

Modélisation De La Précipitation Dans La Partie Sud-Ouest De Madagascar En Fonction De L'Humidité, De L'OLR, De La Chaleur Latente Et De La SST Par Forêt Aléatoire

[Precipitation Modeling In The Southwest Of Madagascar Based On Humidity, OLR, Latent Heat, And SST Using Random Forest]

Aimé Rivesty ANDRIANARIMANANA¹, Honoré RAZAFIMANDIMBY², Miller Audiat HERINDRAINY³, Jean Eugène RANDRIANANTENAINA⁴, Adolphe Andriamanga RATIARISON⁵

^{1,2,3,4,5} Laboratoire Dynamique de l'Atmosphère et des Océans (DyACO), Physique et Applications, Sciences et Technologies

^{1,2,3,4} Université Tuléar

⁵ Université d'Antananarivo

¹arivesty@gmail.com, ²zoenkary@gmail.com, ³millerhery1@gmail.com, ⁴rjendria02@yahoo.fr,

⁵adolphe.ratiarison@univ-antananarivo.mg



Résumé : Cette étude vise à modéliser la précipitation dans la partie Sud-Ouest de Madagascar en utilisant les variables climatiques explicatives : humidité, Rayonnement ascendant de grande longueur d'onde (OLR), chaleur latente et Température de Surface de la Mer (SST), à l'aide d'un modèle de Forêt Aléatoire. L'objectif est de comprendre l'impact de ces variables sur la variabilité des prélèvements et de prédire ces dernières en fonction des conditions climatiques locales. L'Indice Standardisé de Précipitation (SPI) a été utilisé pour caractériser les périodes climatiques, distinguant les années normales, les années sèches et les années humides sur la période 1979-2018. Les performances du modèle ont été réalisées avec un $R^2 = 0.865$, indiquant une bonne capacité explicative, et un SMAPE = 32.58%, qui est jugé acceptable pour des prévisions climatiques. Les résultats montrent que les variables comme l'humidité, la chaleur latente et la SST ont une influence notable sur les précipitations dans cette région, mettant en évidence l'importance de ces facteurs pour comprendre la dynamique climatique locale.

Mots clés : Précipitations, modélisation, Forêt Aléatoire, humidité, OLR, chaleur latente, SST.

Abstract : This study aims to model precipitation in the southwest of Madagascar using explanatory climatic variables : humidity, Outgoing Longwave Radiation (OLR), latent heat, and Sea Surface Temperature (SST), through a Random Forest model. The goal is to understand the impact of these variables on precipitation variability and to predict them based on local climatic conditions. The Standardized Precipitation Index (SPI) was used to characterize climatic periods, distinguishing between normal years, dry years, and wet years from 1979 to 2018. The model's performance was evaluated with an $R^2 = 0.865$, indicating a good explanatory capacity, and a SMAPE = 32.58%, which is deemed acceptable for climate forecasting. The results show that variables such as humidity, latent heat, and SST have a significant influence on precipitation in this region, highlighting the importance of these factors for understanding the local climate dynamics.

Keywords : Precipitation, modeling, Random Forest, humidity, OLR, latent heat, SST.

Introduction

La partie Sud-Ouest de Madagascar est régulièrement confrontée à des épisodes de sécheresse qui affectent non seulement l'agriculture, mais aussi la disponibilité des ressources en eau pour la population locale. Dans cette région, les précipitations sont principalement associées aux cyclones, ce qui complique la prévision des événements pluvieux, d'autant plus qu'elles sont peu fréquentes et souvent imprévisibles. Cette variabilité des précipitations rend la gestion des ressources hydriques encore plus complexe, nécessitant des outils de prévision climatiques plus précis et adaptés à la spécificité de la région.

L'objectif de cette étude est de déterminer dans quelle mesure certaines variables climatiques, telles que la température de surface de la mer (SST), l'humidité relative, le rayonnement ascendant de grande longueur d'onde (OLR) et la chaleur latente, peuvent influencer les précipitations dans cette zone. Ces variables sont choisies en raison de leur rôle central dans les processus atmosphériques et climatiques, influençant directement la dynamique des précipitations. À travers cette étude, nous cherchons à évaluer l'efficacité du modèle statistique et d'apprentissage automatique pour prédire les précipitations en fonction de ces variables climatiques. L'analyse des résultats permet de mieux comprendre les mécanismes sous-jacents et d'améliorer les prévisions climatiques pour cette région, afin d'aider à une gestion plus optimale des ressources en eau et de l'agriculture.

I. Méthodologie

I.1 Zone d'étude

La zone d'étude se situe dans le Sud-Ouest de Madagascar, entre 22°S et 26°S de latitude et 43°E et 46°E de longitude (figure 1). Cette région, bordée par le canal de Mozambique, est caractérisée par un climat semi-aride à aride avec une forte variabilité de précipitation.

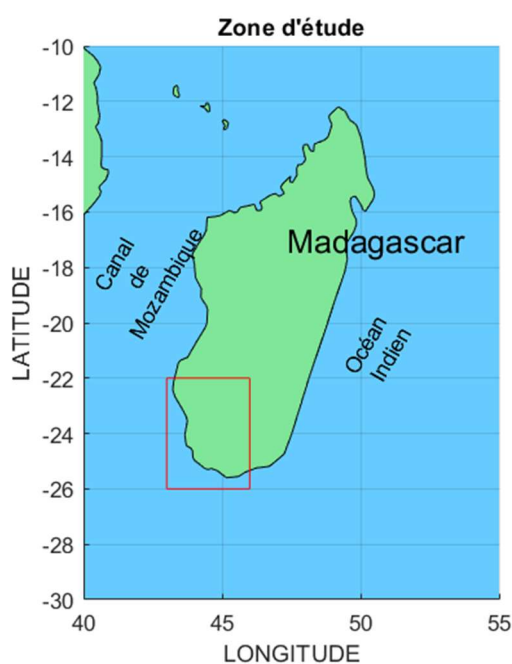


Figure 1 : Représentation de la zone d'étude

I.2 Donnée utilisée :

Les données de réanalyse que nous exploitons comprennent le rayonnement sortant de grande longueur d'onde (OLR), la température de la surface de la mer (SST) et l'humidité. Elles proviennent du Centre européen pour les prévisions météorologiques à moyen terme (ECMWF) et sont structurées en trois dimensions, avec une résolution spatiale définie par la longitude et la latitude, ainsi qu'une dimension temporelle en jours sur la période : 1979-2018.

I.3 Indice Standardisé de la précipitation (SPI)

L'indice standardisé de la précipitation (SPI, pour *Standardized Precipitation Index*) est un indicateur utilisé pour évaluer l'anomalie des précipitations sur une période donnée, en tenant compte des variations naturelles du climat [1] [2]. Il permet de classer les conditions de sécheresse ou d'excès de précipitations. L'indice SPI donne des valeurs qui peuvent être interprétées de la manière suivante :

Tableau 1: Valeur de l'indice de SPI

$2 < \text{SPI}$	Extrêmement humide
$1,5 < \text{SPI} < 1,99$	Très humide
$1 < \text{SPI} < 1,49$	Modérément humide
$-0,99 < \text{SPI} < 0,99$	Proche de la normale
$-1,49 < \text{SPI} < -1$	Modérément sec
$-1,99 < \text{SPI} < -1,49$	Très sec
$-2 > \text{SPI}$	Extrêmement sec

I.4 Corrélation

La corrélation entre deux variables mesure le degré auquel elles varient ensemble. Lorsqu'il y a une corrélation positive, cela signifie que lorsque la valeur d'une variable augmente, la valeur de l'autre variable augmente également. En revanche, une corrélation négative indique que lorsque la valeur d'une variable augmente, la valeur de l'autre variable diminue. Le coefficient de corrélation (noté r) est souvent utilisé pour quantifier cette relation. Il prend des valeurs entre -1 et 1. Un coefficient de corrélation de 1 indique une corrélation positive parfaite, -1 indique une corrélation négative parfaite, et 0 indique qu'il n'y a aucune corrélation linéaire entre les deux variables [3].

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

où :

n est le nombre de paires de données, x et y sont les valeurs des deux variables.

I.5 Régression par Forêt Aléatoire

La forêt aléatoire est une méthode d'apprentissage automatique qui combine les prédictions de plusieurs arbres de décision pour améliorer la précision et réduire le risque de surapprentissage.

L'algorithme de forêt d'arbres décisionnels est composé d'un ensemble d'arbres de décision, et chaque arbre de l'ensemble est constitué d'un échantillon de données tiré d'un ensemble d'entraînement avec une option de remplacement, appelé bootstrapping. Sur cet échantillon d'entraînement, un tiers est mis de côté en tant que données de test, connu sous le nom d'échantillon hors sac

(oob), sur lequel nous reviendrons plus tard. Un autre cas de mesure aléatoire est ensuite injecté par le biais de l'ensachage de fonctionnalités, ajoutant plus de diversité au jeu de données et réduisant la corrélation entre les arbres de décision. Selon le type de problème, la détermination de la prévision variera. Pour une tâche de régression, la moyenne des arbres de décision individuels sera calculée et, pour une tâche de classification, un vote majoritaire, c'est-à-dire la variable catégorielle la plus fréquente, donnera la classe prédite. Enfin, l'échantillon oob est ensuite utilisé pour la validation croisée, finalisant cette prédiction [4] [5].

Les arbres de décision, bien que simples et interprétables, sont sensibles aux variations des données, ce qui peut provoquer un surapprentissage. La forêt aléatoire réduit la variance et améliore la généralisation en combinant plusieurs arbres. La figure 2 montre l'agrégation des prédictions de ces arbres.

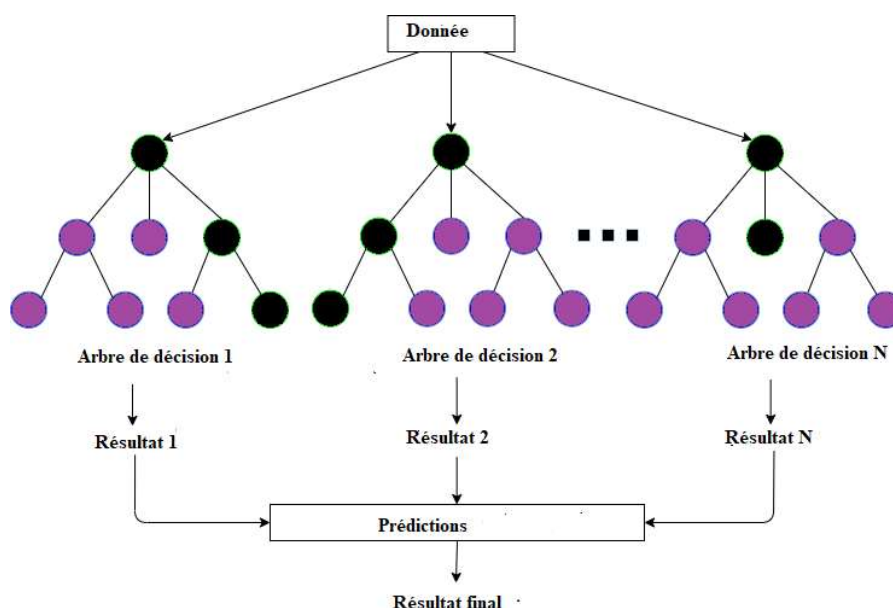


Figure 2 : Agrégation des prédictions dans une forêt aléatoire

II. Résultats

II.1 Indice Standardisé de la précipitation (SPI)

La figure suivante représente la variation des indices standardisés de précipitation entre 1979 et 2018, elle offre un aperçu de l'évolution des régimes pluviométriques au cours de cette période. Elle met en évidence que sur ces 40 années d'observation, 27 années ont affiché des valeurs proches de la normale, caractérisées par des précipitations correspondant en moyenne aux valeurs attendues pour cette période. Ces années pourraient refléter des conditions climatiques relativement stables, sans influences majeures d'événements extrêmes. En revanche, 9 années ont été marquées par des périodes sèches, dont 2 années particulièrement sèches, avec des déficits de précipitation bien en dessous de la normale, ce qui pourrait suggérer une intensification des phénomènes de sécheresse, probablement en lien avec des variations naturelles ou anthropiques, telles que le changement climatique. Les 7 autres années modérément sèches, bien qu'elles présentent des déficits moins extrêmes, montrent néanmoins une tendance à des périodes moins fertiles et peuvent impacter les écosystèmes et l'agriculture. Enfin, 4 années ont été associées à des périodes humides, dont 1 année extrêmement humide, où les précipitations ont largement excédé la normale. Cela pourrait indiquer la présence d'événements climatiques extrêmes comme des cyclones ou des perturbations atmosphériques particulièrement intenses. Les 2 années très humides, bien que marquées par des excédents de précipitations significatifs, sont moins extrêmes que l'année extrêmement humide, et la présence d'une année modérément humide, avec des précipitations supérieures à la moyenne sans

atteindre des niveaux exceptionnels, peut suggérer une augmentation progressive des conditions favorables à la pluie dans certaines régions.

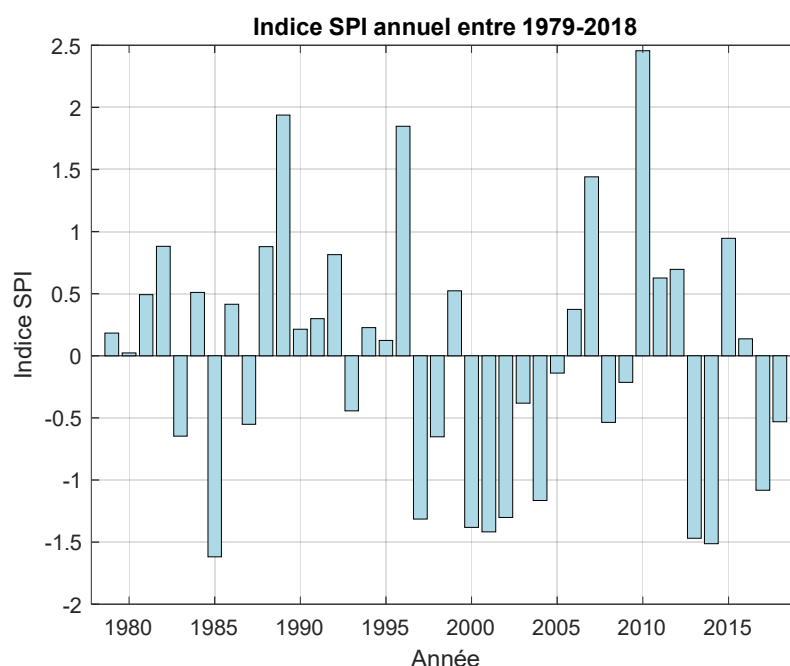


Figure 3 : Variation des indices standardisés de précipitation SPI

II.2 Corrélation entre les variables

Le tableau 2 résume les valeurs de coefficients de corrélation entre la température à la surface de la mer, l'humidité relative à 1000 hPa, l'OLR, la chaleur latente et la précipitation. Les résultats montrent que **la température de surface de la mer (SST) et l'humidité relative sont positivement corrélées avec les précipitations**, ce qui suggère que l'évaporation océanique et l'accumulation d'humidité jouent un rôle clé dans l'apport en précipitations. Cependant, bien que l'humidité relative moyenne soit élevée (environ **80 %**), les précipitations restent faibles, ne dépassant pas **80 mm** en saison pluvieuse. Cela indique que l'humidité seule ne suffit pas à générer des précipitations abondantes.

L'OLR (rayonnement ascendant de grande longueur d'onde) est fortement corrélé négativement avec la précipitation, lorsque des nuages denses se forment et bloquent le rayonnement thermique, les conditions deviennent plus favorables aux précipitations. Toutefois, la **chaleur latente présente une corrélation négative avec les précipitations**, une accumulation d'énergie sous forme de chaleur latente pourrait inhiber la convection et limiter la formation de nuages pluvieux.

Ces résultats révèlent que, malgré une humidité relative élevée, les précipitations restent limitées en raison d'un manque de **forçage dynamique** et de **conditions atmosphériques peu favorables à la convection**, sauf lors du passage des cyclones.

Tableau 2 : Corrélation entre les variables

Coefficient	Précipitation	SST	Humidité relative	OLR	Chaleur latente
Précipitation	1	0,53	0,66	-0,73	-0,59
SST	0,53	1	0,78	-0,67	-0,82
Humidité relative	0,66	0,78	1	-0,83	-0,80
OLR	-0,73	-0,67	-0,83	1	-0,74
Chaleur latente	-0,59	-0,82	-0,80	-0,74	1

II.3 Présentations des figures issues du modèle

La figure 3 illustre les résultats issus du modèle de Forêt Aléatoire, qui parvient à capturer la tendance générale des prélèvements, en reproduisant correctement les périodes de pluie et de sécheresse. Cependant, il a tendance à sous-estimer les précipitations extrêmes, ce qui se traduit par des pics observés plus élevés que ceux prédits. Malgré une bonne représentation des faibles précipitations, les écarts observés suggèrent que les variables utilisées n'expliquent pas pleinement la variabilité des précipitations dans cette région, probablement en raison de corrélations seulement moyennes avec ces dernières.

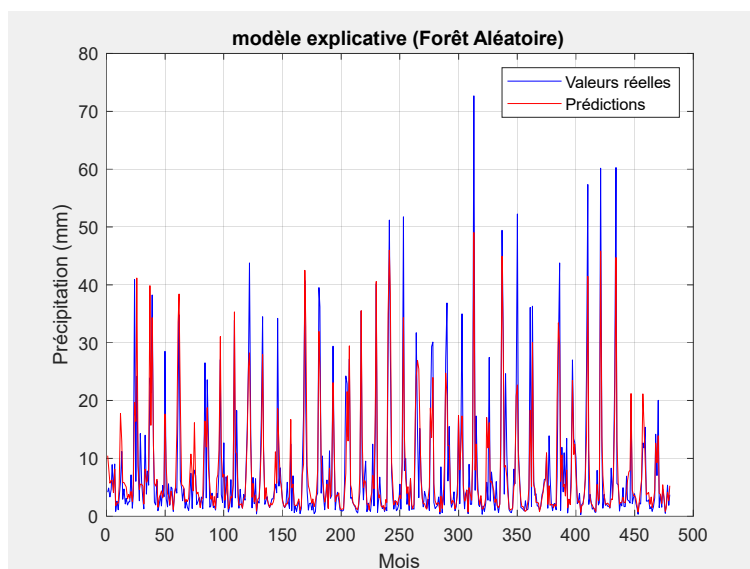


Figure 3 : Résultat du modèle par Forêt aléatoire

II.4 Importance des variables

La figure 4 présente l'importance des variables pour prédire la précipitation. L'OLR et l'humidité sont les deux variables les plus importantes pour prédire les précipitations dans notre modèle. Cela est cohérent avec la physique du climat, car l'OLR et l'humidité atmosphérique sont des facteurs clés dans la formation des précipitations. La SST et la chaleur latente jouent également un rôle, mais leur influence est moindre comparée à l'OLR et à l'humidité.

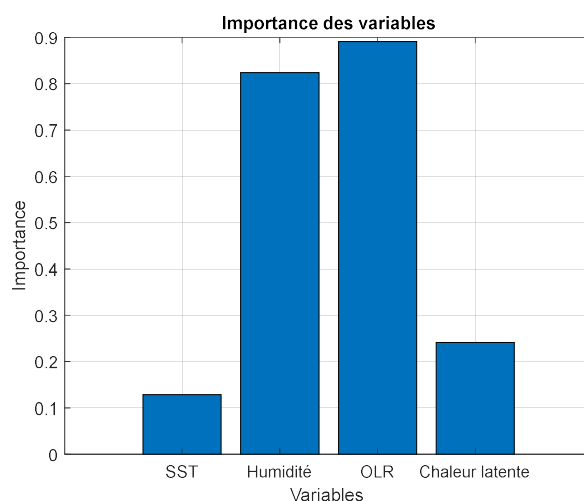


Figure 4 : Importance des variables dans la prédiction des précipitations

II.5 Performance du modèle

Le SMAPE (Symmetric Mean Absolute Percentage Error) est un indicateur qui évalue la performance du modèle en termes d'erreur absolue, mais il est symétrique, ce qui le rend plus robuste que le MAPE traditionnel, surtout en cas de valeurs proches de zéro. La formule du SMAPE est donnée par [6]:

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2}$$

Où :

- y_i : Valeur réelle pour l'observation i .
- $\hat{y}_i$: Valeur prédite pour l'observation i .
- n : Nombre total d'observations.

Tableau 3 : Reference de la qualité du modèle

SMAPE : Taux de croissance de la population (%)	Qualité du modèle
< 10%	Excellente prédiction
10 - 20%	Bonne prédiction
20 - 50%	Prédiction acceptable
> 50%	Mauvaise prédiction

Dans notre cas, nous avons la valeur **SMAPE = 32.52%**, elle suggère une **erreur modérée**, mais acceptable dans un cadre climatique surtout lorsqu'il s'agit de données complexes et de phénomènes non linéaires.

Le **R² (coefficient de détermination)** mesure la proportion de la variance des données recueillies qui est expliquée par le modèle. L'analyse nous a donné la valeur **R² de 0.8650**, cela signifie que le modèle explique environ **86.5% de la variance** des données collectées dans les données testées. Un R² de 0.8650 est généralement considéré comme très bon dans le cadre d'un modèle de régression. La majorité des variations de la précipitation sont bien capturées par le modèle.

III. Discussion

Les résultats obtenus montrent que les précipitations dans la partie Sud-Ouest de Madagascar ne sont pas entièrement expliquées par les variables considérées (SST, humidité relative, OLR, chaleur latente). En effet, la formation des précipitations dépend fortement de processus atmosphériques plus complexes, comme la convection, la circulation des vents et les variations de pression, qui n'ont pas été intégrés dans le modèle. De plus, les précipitations dans cette région surviennent principalement lors des cyclones, des événements extrêmes qui ne sont pas capturés par une approche statistique classique. Des influences climatiques à grande échelle, telles que le phénomène ENSO ou le dipôle de l'océan Indien, pourraient également jouer un rôle dans la variabilité des précipitations.

Par ailleurs, les relations entre les variables climatiques et la précipitation sont probablement non linéaires, ce qui limite l'efficacité des modèles linéaires testés. La faible performance des régressions classiques et des forêts aléatoires suggère que ces méthodes ne suffisent pas à capturer la complexité des interactions climatiques. Pour améliorer la prédiction des précipitations, il serait pertinent d'explorer des modèles plus sophistiqués, comme les réseaux neuronaux ou des approches basées sur la modélisation climatique dynamique, tout en intégrant des variables supplémentaires plus représentatives des processus météorologiques en jeu.

Conclusion

Cette étude a permis d'analyser l'influence de plusieurs variables climatiques sur les précipitations dans la partie Sud-Ouest de Madagascar et d'évaluer l'efficacité des modèles statistiques et d'apprentissage automatique pour leur prévision. Les résultats ont mis en évidence que l'OLR et l'humidité relative sont les principaux déterminants des précipitations dans cette région, soulignant leur rôle clé dans la dynamique atmosphérique locale.

L'analyse du SPI (Standardized Precipitation Index) sur la période étudiée (1979-2018) a révélé une prépondérance des années normales (27 ans), suivie des années sèches (9 ans) et humides (4 ans). Le modèle développé a démontré une performance robuste avec un coefficient de détermination $R^2=0.865$, traduisant une bonne capacité explicative, et un SMAPE de 32.58 %, considéré comme acceptable pour des prévisions climatiques. Ces résultats soulignent l'intérêt d'intégrer des variables atmosphériques spécifiques pour améliorer la prédiction des envois dans cette région vulnérable aux variations climatiques.

Références

- [1] B. M. Thomas, . J. D. Nolan et John Kleist, «THE RELATIONSHIP OF DROUGHT FREQUENCY,» chez *Eighth Conference on Applied Climatology*, Anaheim, California, 1993.
- [2] Copernicus, «Standardized Precipitation Index (SPI),» *Copernicus European Drought Observatory (EDO)*: <https://edo.jrc.ec.europa.eu/>, 2020.
- [3] R. Rakotomalala, «Analyse de corrélation,» Université Lumière Lyon 2, Lyon, 2012.
- [4] L. Balaji , Écrivain, *Mondrian Forests: Efficient Online Random Forests*. [Performance].
- [5] B. Leo, «Random Forests,» *Machine Learning*, vol. 45, p. 5–32, 2001.
- [6] K. Vladik , T. N. Hung et O. Rujira , «How to Estimate Forecasting Quality: A System-Motivated Derivation of Symmetric Mean Absolute Percentage Error (SMAPE) and Other Similar Characteristics,» *scholarworks.utep*, Vols. %1 sur %27-2014, pp. UTEP-CS-14-53 , 2014.