. La base CNIGS_DB

La base de données CNIGS_DB comporte 165 stations couvrant des segments variables de 1863 à 1989. 3 stations n'ont pas de coordonnées et pas de noms et sont pour l'instant inutilisables dans la plupart des analyses. Il s'agit des stations #211030M.csv¹, #605100M.csv et #703070M.csv. C'est dommageable surtout pour la station #605100M.csv qui est longue et comporte peu de manques. Il serait donc très utile de pouvoir la nommer et la localiser.

Un premier test a été de vérifier les doublons sur une période commune afin de valider l'enregistrement. Il s'agit de "Grande Rivière du Nord" (#105020M.csv et #105070M.csv), "Port de Paix" (#207010M.csv et #207090M.csv), "St Louis du Nord" (#208010M.csv et #208020M.csv), "Camp Perrin" (#605020.csv et #605030.csv). Dans tous ces cas, les données sont identiques dans les deux fichiers à 1 mm près et on peut considérer qu'il s'agit donc de la même station. Dans le cas de "Saltadère" (#302810M.csv et #302840M.csv), il n'y a pas de période commune et il est donc impossible de dire si il s'agit de la même station *a priori* mais elle sera confondue dans ce qui suit. Dans le cas de "Saint Marc" (#304010M.csv et #304020M.csv), la correspondance est correcte mais les données semblent fausses après 1982 car il y a alors des doublons temporels et les données d'un même mois ne correspondent pas. Cette station sera vérifiée avec la base "Chemonics" Il semble donc prudent de ne pas considérer les données de "Saint Marc" après 1982 dans cette base. Ce type de problème concernent aussi de façon plus générale les enregistrements de "Pétion Ville" (#404060M.csv) et "Les Cayes" (#605010.csv). Dans ces deux cas, de nombreux mois se répètent dans la base avec des valeurs très différentes (> 10 mm d'écart entre les valeurs affichées pour le même mois) Il y a parfois 3 fois le même mois, parfois 4. Ces deux stations sont donc inutilisables en l'état. Certaines stations comportent aussi des redoublements de certains mois, mais avec des valeurs quasiment égales (à 1 mm près) et ne sont pas considérées comme fausses *a priori*. On retrouve ce cas par exemple à "Ganthier" (#407030M.csv) ou "Chantal" (#604010M.csv).

Enfin, certaines stations présentent un taux élevé de valeur nulle la dernière année, même durant la

1 Les stations sont identifiées à partir du nom du fichier correspondant.

saison des pluies, et cette dernière année est donc considérée comme douteuse. Cela concerne "Limbe" (1987, #103020M.csv), "St Louis du Nord" (1988, #208020M.csv), "Leogane" (1988, #401010.csv), "Serrey" (1966, #402010M.csv), "Bourdon" (1989, #404040M.csv), "Juvenat" (1988, #404050.csv). Il semble prudent de ne pas tenir compte de l'année dans ce cas même si un total nul n'est pas forcément aberrant. Ces données seront vérifiées dans la mesure du possible dans la base "Chemonics" (cf. Section 1.4).

1.2. La base CNSA

Cette base comporte 12 stations couvrant des segments variables de 1948 à 2006, même si le démarrage est 11/12 en 1970 ou bien après. Certaines années sont douteuses avec une suite de plusieurs valeurs nulles notamment hors de la période janvier-mars. Cela concerne les années suivantes : 1984 pour Bombardopolis, 1988 à Cap Haitien, 1977 et 1979 à Jacmel, 1958, 1977, 1988 à Port de Paix. A priori, ces années douteuses ne sont pas prises en compte dans les calculs suivants (avant validation entre les 3 bases de données).

Par ailleurs, "La vallée Jacmel" n'ont pas de coordonnées géographiques dans la base de données CNIGS_DB. La coordonnée de "La vallée Jacmel" est 18.27°N et 72.67°W dans la base de données <http://www.worldatlas.com/aatlas/indlatlong.htm>. L'altitude est de 526m d'après la fiche Wikipedia de cette commune (http://fr.wikipedia.org/wiki/La_Vallée-de-Jacmel). Après coup, j'ai remarqué que cette localisation géographique (mais avec une altitude différente) correspondait à la station de Ridore dans la base CNIGS_DB. La corrélation sur la période commune entre la Vallée de Jacmel de la base CNSA et Ridore de la base CNIGS_DB est de 1 et ces deux stations peuvent donc être fusionnées avec une altitude de 800m.

Enfin, un doute subsiste également sur la station de "Thiotte Madeleine" dont on peut se demander si elle correspond à la station de "Thiotte" dans la base CNIGS_DB. On ne trouve pas de lieu-dit "Thiotte Madeleine" dans la base de données <http://www.worldatlas.com/aatlas/indlatlong.htm> et jusqu'à preuve du contraire, on considérera que la station dénommée "Thiotte Madeleine" réfère à celle de "Thiotte".

1.3. La base Chemonics

Cette base comporte 154 stations couvrant la période 1923-1998 au maximum avec des précipitations par décades. Elle est difficile à utiliser en tant que telle puisqu'il y a de très nombreuses plages de zéros sans que cela se situe logiquement lors de la saison sèche. Par exemple dans le premier fichier de la base "Aubert", la période mars-décembre 1990 est composée uniquement de zéros. Il est possible que cela soit dans certains cas conforme à la réalité mais on ne peut pas imaginer 10 mois consécutifs totalement secs. De même dans la seconde station "Baie de Henne", les périodes suivantes semblent douteuses; janvier-août 1977, octobre-décembre 1984, janvier-septembre 1986, juin-décembre 1988. Il ne semble pas possible d'utiliser cette base de données en l'état car il est excessivement difficile d'imaginer des critères objectifs afin de distinguer les "vrais" des "faux" 0s. Cette base de données est toutefois utilisée dans ce qui suit pour compléter et vérifier des séries temporelles issues des bases de données précédentes, notamment les stations qui posaient un certain nombre de problèmes dans la base CNIGS_DB. On examinera aussi la possibilité d'étendre les données de la base CNIGS_DB au-delà de 1989.

1.4. La fusion des 3 bases de données

La base générale est constituée des 165 stations de la base "CNIGS_DB". Cette station a été retenue car elle couvre une période longue relativement aux autres. Par rapport à cette base, il y a 107 stations communes dans la base "Chemonics" et 12 stations communes dans la base "CNSA". Les données commencent dans l'absolu en janvier 1863 et se terminent en décembre 2006. Sur cette période, le pourcentage des valeurs manquantes atteint 81.9% dans la base "CNIGS_DB". Un

premier travail a été de comparer les différentes bases de données sur les périodes communes. Il est à noter que l'ensemble de ce travail analytique est entièrement automatisé et il est donc possible d'améliorer la base à partir de critères objectifs.

Une première cartographie des données a permis de détecter deux erreurs avec des stations situées en pleine mer. Les coordonnées ont été recherchées sur <http://www.worldatlas.com/aatlas/indlatlong.htm>; Cornillon est situé à 18.68°N et 71.94°W (au lieu de 18.66°N et 72.93°W dans la base de données) et Petit Goave est situé à 18.43°N et 72.86°W (au lieu de 18.48°N et 72.85°W). Enfin la station de Soissons II n'a pas de coordonnées dans la base CNIGS. Sa localisation est de 18°32'N et 72°15'W selon <http://www.maplandia.com/haiti/ouest/soissons/>

La figure 1 présente les mois disponibles dans deux bases à la station de Cayes. On remarque immédiatement qu'il y a un grand nombre de valeurs nulles dans « Chemonics » qui correspondent à des précipitations mesurées dans les deux autres bases. Ce cas se présente aussi à une autre station (Damien) de façon toutefois moins prononcée (Figure 2). Une première correction a donc été de considérer qu'une valeur nulle dans une des bases alors que la valeur est non nulle dans une ou les deux autres bases (dans le cas des 12 stations disponibles dans les 3 bases) était fausse. Cela concerne surtout la base « Chemonics » mais aussi la base « CIGS » de façon ponctuelle. Quand on compare systématiquement les totaux entre les 3 bases avant cette correction (Figure 3), les corrélations sur les périodes communes sont globalement satisfaisantes mais < 0.8 dans un certain nombre de cas. L'application de la simple règle d'élimination des zéros permet d'augmenter sensiblement les corrélations (Figure 4) et d'éliminer la plupart des cas < 0.8 . Cela ne signifie cependant pas que ce problème n'existe plus, notamment dans le cas où la station n'est disponible que dans la base CNIGS_DB. De façon ponctuelle, la comparaison entre les bases a montré également des totaux $\times 10$. Cette erreur est assez fréquente mais assez simple à corriger dans le cas de stations disponibles dans au moins deux bases mais là encore, on ne peut pas exclure que ce type d'erreur soit totalement exclu notamment dans le cas des stations uniquement disponibles dans la base CNIGS.

Au final, il y a 155 stations c'est-à-dire les 165 de départ de la base CNIGS – les 3 sans coordonnées géographiques ni noms, et 7 doublons (St Marc, Gde Rivière du Nord, Port de Paix, St Louis du Nord, Camp Perrin et Saltadère,???)

2. Analyse des précipitations annuelles

La figure 5 présente les 155 stations de la base fusionnée superposée au relief issu de la base GTOPO30 (résolution horizontale de 0.0084°). La répartition géographique est évidemment loin d'être parfaite notamment sur la portion la plus montagneuse. Certains secteurs comme par exemple le centre de la Péninsule méridionale sont malheureusement assez mal documentés. La disponibilité est maximale (~ 80 stations disponibles) entre les années 20 et les années 70 (Figure 4). Les précipitations annuelles moyennes et le cycle annuel moyen sont calculées selon deux approches : (1) utilisation de toutes les données sans restriction de période disponible. Cette approche maximise la couverture spatiale mais en même temps est biaisée à partir du moment où les précipitations moyennes ne sont pas estimées sur les mêmes périodes et un enregistrement durant une période humide peut par exemple engendrer un biais humide par rapport à un enregistrement dans une phase sèche. Ce biais potentiel n'existe pas seulement si la pluviométrie est non-stationnaire à l'échelle interannuelle mais apparaît simplement dès que la période disponible pour chaque station est brève et donc susceptible d'être contaminée par une erreur liée à l'échantillonnage temporel; (2) utilisation de la période 1921-70 en se restreignant aux stations ayant au moins 10 ans de données durant cette période, soit 96 stations (Figure 4). Ce choix n'exclut bien évidemment pas les biais de la première méthode, puisque les 10 années disponibles peuvent documenter des phases plus ou

moins humides.

Les figures 6 et 7 présentent la moyenne annuelle sur la base des 155 stations et des 96 stations. Les maximas se situent sur l'ouest de la péninsule sud, la partie nord centrale au nord de la chaîne du nord (excepté la plaine du Nord) ainsi que des stations isolées, plutôt en hauteur (Par exemple Perrodin). Des minimas très bas < 500 mm apparaissent sur la Péninsule septentrionale ainsi que sur la vallée de la rivière Artibonite sous le vent des chaînes du Nord. Certaines stations dans la dépression E-W qui s'allonge depuis Port au Prince sont également anormalement sèches (< 1000 mm). Il existe au moins deux climatologies anciennes, qui reposent sur des données partiellement identiques (Figure 8). Ces deux climatologies sont raisonnablement en accord et on peut donc estimer que les précipitations moyennes annuelles sont vraisemblablement comprises entre 500 mm pour les secteurs les plus secs à 2500-3000 mm pour les secteurs les plus arrosés (Péninsule méridionale, chaîne de montagnes notamment au vent des flux d'E-NE au nord).

La question de la relation entre pluie annuelle et altitude n'est pas simple car (1) on sait que dans un air humide proche de la saturation, les précipitations maximales sont plutôt dans les basses pentes et (2) l'orientation au flux humide peut être très (voire plus) importante que l'altitude brute ou relative (c'est-à-dire par rapport au talweg local). Le nuage de point entre l'altitude et les précipitations moyennes annuelles (Figure 9) est effectivement compliqué avec des précipitations moyennes annuelles très variables dans les basses couches (très arrosé au SW et très sec au NW et dans la vallée de l'Artibonite) et un maximum apparemment entre 800 et 1200 m sans que l'ajustement polynomial ne soit très satisfaisant (Figure 9).

En conséquence, il semble difficile, voire impossible, de simuler correctement un champ de pluie, même climatologique, en se basant uniquement sur les altitudes. D'autre part, la complexité de la pluviométrie en fonction de la position interdit des règles simples et linéaires. La simulation des pluies à haute résolution est possible soit avec du krigeage (mais le variogramme doit être fortement modulé en fonction de la position avec des gradients parfois très forts) soit avec des méthodes non-linéaires. J'ai testé une méthode de simulation des pluies annuelles/mensuelles avec un réseau de neurones avec un seul niveau caché afin de ne pas trop compliquer le réseau et les tests. Le modèle a été testé par une méthode récursive avec un apprentissage initial avec 80% des stations disponibles et une validation sur 20% des stations. Ceci est répété 100 fois en choisissant à chaque fois l'ensemble d'apprentissage et de validation de façon aléatoire. Dans une première étape, aucun test n'est réalisé afin de profiter de la plus grande quantité de données. Pour chaque modèle, le prédictand est la moyenne annuelle (et mensuelle ensuite) des pluies et les prédicteurs peuvent être : la longitude (X), la latitude (Y), l'altitude (Z), la direction (D) et l'intensité (S) de la pente, ces deux derniers paramètres étant calculés à partir du fichier GTOPO30² et par recherche des valeurs au point de grille le plus proche de la station pluviométrique. 5 ensembles de prédicteurs ont été utilisés : Z, X+Y, X+Y+Z, X+Y+D+S et les 5 ensemble. Le modèle est validé en comparant chaque station avec l'estimation offerte par la simulation avec les paramètres du fichier GTOPO30 au point de grille le plus proche. Avec une résolution de 1 km, cela signifie que le point de grille est donc à moins de 1 km de la station pluviométrique. Enfin, les réseaux de neurones ont été paramétrés avec 15 neurones dans la couche cachée, mais des tests ont montré que les résultats étaient très robustes avec de 10 à 30 neurones. Je n'ai pas testé pour l'instant une décomposition entre plusieurs couches, mais il n'est pas évident que la taille de l'échantillon soit adaptée à un modèle très complexifié.

La figure 10 présente les scores de corrélation (entre les moyennes annuelles observées aux 155 stations et les moyennes simulées aux points de grille les plus proches). On voit nettement qu'un modèle n'utilisant que Z est insatisfaisant alors que les 4 autres modèles aboutissent à une qualité assez proche ($r \sim 0.75-0.8$ si on considère la corrélation entre la moyenne des 100 simulations et les

2 On calcule d'abord le gradient de la pente dans le sens zonal et méridien à partir des altitudes, puis la direction et l'intensité de la pente.

observations). La figure 11 présente les nuages de points entre observation et moyenne des 100 simulations. La RMSE³ est assez élevée (de l'ordre de 350-400 mm) mais on doit pouvoir corriger en partie ce biais en calibrant la moyenne des 100 simulations par rapport aux observations, c'est-à-dire en multipliant les simulations par le ratio entre l'écart-type observé et l'écart-type simulé. Cela revient à « abaisser » les précipitations simulées les plus faibles et « augmenter » les précipitations simulées les plus élevées de telle sorte qu'on se rapproche de la diagonale en moyenne. Pour l'instant, cette correction n'a pas été appliquée car elle n'apporte rien par rapport à la moyenne spatiale. Ne considérer que 96 stations ayant au moins 10 ans de données entre 1921 et 1970 (Figures 12 et 13) apportent des résultats assez proches. La simulation est la plus satisfaisante avec X, Y, D et S confirmant donc que l'altitude brute n'est pas un facteur déterminant important des précipitations à Haïti. La figure 14 présente la simulation des pluies annuelles avec le modèle $Pan = f(X, Y, S, D)$ à la résolution native du fichier GTOPO30, c'est-à-dire environ 1 km. Les pluies les plus faibles (< 1000 mm et parfois < 600 mm) sont simulées sur la péninsule septentrionale, la vallée de la rivière Artibonite, la dépression de Port au Prince et l'île de la Gonave alors que les maximas (> 1800 mm) s'orientent selon les pentes orientées à l'E-NE au nord et au centre alors que l'extrémité de la péninsule méridionale apparaît comme assez uniformément arrosée.

[Des tests ont été réalisés sur des résolutions dégradées (0.05°) et il est évidemment très simple de réaliser ces cartes de façon automatique. En théorie, il faudrait aussi tester l'impact du nombre de simulations (de 1000 à 10000 par exemple) et d'examiner l'impact des corrections basées sur la variance. On peut aussi dériver de ces cartes des quantités d'eau tombées sur des bassins versants qui pourraient être comparées à des écoulements mesurés (mais il me faut alors des limites de bassin-versant à la résolution native de 0.0085° du fichier GTOPO30). Au niveau des prédictors, je ne pense pas pouvoir faire beaucoup mieux, mais on peut toujours imaginer d'autres quantités basées sur la topographie avec des tests. On peut discuter cette approche car les tests sont très rapides à faire sur matlab tant que le nombre de simulations n'est pas trop prohibitif. Je pense également qu'il vaut sans doute mieux utiliser le maximum de stations même si certaines stations sont disponibles sur des périodes très brèves car les réseaux de neurones sont en général sensibles à la taille de l'échantillon. Il faut peut-être considérer une correction en fonction de l'anomalie moyenne annuelle à l'échelle de l'île ? Enfin, il est peut-être souhaitable de tester le rapport apprentissage/validation (ici 80 %/20%) sur la qualité de l'estimation mais je ne pense pas que cela modifie beaucoup les résultats].

3. Analyse des précipitations mensuelles

Les pluies moyennes mensuelles ont été analysées de la même façon que les pluies annuelles. Les figures 15 et 16 présentent la moyenne climatologique d'après l'ensemble des 155 stations et le sous-ensemble des 96 stations ayant au moins 10 ans disponibles entre 1921 et 1970. Les précipitations mensuelles respectent dans l'ensemble un cycle bi-annuel classique dans cette partie des Caraïbes avec des pluies abondantes surtout d'avril à novembre sans qu'elles soient totalement exclues durant l'hiver boréal avec une baisse relative intermédiaire en juillet. Cette baisse est généralement appelée « Mid-summer Drought » (MSD) dans la littérature et est habituellement associée avec le décalage/gonflement de l'anticyclone tropical de l'Atlantique Nord au cœur de l'été boréal. On peut toutefois noter sur les figures 15-16 des précipitations hivernales un peu plus abondantes sur deux secteurs : la partie centrale du nord, surtout la partie au vent du flux d'ENE à N. C'est peut-être lié à l'activité des perturbations tempérées encore qu'il faudrait dans ce cas descendre à des échelles temporelles plus fines afin de cerner au mieux les mécanismes impliqués. La partie extrême-occidentale de la péninsule méridionale connaît aussi des pluies d'hiver.

Une façon de synthétiser le cycle annuel des précipitations est de réaliser une classification des

3 Root Mean Square Error

cycles annuels moyens (Figures 17-18). Chaque total mensuel a d'abord été standardisé par rapport à la moyenne générale (sur tous les mois et toutes les stations) afin de distinguer non seulement les phases sèches/arrosées mais aussi les stations sèches/arrosées. J'ai ensuite synthétisé la moyenne mensuelle des 155 et des 96 stations grâce à une ACP résumant 90 % de la variance totale. Ensuite, une classification par nuées dynamiques a été réalisée en faisant varier le nombre de classes de 2 à 10. La classification « optimale » a été déterminée grâce à l'indice de « classifiabilité » défini dans Michelangeli et al. (1995). Dans les deux bases de données, cet indice présente un pic peu marqué ($\sim 0.98^4$) pour 4 classes. On remarque que les deux classifications coïncident bien, c'est-à-dire qu'en général, une station classée dans une groupe donné avec 155 stations reste dans le même groupe quand la classification est réalisée sur les 96 stations.

3 cycles annuels approximativement identiques (c'est-à-dire pic principal en mai, pic secondaire plus étalé en septembre-octobre, saison sèche maximale en décembre-mars et MSD en juillet) ressortent avec une simple distinction due à la quantité moyenne des précipitations. Les stations les plus sèches (en rouge sur les figures 17-18) se localisent sur le NW, la dépression au NE de la chaîne du nord, la dépression de la rivière Artibonite, certaines stations de la dépression W-E de Port au Prince, et quelques stations de la côte sud orientée vers le SW. Les stations les plus arrosées (en vert sur les figures 17-18) se placent généralement en altitude. Un cycle un peu différent caractérise le dernier groupe (en noir sur les figures 17-18) avec des pluies hivernales plus abondantes qu'ailleurs et un second pic plus important que le premier et décalé en novembre. Ce style concerne les stations dans la partie centrale nord, notamment les versants orientés au N et au NE, ainsi que les stations au nord de la côte septentrionale de la péninsule méridionale. Il y a sans doute un mélange de différents facteurs qui individualisent un peu ces deux secteurs particuliers : le rôle éventuel des perturbations tempérées, mais sans doute aussi celui de perturbations tropicales en fin de saison dont le rôle doit être particulièrement exacerbé sur ces secteurs.

La figure 19 présente la simulation des pluies mensuelles selon exactement les mêmes principes que pour les précipitations annuelles, à savoir un réseau de neurones paramétré grâce aux coordonnées géographiques et aux directions et intensité des pentes avec un apprentissage sur 80 % des stations et une validation sur les 20 % restants. L'estimation est encore une fois donnée par 100 simulations avec un découpage aléatoire du sous-ensemble d'apprentissage et de validation. Le test indépendant est représenté par la comparaison entre le total mensuel observé et l'estimation fournie au point de grille le plus proche à partir des paramètres GTOPO30. La corrélation entre observation et simulation (= moyenne des 100 simulations) est comprise entre 0.75 et 0.88 selon les mois avec une baisse de la qualité de l'estimation en été, ce qui est assez attendu puisque les variations sont faibles et très bien contraintes géographiquement par la longitude et la latitude en hiver.

[Comme pour les précipitations annuelles, on peut sans doute affiner cette approche si il y a un intérêt et produire des séries par BV].

4. La variabilité des pluies

Une première approche est de considérer l'échelle annuelle qui est la plus intégrative. Par ailleurs, la saison des pluies étant centrée sur l'été boréal, il y a une certaine logique à considérer l'année civile comme première unité temporelle de base. Un premier problème est le calcul de la matrice de covariance avec des séries incomplètes. Le choix a de d'abord sélectionner les 70 stations ayant plus de 30 ans de données. Ce choix a un biais évident : il crée une hétérogénéité assez importante avec une documentation moins bonne pour la péninsule méridionale, ce qui va d'une certaine façon contraindre les axes principaux là où la documentation est un peu meilleure. Ce biais s'ajoute à celui

4 L'indice de classifiabilité vaut 1 si la classification est totalement indépendante du choix initial des centroïdes. Autrement dit, un indice de 1 indique une classification parfaitement robuste alors qu'un indice de 0 indique que la classification dépend uniquement du choix initial aléatoire des centroïdes.

de la géographie d'Haïti puisque la forme générale complexe avec deux péninsules induit d'abord une déformation qui contraint d'une certaine manière les axes principaux.

Une fois le sous-ensemble défini, la corrélation entre chaque couple de stations a été calculé sur la période commune. Une ACP a alors permis d'extraire les axes principaux de cette matrice de corrélation. Les composantes principales ont ensuite été calculées en projetant les données disponibles sur ces axes principaux qui par construction ne sont pas strictement orthogonaux (puisque basés sur des périodes différentes). Les composantes principales ont été calibrées en fonction du nombre de station disponibles. Le premier axe principal résume 21% de la variance totale (Figure 20) avec une évolution en phase de la majorité des stations (mais des corrélations < 0.8 en général donc avec un certain niveau de bruit local) sauf quelques stations isolées. Le second axe, qui explique à peu près 11 % de la variance totale (Figure 20) montre une structure tripolaire avec une opposition entre le centre et les deux extrémités méridiennes du pays. On ne sait pas si cette structure est dictée par des causes physiques ou si c'est le fruit méthodologique de l'ACP, car on sait que les axes principaux, en-dehors du premier, se définissent en partie à partir des résidus des axes précédents (Richman, 1986). Il est difficile de conclure sur cet aspect avec cet ensemble de données. Les axes suivants n'offrent pas de structure simple même si l'axe #3 est plutôt une opposition N-S et l'axe #4 est majoritairement une opposition entre le centre nord et le centre sud. Les variances associées sont dégénérantes et il faut se garder d'une interprétation physique de ces axes. Quand on utilise une classification en nuées dynamiques afin de détecter d'éventuelles sous-régions, le résultat n'est pas satisfaisant car l'indice de classifiabilité défini plus haut varie de façon monotone sans maxima bien définis entre 2 et 10 classes, ce qui suggère qu'un découpage d'Haïti en « sous-régions » cohérentes du point de vue de la variabilité interannuelle est sans doute un exercice un peu vain.

[l'ACP peut servir à reconstruire les données stationnelles, par exemple en utilisant une régression multivariée à partir des composantes principales ou bien en utilisant une ACP « probabiliste » mais je ne sais pas si cette approche a un quelconque intérêt. Pour la régionalisation, on peut utiliser une approche « floue » sans doute plus pertinente mais là encore, je ne sais pas si il y a un intérêt].

Une seconde approche a été de considérer une ACP « étendue » selon le modèle de Camberlin et al. (2003). Ici on considère la covariance non seulement entre les stations d'un réseau, mais aussi entre les mois de l'année. Cela sert à analyser notamment comment la variabilité d'un mois (ou une saison) se combine avec celle d'un (ou d'une) autre. L'approche a été de considérer ici les mois individuels. Dans ce cas, un axe vectoriel n'est pas une seule carte mais une série de 12 cartes. Le premier axe (Figure 21) est dominé par la variabilité interannuelle de février à juin (surtout à avril) puis de novembre. Par contre le mois de décembre semble être en opposition avec ces mois, (surtout avril). Le bimestre août-septembre est relativement indépendant des autres sur cet axe. Le second axe vectoriel (Figure 22) oppose plutôt février et mars à mai et novembre. La variance expliquée au cœur de l'été est faible et variable dans l'espace. Les axes suivants sont très variables. Le mois de juillet, représentant le cœur de la MSD ne ressort jamais franchement sur les 2 premiers axes alors que le début de la première saison des pluies ressort comme la période la plus cohérente (notamment de février à avril).

Une troisième approche a été de considérer de façon indépendante la variabilité interannuelle des totaux trimestriels glissants. Le premier axe des 12 trimestres est représenté sur la figure 23. On remarque sans surprise que le trimestre le plus cohérent, c'est-à-dire celui où le premier axe explique le plus de variance, se localise en février-avril (variance expliquée = 37% de la variance totale), alors que les trimestres les moins cohérents (variance plus faible et corrélations variables dans l'espace) sont JJA, JAS et ASO, c'est-à-dire ceux entre la MSD et le début de la seconde saison des pluies. Les second et troisième axes principaux de chaque trimestre sont présentés sur les figures 24 et 25. Le second axe représentant environ 10 % de la variance est caractérisé par une

structure dipolaire N-S ou bien un tripôle (Fig. 24). Le troisième axe (Fig. 25) n'a pas de structures bien définies et/ou persistantes entre les trimestres glissants.

Les liens entre la variabilité interannuelle des trimestres et les TSO ont été analysés en premier avec une CCA entre les trimestres et les TSO issues du fichier ERSST v3b (Smith et al. 2008). Cette approche nécessite de combler préalablement les données manquantes du champ de pluie. Cela a été réalisé avec une régression multivariée avec les 5 premières composantes principales de chaque trimestre. Cette approche tend à réduire le « bruit » spatial et offre une reconstruction satisfaisante (la corrélation entre reconstruction et observation est de l'ordre de 0.8 en moyenne). Ensuite une CCA a été calculée entre les pluies et les TSO synchrones sur les 12 trimestres glissants en pré-filtrant avec une ACP en ne retenant que les axes expliquant plus de 50 % de la variance totale de chaque champ (Barnett et Preisendorfer, 1987). Seuls sont montrés ici les deux premiers axes de la CCA pour FMA (début de la première saison des pluies et saison la plus cohérente), AMJ (centre de la première saison des pluies), JJA (centré sur la MSD) et SON (centré sur la seconde saison des pluies).

En FMA, le premier axe canonique (Fig. 26) suggère que des précipitations globalement excédentaires sont associées plutôt à des phases chaudes de l'ENSO (et inversement). Les corrélations sont cependant modérées sur le Pacifique est et les plus élevées sont décalées au sud de l'équateur. Les scores temporels montrent par ailleurs une phase positive (donc moins de pluie et plutôt une tendance froide dans le Pacifique Est) de 1940 au milieu des années 70. La variabilité temporelle est notablement plus importante avant 1940 et après 1975. Le second axe de la CCA (Fig. 27) associe une signature plus typique des ENSO (phase chaude associée à un réchauffement dans l'Atlantique tropical Nord) à un dipôle pluviométrique à Haiti avec plus de pluies sur le Nord de l'île en phase chaude. Cela peut être mis en relation avec des perturbations tempérées plus actives sur le Golfe du Mexique en phase chaude du fait du glissement méridien de l'axe des perturbations et d'affaiblissement concomitant de l'anticyclone des Açores en liaison avec une phase négative de l'oscillation nord-atlantique.

En AMJ, le premier axe canonique offre une rémanence de la relation ENSO-Atlantique Tropical Nord (i.e. anomalies chaudes sur l'Atlantique tropical Nord associé à un ENSO chaud comme sur le CCA#2 en FMA) qui couple alors événement chaud ENSO avec des anomalies pluviométriques positives sur la plupart d'Haiti (Fig. 28). Il semble d'après la chronique des scores temporels, qu'il s'agissent plus certainement d'ENSO « finissants » que d'ENSO « démarrants » (cf. score négatif en 1983-84, 1987-88, 1993, 1998 etc.). Cela indique donc que le printemps boréal suivant la fin d'un ENSO serait plutôt associé à des anomalies pluviométriques positives. Le coefficient de corrélation canonique n'est cependant pas très élevé et l'estimation du potentiel en terme de prévision saisonnière n'est sans doute pas très élevée non plus. Le second axe canonique (Fig. 29) montre une forte contribution de la variabilité multidécennale et associe des anomalies chaudes sur l'Atlantique tropical (surtout sa partie sud), notamment depuis les années 1960s, a plutôt moins de pluies sur Haiti, mais il y a de fortes variations spatiales même sur des distances courtes [*je me pose des questions sur ce type de structures et me demande dans quelle mesure cela ne peut pas être le fruit des estimations des pluies*].

En JJA, le premier axe associe anomalie chaude sur l'Atlantique tropical Nord (+ une phase ENSO froide) et plutôt des anomalies pluviométriques positives sur Haiti. Là encore, quelques stations ont un comportement semble-t-il différent des autres. Dans quelle mesure cela peut-il être associé aux erreurs (?), à la reconstruction de la variabilité interannuelle pluviométrique (?) ? Il y aurait donc une inversion, déjà partiellement documentée à l'échelle des Caraïbes (refs), du signal de l'ENSO (chaud = plus de pluies au printemps boréal puis l'inverse en été boréal). On remarque en tous les cas, que la synchronicité des deux courbes temporelles est plus importante au début et à la fin de la période disponible. Le second axe est plus difficile à analyser notamment pour les pluies avec une

très forte variabilité spatiale (Fig. 31).

En SON, on retrouve quelque chose ressemblante avec le signal estival de l'ENSO mais avec pas mal de variabilité multidéennale.

Références

Alpert L. (1941) The areal distribution of mean annual rainfall over the island of Hispaniola, *Monthly Weather Review*, **69**, 201-204.

Fassig O.L. (1929) A tentative chart of annual rainfall over the island of Haiti-San Domingo, *Monthly Weather Review*, **57**, 296.

Huffman et al. TRMM

Michelangeli et al., (1995)