

Modélisation De La Surface Occupée Par Les Tourbillons Océaniques (Cyclones) Au Sud Et Sud-Est De Madagascar

[Modelling The Surface Area Affected By Cyclones In Southern And Southeastern Madagascar]

NIRIKO Henri¹, RATSIMAVO Jacques Chrysologue², LALAOHARISOA Sahoby³,

RANDRIANANTENAINA Jean Eugène⁴, RATIARISON Adolphe⁵

^{1,2,4} Université de Tuléar, Laboratoire de la Dynamique de l'Atmosphère, du Climat et des Océans (DyACO), B.P.
185, Antananarivo, Madagascar,

^{3,5} Université d'Antananarivo, Laboratoire de la Dynamique de l'Atmosphère, du Climat et des Océans (DyACO),
B.P. 906, Antananarivo, Madagascar,

¹nirikohenri@gmail.com , ²rat.chryso@gmail.com , ³sahoby85@gmail.com , ⁴rjendria02@yahoo.fr ,

⁵adolphe.ratiarison@univ-antananarivo.mg



Résumé : Ce travail vise à construire un modèle basé sur la forêt aléatoire afin d'analyser les relations entre plusieurs variables océaniques, concernant la surface occupée par les tourbillons (ou cyclones) au Sud et au Sud-Est de Madagascar. Les variables considérées incluent le vent zonal (u, m/s), le vent méridional (v, m/s), les courants marins (m/s), la hauteur de la surface de la mer (SSH, m), la température de la surface de la mer (SST, K), la salinité de surface de la mer (SSS, PSU), la pression atmosphérique (Pa) et la surface relative occupée par les tourbillons. Un prétraitement des données a été effectué, suivi d'une évaluation de la multicollinéarité à l'aide du Facteur d'Inflation de la Variance (VIF) et de l'analyse des corrélations entre les variables. Ensuite, l'analyse de causalité basée sur le test de Granger a permis de sélectionner les variables les plus significatives à inclure dans le modèle. D'après ces analyses, les variables retenues pour la modélisation sont la salinité de surface de la mer (SSS), la hauteur de la mer (SSH), la température de surface de la mer (SST) et le courant méridional. L'évaluation des performances du modèle montre un MAPE de 1,6008 % et un coefficient de détermination R^2 de 0,8523, attestant de sa précision et de sa fiabilité.

Mots clés : Surface occupée par les tourbillons océaniques, VIF, corrélation, causalité de Granger, forêt aléatoire, MAPE

Abstract : This work aims to build a model based on a random forest to analyze the relationships between several oceanic variables concerning the area occupied by eddies (cyclones) in the South and Southeast of Madagascar. The variables considered include zonal wind (u, m/s), meridional wind (v, m/s), ocean currents (m/s), sea surface height (SSH, m), sea surface temperature (SST, K), sea surface salinity (SSS, PSU), atmospheric pressure (Pa), and the relative area occupied by eddies. Data preprocessing was performed, followed by an evaluation of multicollinearity using the Variance Inflation Factor (VIF) and correlation analysis between variables. Subsequently, causality analysis based on the Granger test allowed the selection of the most significant variables to include in the model. Based on these analyses, the variables retained for modeling are sea surface salinity (SSS), sea surface height (SSH), sea surface temperature (SST), and meridional current. The evaluation of the model's performance shows a MAPE of 1.6008% and a coefficient of determination R^2 of 0.8523, attesting to its accuracy and reliability.

Keywords : Surface occupied by oceanic eddies, VIF, correlation, Granger causality, random forest, MAPE

Introduction

L'étude de la surface relative, définie comme la zone occupée par chaque tourbillon dans l'océan, nécessite l'intégration de plusieurs paramètres océanographiques et climatiques. Afin de modéliser cette surface de manière précise, un ensemble d'analyses et de prétraitements a été mis en place pour optimiser la sélection des variables et garantir la robustesse du modèle.

L'étude se concentre sur la région située au Sud et au Sud-Est de Madagascar. Les variables étudiées dans cette analyse comprennent le vent zonal (u), le vent méridional (v), les courants marins, la hauteur de la surface de la mer (SSH), la température de la surface de la mer (SST), la salinité de surface de la mer (SSS), la pression atmosphérique (Pa) et la surface relative occupée par les tourbillons. Le travail est structuré comme suit : dans un premier temps, nous présentons la zone d'étude. Ensuite, nous détaillons le prétraitement des données et les méthodes utilisées, notamment l'étude de la stationnarité, le test de causalité de Granger, ainsi que l'analyse de la multicolinéarité avec le facteur d'inflation de la variance (VIF) et les corrélations. Enfin, dans les résultats, nous présentons la sélection des variables pertinentes, le modèle obtenu, et l'analyse de la performance du modèle.

1. Zone d'étude et données

➤ Délimitation de la zone d'étude :

L'étude se focalise sur une région du Sud et Sud-Est de l'Océan Indien, englobant Madagascar, en définissant deux zones d'analyse : la Zone A, située entre $40^{\circ}E$ et $75^{\circ}E$ de longitude et $25^{\circ}S$ à $30^{\circ}S$ de latitude, qui couvre des eaux profondes influencées par le Courant des Aiguilles et des structures tourbillonnaires, et la Zone B, s'étendant entre $47^{\circ}E$ et $75^{\circ}E$ de longitude et $18^{\circ}S$ à $25^{\circ}S$ de latitude, qui comprend le plateau continental de Madagascar. La figure ci-dessous illustre les limites des zones A et B superposées à la carte des côtes de Madagascar et des îles environnantes.

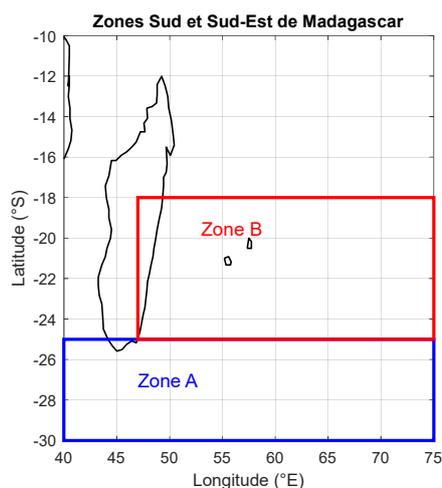


Figure 1 : Zone d'étude

➤ Présentation des données utilisées

Les données exploitées dans cette étude proviennent des bases Copernicus Marine et Copernicus Climate Data Store, qui fournissent des informations détaillées sur les variables océanographiques et climatiques utilisées dans l'analyse. Ces données sont structurées sous forme de matrices tridimensionnelles (longitude, latitude, temps) afin de permettre une analyse spatio-temporelle précise des phénomènes étudiés.

2. Variables étudiées

L'étude repose sur plusieurs variables océanographiques et climatiques, notamment le vent zonal (u, m/s), le vent méridional (v, m/s), les courants marins (m/s), la hauteur de la surface de la mer (SSH, m), la température de la surface de la mer (SST, K), la salinité de surface de la mer (SSS, PSU), la pression atmosphérique (Pa) et la surface relative occupée par les tourbillons.

3. Méthodologie

Pour modéliser la surface occupée par les tourbillons dans la région sud et sud-est de Madagascar, une méthodologie rigoureuse a été mise en œuvre. Celle-ci comprend plusieurs étapes clés, allant du prétraitement des données à la modélisation par forêt aléatoire, en passant par une analyse approfondie des variables.

3.1 Prétraitement des données

➤ *Calcul des moyennes climatologiques journalières*

Pour réduire les fluctuations saisonnières et examiner les tendances climatiques à long terme, les moyennes climatologiques journalières ont été calculées en tenant compte des dimensions spatiales (longitude et latitude) ainsi que des dimensions temporelles (jours et années) sur la période d'étude allant de 1993 à 2020.

La formule employée pour le calcul de la moyenne spatio-temporelle journalière est la suivante :

$$\overline{X}_{spatio-temporelle} = \frac{1}{N_{lat} N_{lon} N_t} \sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} \sum_{t=1}^{N_t} X_{i,j,t}$$

N_{lat} , N_{lon} : Dimensions spatiales.

N_t : Nombre total de périodes temporelles.

$\overline{X}_{spatio-temporelle}$: la moyenne spatio-temporelle journalière

$X_{i,j,t}$: Valeur de la variable à la position spatiale (longitude, latitude) et au temps t.

➤ *Analyse de stationnarité des données temporelles*

Une fois les moyennes spatio-temporelles des variables calculées, nous avons vérifié la stationnarité des séries temporelles en utilisant le test de Dickey-Fuller Augmenté (ADF). Ce test, introduit par Dickey et Fuller en 1979, permet d'identifier la présence d'une racine unitaire, ce qui indique une non-stationnarité.

Hypothèses :

- H_0 : La série possède une racine unitaire (non stationnaire).
- H_1 : La série est stationnaire.

Critère de décision :

Si la p-value est inférieure à 0,05, cela entraîne le rejet de l'hypothèse nulle, indiquant ainsi une stationnarité de la série. Autrement, on considère que la série n'est pas stationnaire.

Pour rendre stationnaires les séries non stationnaires, une différenciation du premier ou deuxième ordre a été mise en œuvre.

3.2 Analyse exploratoire des données

a) Analyse de causalité de Granger

Le test de causalité de Granger a été effectué pour déterminer les relations causales entre les variables. Ce test aide à identifier si une variable peut servir à prédire une autre [1].

Les hypothèses :

- H_0 : La variable X n'est pas une cause de Granger de Y.
- H_1 : La variable X est une cause de Granger de Y.

Seulement les relations considérées comme significatives ($p < 0,05$) sont prises en compte pour l'inclusion des variables dans le modèle. Les données indiquent qu'il existe une relation directionnelle forte entre certaines variables, soulignant leur pertinence pour la modélisation.

b) Analyse de corrélation

Une étude de corrélation a été réalisée afin d'examiner les liens entre diverses variables et de déterminer celles qui sont significatives pour le modèle. Cette approche évalue la puissance et le sens des corrélations linéaires, simplifiant ainsi le choix des éléments significatifs. La formule suivante a été utilisée pour calculer les coefficients de corrélation, désignés par la lettre r :

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

où $Cov(X, Y)$ est la covariance entre les variables X et Y , et σ_X et σ_Y représentent leurs écarts-types. Cette analyse a permis d'éliminer les variables présentant des corrélations élevées pour éviter la multicolinéarité et optimiser les performances du modèle, guidant ainsi le choix des variables à retenir dans la modélisation pour garantir une interprétation fiable des résultats.

c) Analyse de la multicolinéarité (VIF)

L'examen de la multicolinéarité entre les variables explicatives a été effectué en utilisant le facteur d'inflation de la variance (VIF). La formule suivante est utilisée pour calculer le VIF [2] :

$$VIF = \frac{1}{1 - R_i^2}$$

R_i^2 : le coefficient de détermination obtenu en régressant la variable i sur toutes les autres variables explicatives du modèle.

Le VIF évalue dans quelle mesure une variable est linéairement associée aux autres variables explicatives du modèle. On utilise les critères suivants :

- **VIF < 5** : Colinéarité faible, variable maintenue.
- **$5 \leq VIF \leq 10$** : Présence d'une colinéarité modérée, une analyse plus approfondie est requise.
- **VIF > 10** : Présence de colinéarité élevée, variable qui pourrait être retirée ou modifiée.

On a géré les variables ayant une multicolinéarité élevée par :

- Élimination de certaines variables répétitives.
- Changement des variables liées par corrélation

3.3 Sélection des variables :

Sur la base des résultats des analyses précédentes (corrélation, causalité de Granger, VIF), les variables les plus pertinentes ont été retenues, tandis que celles présentant une redondance ou un impact causal faible ont été éliminées.

3.4 Modélisation par forêt aléatoire

La forêt aléatoire (Random Forest) est une méthode d'apprentissage automatique basée sur l'agrégation d'arbres de décision. Elle est largement utilisée pour des tâches de classification et de régression, grâce à sa robustesse, sa simplicité et sa capacité à gérer des données complexes [3].

❖ *Principe Mathématique*

La forêt aléatoire combine les prédictions des B arbres de décision indépendants. Pour un ensemble de données $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ où x_i représente les caractéristiques et y_i la cible, chaque arbre T_b est entraîné sur un sous-ensemble bootstrap D_b de D [4] [5].

- **Classification** : La prédiction finale est obtenue par un vote majoritaire :

$$\hat{y} = \text{mode}\left(\{T_b(x)\}_{b=1}^B\right)$$

- **Régression** : La prédiction finale est la moyenne des prédictions des arbres :

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

❖ **Importance des Variables**

L'importance d'une variable j est calculée en mesurant la diminution moyenne de l'impureté (Gini ou entropie pour la classification, variance pour la régression) induite par cette variable sur tous les arbres :

$$\text{Importance}(j) = \frac{1}{B} \sum_{b=1}^B \sum_{t \in T_b} \Delta I(j, t)$$

où $\Delta I(j, t)$ est la réduction d'impureté due à la variable j au nœud t .

❖ **Avantages**

- **Robustesse** : Moins sensible au surajustement grâce à l'agrégation et à la sélection aléatoire des variables.
- **Interprétabilité** : Mesure de l'importance des variables.
- **Polyvalence** : Applicable à des problèmes de classification et de régression.

4. Résultats

4.1 Analyse de sélection et filtrage des variables pour la modélisation

Après avoir mené plusieurs phases de prétraitement des données, incluant des tests de stationnarité, une analyse de causalité selon Granger et une étude de la multicollinéarité (via le **VIF**, *Variance Inflation Factor*), un ensemble final de variables explicatives a été sélectionné pour modéliser la **surface relative**. Cette sélection rigoureuse garantit que les variables retenues sont à la fois pertinentes et non redondantes, optimisant ainsi la performance du modèle.

4.2. Stabilité des séries temporelles

La stationnarité des séries temporelles a été évaluée à l'aide du **test Augmented Dickey-Fuller (ADF)**. Les résultats initiaux ont montré que toutes les variables étaient **non stationnaires** ($p\text{-value} > 0,05$). Pour remédier à cela, les séries ont été différenciées jusqu'à l'**ordre 2**, ce qui a permis d'obtenir des séries stationnaires, condition essentielle pour une modélisation fiable.

4.3. Causalité de Granger

Le test de causalité de Granger a été appliqué pour identifier les relations de causalité entre les variables. Les résultats ont révélé que parmi les huit variables étudiées, une seule était causée par la surface relative. Plus précisément :

Surface relative → Salinité (P-value : 0,01207) : La surface relative a un effet causal significatif sur la salinité. Cette relation unidirectionnelle souligne l'importance de la surface relative dans l'explication des variations de salinité de surface de la mer.

4.5 Corrélation entre les variables

L'analyse des corrélations entre les variables permet de comprendre les relations linéaires entre elles et d'identifier les variables les plus pertinentes pour la modélisation de la surface relative. La synthèse des résultats est représentée dans le tableau I:

Tableau I : Corrélation entre les variables

	surface	vague	pression	SSH	SST	ventU	ventV	courantU	courantV	salinité
surface	1,000	-0,413	-0,684	0,671	0,805	0,080	-0,372	0,055	-0,483	-0,607
vague	-0,413	1,000	0,804	-0,774	-0,549	-0,341	0,895	-0,627	0,734	-0,015
pression	-0,684	0,804	1,000	-0,909	-0,908	-0,196	0,794	-0,427	0,822	0,476
SSH	0,671	-0,774	-0,909	1,000	0,833	0,315	-0,758	0,450	-0,713	-0,312
SST	0,805	-0,549	-0,908	0,833	1,000	0,091	-0,522	0,141	-0,660	-0,751
ventU	0,080	-0,341	-0,196	0,315	0,091	1,000	-0,172	0,662	0,032	0,167
ventV	-0,372	0,895	0,794	-0,758	-0,522	-0,172	1,000	-0,567	0,855	-0,009
courantU	0,055	-0,627	-0,427	0,450	0,141	0,662	-0,567	1,000	-0,357	0,341
courantV	-0,483	0,734	0,822	-0,713	-0,660	0,032	0,855	-0,357	1,000	0,260
salinité	-0,607	-0,015	0,476	-0,312	-0,751	0,167	-0,009	0,341	0,260	1,000

4.4 Analyse de la multicollinéarité (VIF)

L'analyse de la multicollinéarité via le VIF a permis d'évaluer les dépendances linéaires entre les variables (Tableau II). Voici les résultats détaillés :

Vague (VIF = 3,16) : La variable vague présente une multicollinéarité modérée, indiquant une certaine corrélation avec d'autres variables. Cependant, cette corrélation n'est pas excessive, et la variable peut être conservée pour la modélisation.

Pression (VIF = 6,64) : La variable pression montre une multicollinéarité élevée, suggérant une forte corrélation avec d'autres variables, potentiellement SST (Température de Surface de la Mer) ou SSH (Hauteur de Surface de la Mer). Cette redondance pourrait justifier son exclusion si d'autres variables la représentent suffisamment.

SSH (VIF = 0,65) : La variable SSH présente une multicollinéarité très faible, ce qui signifie qu'elle apporte une information unique au modèle. Elle peut être conservée sans problème.

SST (VIF = 7,88) : La variable SST a le VIF le plus élevé, indiquant une forte redondance avec d'autres variables. Elle pourrait être corrélée à la pression ou à d'autres variables, et son exclusion pourrait être envisagée si d'autres variables capturent suffisamment son information.

VentU (VIF = 1,59) et VentV (VIF = 1,27) : Les composantes du vent (VentU et VentV) présentent des VIF faibles, ce qui indique une faible redondance avec les autres variables. Elles peuvent être conservées sans risque de multicollinéarité.

CourantU (VIF = 5,76) : La variable CourantU montre un VIF légèrement élevé, suggérant une certaine redondance. Une réévaluation de cette variable est recommandée, notamment en la comparant à des variables similaires comme CourantV ou VentV.

CourantV (VIF = 3,19) : La variable CourantV présente un VIF modéré, ce qui indique qu'elle peut être conservée dans le modèle sans problème majeur.

Salinité (VIF = 5,97) : La variable salinité montre une multicolinéarité modérée. Cependant, étant donné son importance identifiée par le test de causalité de Granger, elle pourrait être conservée dans le modèle. Une vérification supplémentaire de sa relation avec d'autres variables (comme SST) est recommandée pour éviter toute redondance excessive.

Tableau II : Tableau des Valeurs de VIF pour les Variables Explicatives

	vague	pression	SSH	SST	ventU	ventV	courantU	courantV	salinité
VIF	3,16	6,64	0,65	7,88	1,59	1,27	5,76	3,19	5,97

4.6 Sélection des variables à retenir et à exclure dans le Modèle

En s'appuyant sur des évaluations de corrélation, de multicolinéarité (VIF) et de causalité de Granger, nous avons effectué un choix minutieux des variables afin d'améliorer l'efficacité et la lisibilité du modèle. Les variables sélectionnées montrent une corrélation notable avec la surface relative, une redondance minimale entre elles et une signification physique dans le cadre de la recherche. Par contre, les variables non retenues ont été éliminées pour leur contribution négligeable, l'absence de relation causale ou leur redondance avec d'autres indicateurs plus significatifs. Le tableau III résume les variables à intégrer ou à exclure, avec les raisons correspondantes.

Tableau III : Les variables à intégrer ou à exclure

Variable	Statut	Raison
Salinité (SSS)	retenue	Corrélation modérée, causalité significative, VIF modéré.
SSH	retenue	Corrélation modérée, VIF très faible, pertinence physique.
SST	retenue	Corrélation élevée, pertinence physique (exclure la pression pour éviter la redondance).
CourantV	retenue	Corrélation modérée, VIF acceptable, pertinence physique.
Pression	exclue	Multicolinéarité élevée avec SST, redondance.
Vague	exclue	Corrélation modérée, absence de causalité, faible contribution.
VentU et VentV	exclue	Corrélation faible, absence de causalité, faible contribution.
CourantU	exclue	Corrélation très faible, absence de causalité, faible contribution.

Les variables SST, Salinité, SSH et CourantV sont les meilleures candidates pour prédire la surface relative. Elles présentent une bonne corrélation avec la surface, un faible niveau de redondance entre elles, et leurs effets causaux sont significatifs. En revanche, les variables Vague, VentU, VentV et CourantU ne contribuent pas de manière significative à la modélisation de la surface relative, soit en raison de leur faible impact causal, soit de leur redondance avec d'autres variables. Par conséquent, ces dernières peuvent

être écartées du modèle sans perte d'information importante. Cette sélection rigoureuse garantit un modèle robuste, interprétable et performant.

Remarque : La vague est principalement influencée par le vent et les conditions météorologiques locales. Bien qu'elle puisse avoir un impact sur la surface relative, cet impact est souvent indirect et moins significatif que celui des courants ou de la température de l'eau.

4.7 Modélisation

❖ Paramètres du modèle

- **Variables d'entrée (prédicteurs) :**

- ✓ **SST (Température de Surface de la Mer) :** Influence la densité de l'eau et les courants marins, avec une corrélation élevée avec la surface relative.
- ✓ **Salinité (SSS) :** Contribue à la densité de l'eau et présente une relation causale significative avec la surface relative.
- ✓ **SSH (Hauteur de Surface de la Mer) :** Directement liée aux variations du niveau de la mer, avec une faible multicolinéarité et une corrélation modérée.
- ✓ **CourantV (Composante V du Courant de Surface) :** Capture les mouvements verticaux de l'eau, apportant une information complémentaire sur les variations de la surface relative.

- **Variable de sortie (cible) :**

- ✓ **Surface relative (surface occupée par les tourbillons) :** Variable à expliquer, représentant la surface occupée par les tourbillons océaniques.

❖ Arbres

L'arbre de décision généré par MATLAB segmente les données en fonction de plusieurs variables explicatives pour prédire la surface relative. Construits à partir de 100 échantillons distincts, ces arbres de la forêt aléatoire capturent différentes variations et combinent leurs résultats pour une estimation robuste. L'analyse met en évidence l'importance du courant méridional (x_3) et de la salinité de surface de la mer (x_5) dans la classification, révélant des relations non linéaires influençant fortement la surface relative.

Voici un exemple d'arbre de décision généré par MATLAB, illustrant la segmentation des données en fonction des variables explicatives et la prédiction de la surface relative (Figure 2).

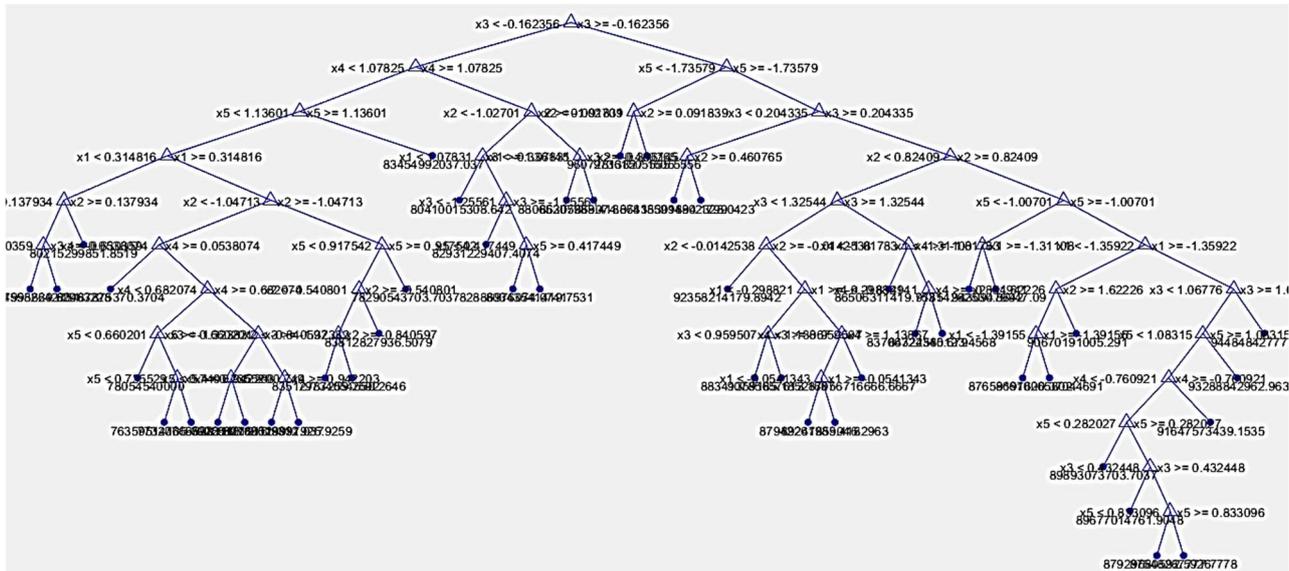


Figure 2: Exemple d'arbre de décision pour la prédiction de la surface relative

❖ Répartition des Données pour l'Apprentissage et le Test

La préparation des données pour un modèle de machine learning, nécessite une étape fondamentale qui consiste à diviser en deux groupes bien distincts :

- l'ensemble d'apprentissage
- l'ensemble de test.

Dans ce travail, 80 % des données sont utilisées pour l'apprentissage, permettant au modèle de s'entraîner et de découvrir les patterns cachés dans la majorité des informations disponibles. Les 20 % restants, quant à eux, servent à tester le modèle, offrant une évaluation impartiale de ses performances sur des données qu'il n'a jamais rencontrées auparavant.

Par cette structure de données, nous garantissons que le modèle conserve sa solidité et son efficacité, même lorsqu'il est confronté à des informations inédites.

5 Performance du modèle : indicateurs de qualité d'ajustement

Dans cette recherche, nous avons examiné l'efficacité du modèle de forêt aléatoire à travers deux indices majeurs :

- le coefficient de détermination (R^2)
- MAPE (*Mean Absolute Percentage Error*).

Ces indicateurs évaluent à la fois la capacité du modèle à expliquer la variation des données et l'exactitude de ses prévisions.

Le coefficient de détermination R^2 est de **0,8523**, ce qui signifie que le modèle explique environ **85,23 %** de la variabilité des données. Cela témoigne d'un bon ajustement et d'une forte capacité à saisir les relations entre les variables.

Le MAPE est une mesure de l'erreur relative moyenne en pourcentage, calculée à l'aide de la formule suivante [6]

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

où :

- n est le nombre total d'observations,
- y_i est la valeur réelle de l'observation i ,
- \hat{y}_i est la valeur prédite pour l'observation i .

Pour notre modèle, le MAPE est de 1,6008%, ce qui montre que, en moyenne, les prédictions du modèle s'écartent de seulement de 1,6008% des valeurs réelles. Cette faible erreur relative confirme la haute précision du modèle.

L'association de ces deux critères souligne une performance globale remarquable : le modèle se révèle être à la fois fiable (grâce à un R^2 important) et précis (en raison d'un MAPE très bas). Ces conclusions indiquent que l'algorithme de forêt aléatoire s'ajuste parfaitement aux données et peut produire des prévisions solides.

5 Présentation de la figure issue du modèle

La figure 3 illustre les résultats obtenus à partir du modèle, mettant en évidence les tendances, les performances et les éventuelles variations des prédictions par rapport aux valeurs réelles. Elles permettent d'évaluer visuellement la qualité de l'ajustement.

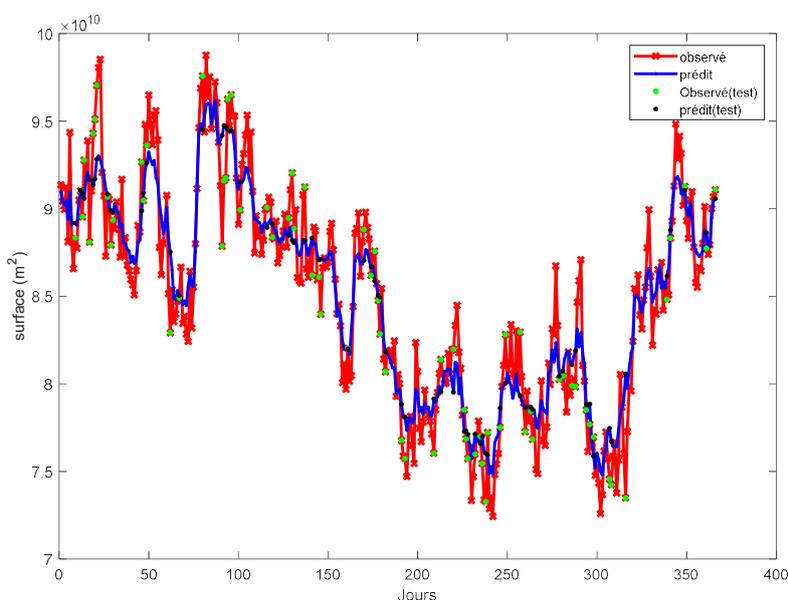


Figure 3 : Représentation du modèle et de la valeur observée

Conclusion

En conclusion, cette étude a permis de développer un modèle de forêt aléatoire intégrant plusieurs variables océaniques et climatiques pour mieux comprendre les dynamiques marines, notamment la surface occupée par les tourbillons au Sud et Sud-Est de Madagascar. L'objectif principal était de créer un modèle robuste capable de traiter un ensemble complexe de variables, incluant la salinité de la surface, la hauteur de la mer, la température de surface, les courants marins, ainsi que des facteurs climatiques tels que le vent et la pression atmosphérique. L'évaluation des performances du modèle a montré des résultats prometteurs, avec un MAPE de 1,6008 % et un coefficient de détermination R^2 de 0,8523, confirmant sa capacité à fournir des prédictions fiables.

Référence :

- [1] C. Granger, «Investigating causal relations by econometric models and cross-spectral methods,» *Econometrica*, p. 424–438, 1969.
- [2] M. Jacob et K. Eda, «Qu'est-ce que la multicolinéarité,» 21 novembre 2024. [En ligne]. Available: <https://www.ibm.com/fr-fr/topics/multicollinearity#f15>.
- [3] L. Breiman, «Random Forests,» *Machine Learning*, p. 5–32, 2001.
- [4] T. Hastie, R. Tibshirani et J. Friedman, «The Elements of Statistical Learning: Data Mining, Inference, and Prediction,» Springer, 2009.
- [5] A. Géron, «Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow,» O'Reilly Media, 2019.
- [6] S. Kim et H. Kim, «A new metric of absolute percentage error for intermittent demand forecasts,» *International Journal of Forecasting*, vol. 32, n° 13, pp. 669-679, 2016.