

# *Modélisation De La Hauteur De Vague Au Sud Et Sud-Est De Madagascar*

Henri NIRIKO<sup>1</sup>, Sahoby LALAOHARISOA<sup>2</sup>, Jacques Chrysologue RATSIMAVO<sup>3</sup>, Adolphe RATIARISON<sup>4</sup>

<sup>1,2,3,4</sup> Laboratoire Dynamique de l'Atmosphère du Climat et des Océans (DyACO), Physique et Applications, Sciences et Technologies

<sup>1,3</sup> Université de Toliara

<sup>2,4</sup> Université d'Antananarivo

<sup>1</sup>nirikohenri@gmail.com, <sup>2</sup>sahoby85@gmail.com, <sup>3</sup>rat.chryso@gmail.com, <sup>4</sup>adolphe.ratiarison@univ-antananarivo.mg



**Résumé :** Ce travail vise à modéliser la moyenne climatologique de la hauteur des vagues au Sud et au Sud Est de Madagascar en utilisant la méthode de régression linéaire multiple au sein d'un modèle unique. Les variables explicatives considérées incluent le vent (composantes méridionale et zonale), les courants marins, la hauteur de la surface de la mer (SSH), la salinité de la surface de la mer, la température de la surface de la mer (SST) et la pression atmosphérique. Une analyse préalable comprenant le test de causalité de Granger, et l'évaluation de la multicollinéarité via le facteur d'inflation de la variance (VIF) a permis d'éliminer la variable SSH, en raison de la forte valeur de VIF et de l'absence de lien de causalité significatif. Les résultats finaux indiquent que les sept variables explicatives restantes permettent de prédire efficacement la hauteur des vagues, avec un coefficient de détermination ( $R^2$ ) de 0.8916, témoignant d'une forte capacité explicative du modèle.

**Mots clés :** hauteur de vague, test de causalité de Granger, régression linéaire multiple, multicollinéarité, VIF.

**Abstract :** The aim of this work is to model climatological mean wave heights in the south and southeast of Madagascar, using the multiple linear regression method within a single model. The explanatory variables considered include wind (meridional and zonal components), oceans currents, sea surface height (SSH), sea surface salinity, sea surface temperature (SST) and atmospheric pressure. A preliminary analysis including the Granger causality test, and the assessment of multicollinearity via the variance inflation factor (VIF) allowed the SSH variable to be eliminated, due to the high value of VIF and the absence of a significant causal link. The final results indicate that the seven remaining explanatory variables effectively predict wave height, with a coefficient of determination ( $R^2$ ) of 0.8916, testifying to the model's strong explanatory capacity.

**Keywords:** wave height, Granger causality test, multiple linear regression, multicollinearity, VIF.

## **I. Introduction**

La modélisation de la hauteur des vagues est un enjeu clé dans le domaine de l'océanographie et des sciences climatiques, particulièrement dans des régions vulnérables comme le Sud et le Sud Est de Madagascar. Les vagues influencent de manière

significative les écosystèmes côtiers, les activités humaines et les infrastructures maritimes. La compréhension des facteurs qui déterminent la hauteur des vagues, basée sur une analyse climatologique, permet d'améliorer la gestion des risques et optimiser les activités maritimes. Dans cette étude, des données climatologiques pendant 28 ans sont utilisées, pour pouvoir proposer un modèle de régression linéaire multiple prédisant la moyenne climatologique de la hauteur des vagues en intégrant diverses variables explicatives liées aux paramètres atmosphériques et océaniques.

## II. Délimitation de la zone d'étude

La zone d'étude s'étend entre **40° et 75° de longitude Est** et **15° et 30° de latitude A**, couvrant une partie des océans autour de Madagascar (Fig.1). L'analyse se concentre spécifiquement sur deux régions distinctes situées au Sud (A) et au Sud Est (B) de Madagascar, comme illustré dans la figure 1. Elles sont définies géographiquement comme suit :

- **Zone A** : Cette région est délimitée par les latitudes **30°S à 25°S** et les longitudes **40°E à 75°E**. Elle est représentée par un cadre bleu dans la figure.
- **Zone B** : Cette région, située plus à l'Est, s'étend également entre les latitudes **25°S à 18°S**, mais couvre une zone plus orientale, délimitée par les longitudes **47°E à 75°E**. Elle est illustrée par le cadre rouge.

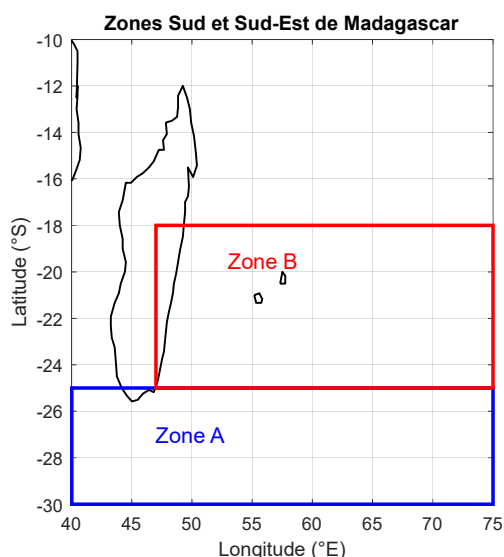


Figure 1 : zone d'étude

## III. Présentation des données utilisées

Les données utilisées dans cette étude proviennent de la base de données du Centre Européen pour les prévisions météorologiques à moyen terme (ECMWF), réputée pour sa précision et sa haute résolution spatiale et temporelle. Ces données couvrent la période allant de l'année 1993 à 2020 et sont disponibles sous forme de séries journalières en trois dimensions pour chaque variable étudiée (Longitude, Latitude et Temps).

#### IV. Variables étudiées

Les variables utilisées pour la modélisation de la hauteur des vagues incluent des paramètres atmosphériques et océaniques qui influencent directement ou indirectement la dynamique des vagues dans les zones étudiées. Ces variables sont :

- **Vent zonal (u, m/s)** : composante horizontale du vent dirigée de l'Ouest vers l'Est.
- **Vent méridional (v, m/s)** : composante verticale du vent dirigée du Sud au Nord.
- **Courants marins (m/s)** : incluant les composantes zonale et méridionale.
- **Hauteur de la surface de la mer (SSH, m)** : mesurant les anomalies de niveau marin.
- **Température de la surface de la mer (SST, °K).**
- **Salinité de surface (SSS, PSU).**
- **Pression atmosphérique (Pa)** : mesurée au niveau de la mer.
- **La hauteur des vagues** (variable dépendante) est mesurée en mètres (m) et représente la hauteur significative des vagues dans la région étudiée.

#### V. Prétraitement des données

Avant l'intégration des variables explicatives dans le modèle de régression, un ensemble d'étapes de prétraitement a été réalisé pour garantir la robustesse et la fiabilité des analyses. Ces étapes incluent le calcul des moyennes climatologiques, des tests de stationnarité, test de causalité de Granger, l'analyse de la multicollinéarité et la sélection rigoureuse des variables :

##### a) Calcul des moyennes climatologiques journalières

Afin de réduire les fluctuations saisonnières et de mieux comprendre les tendances climatiques de long terme, les moyennes climatologiques journalières ont été calculées en prenant en compte les dimensions spatiales (longitude et latitude) et temporelles (jours et années)). Ces moyennes ont été établies à partir des données couvrant toute la période d'étude (1993-2020), permettant ainsi d'avoir les variations journalières pour chaque point géographique.

La formule utilisée pour calculer la moyenne spatio-temporelle journalière est donnée par :

$$\overline{X}_{spatio-temporelle} = \frac{1}{N_{lat} \cdot N_{lon} \cdot N_t} \sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} \sum_{t=1}^{N_t} X_{i,j,t} \quad (1.1)$$

$N_{lat}, N_{lon}$  : 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.

##### b) Analyse de stationnarité des données temporelles

Après le calcul des moyennes des variables spatio-temporelles, nous avons vérifié la stationnarité des séries temporelles à l'aide du test de Dickey-Fuller augmenté (ADF). Ce test, proposé par Dickey et Fuller (1979), permet d'évaluer la présence d'une racine unitaire, indicatrice de non-stationnarité.

### Hypothèses :

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

### Critère de décision :

Une p-value inférieure à 0,05 conduit au rejet de l'hypothèse nulle, indiquant une stationnarité de la série. Dans le cas contraire, la série est considérée comme non stationnaire.

Une méthode de différenciation d'ordres 1 ou 2 va être appliquée à la série non stationnaire pour la rendre stationnaire.

#### c) *Analyse de causalité de Granger*

Pour identifier les relations causales entre les variables, le test de causalité de Granger a été réalisé. Ce test permet de déterminer si une variable peut être utilisée pour en prévoir une autre [1]. Les hypothèses :

- $H_0$  : La variable X n'a pas d'effet causal au sens de Granger de Y.
- $H_1$  : La variable X a un effet causal au sens de Granger de Y.

Seules les relations significatives ( $p < 0,05$ ) sont retenues pour l'intégration des variables dans le modèle. Les résultats montrent que certaines variables présentent une forte relation directionnelle, confirmant leur importance pour la modélisation.

#### d) *Analyse de la multicollinéarité (VIF)*

La multicollinéarité entre les variables explicatives a été examinée à l'aide du facteur d'inflation de la variance (VIF). Le VIF est calculé selon la formule suivante [2] :

$$VIF = \frac{1}{1 - R_i^2} \quad (1.2)$$

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

Le VIF importe dans quelle mesure une variable est linéairement corrélée avec les autres variables explicatives du modèle. Les critères utilisés sont :

- **VIF < 5** : Faible colinéarité, variable conservée.
- **$5 \leq VIF \leq 10$**  : Colinéarité modérée, analyse supplémentaire nécessaire.
- **VIF > 10** : Forte colinéarité, variable susceptible d'être exclue ou transformée.

Les variables présentant une forte multicollinéarité ont été traitées par :

- ✓ Exclusion de certaines variables redondantes.
- ✓ Transformation des variables corrélées

#### e) *Analyse de sélection et filtrage des variables pour la modélisation*

Suite aux différentes étapes de prétraitement des données, incluant les tests de stationnarité, l'analyse de causalité au sens de Granger et l'analyse de la multicollinéarité (VIF), un ensemble final de variables explicatives a été sélectionné pour la modélisation de la hauteur de vague.

### • Stationnarité des séries temporelles

Toutes les séries temporelles ont été initialement testées pour la stationnarité à l'aide du test Dickey-Fuller Augmenté (Augmented Dickey-Fuller ADF). Les résultats ont montré que toutes les variables étaient non stationnaires ( $p\text{-value} > 0.05$ ). Par conséquent, les séries ont été différenciées jusqu'à l'ordre 2 pour assurer leur stationnarité.

### • Causalité de Granger

Les tests de causalité de Granger ont été utilisés pour étudier les relations causales entre les différentes variables. Les résultats ont permis d'identifier plusieurs relations significatives, à la fois unidirectionnelles et bidirectionnelles, qui ont été prises en compte pour la modélisation de la hauteur de vague. En parallèle, l'analyse de la multicollinéarité via le VIF a été réalisée pour vérifier l'absence de colinéarité excessive entre les variables.

#### - Relations causales unidirectionnelles :

- ✓ **Pression → Hauteur de vague** ( $P\text{-value} = 0.0001639$ ) : La pression a un effet causal significatif sur la hauteur de vague.
- ✓ **Hauteur de vague → Vent zonal** ( $P\text{-value} = 0.036758$ ) : La hauteur de vague cause le vent zonal.
- ✓ **Courant zonal → Hauteur de vague** ( $P\text{-value} = 0.007969$ ) : Le courant zonal a un effet causal sur la hauteur de vague.

#### - Relations causales bidirectionnelles :

- ✓ **Hauteur de vague ↔ vent méridional** ( $P\text{-value} = 0.0036193 ; 0.0021941$ ) : La hauteur de vague et le vent méridional ont une relation causale bidirectionnelle.
- ✓ **Hauteur de vague ↔ Courant méridional** ( $P\text{-value} = 0.014627 ; 0.04326$ ) : La hauteur de vague et le courant zonal sont mutuellement causaux.
- ✓ **Hauteur de vague ↔ salinité** ( $P\text{-value} = 1.3952e-09 ; 0.010936$ ) : La hauteur de vague et la salinité (SSS) ont une relation causale bidirectionnelle.

#### - Relations non causales :

- ✓ **Hauteur de vague ~~↔~~ SST** ( $P\text{-value} = 0.072305 ; 0.27412$ ) : la hauteur de la vague et la température de la surface de la mer n'ont pas de relation causale dans les deux sens.
- ✓ **Hauteur de vague ~~↔~~ SSH** ( $P\text{-value} = 0.072305 ; 0.27412$ ) : la hauteur de la vague et la hauteur de la surface de la mer (SSH) n'ont pas de relation causale dans les deux sens.

### • Analyse de la multicollinéarité (VIF) :

L'analyse de la multicollinéarité a été effectuée en calculant le VIF pour chaque variable qui sera regroupé dans le tableau I suivant :

Tableau I : Valeur de VIF

VARIABLES	PRESSION	SSH	SST	VENT U	VENT V	COURANT U	COURANT V	SSS
VIF	7.73	49.11	7.90	28.98	2.55	10.00	4.16	6.26

Les résultats ont montré que certaines variables ont des valeurs de VIF élevées, indiquant une forte colinéarité. En particulier :

- **SSH** a montré un VIF extrêmement élevé (49.11), ce qui justifie son retrait de la modélisation.

- D'autres variables, comme **Vent zonal**, **Vent méridional** et **Courant zonal**, ont présenté des VIF relativement élevés, ce qui a nécessité une attention particulière pour éviter les effets de colinéarité excessive. Cependant, aucune exclusion supplémentaire n'a été faite, car les relations causales observées justifient leur inclusion dans le modèle.

Les variables finales retenues pour la modélisation sont celles qui ont montré des relations causales significatives et un VIF acceptable, assurant ainsi une modélisation robuste et fiable.

#### • Variables retenues pour la modélisation

Les variables retenues pour la modélisation de la hauteur de vague, après les tests de stationnarité, de causalité de Granger et l'analyse de la multicollinéarité, sont :

- **Pression**
- **Vent méridional**
- **Vent zonal**
- **Courant zonal**
- **Courant méridional**
- **Salinité (SSS)**
- **Température de la surface de la mer (SST)**

Ces variables ont montré des relations causales significatives avec la hauteur de vague et ne présentent pas de problèmes de multicollinéarité excessive, ce qui les rend adaptées pour la modélisation de la hauteur de vague dans le cadre de cette étude.

Particulièrement, la température de la surface de la mer (SST) n'a pas d'effet causal significatif sur la hauteur des vagues, bien que son VIF soit de 7,90, ce qui reste inférieur au seuil critique de 10. De même, le vent zonal présente un VIF élevé, mais cela s'explique par son lien de causalité inverse avec la hauteur des vagues. Par conséquent, ces variables sont incluses parmi les variables explicatives.

#### • Variables retirées :

La variable SSH a été retirée de la modélisation, car aucun effet causal significatif n'a été observé avec la hauteur de vague.

### VI. Modèle de régression linéaire multiple

Le modèle de régression linéaire multiple est utilisé pour prédire la hauteur des vagues en fonction des variables environnementales retenues. Ce modèle permet de comprendre l'impact de différents facteurs sur la hauteur des vagues tout en contrôlant les effets des autres variables.

#### ✓ Formulation du modèle

Le modèle de régression linéaire multiple est formulé comme suit :

$$\text{vague} = \alpha + \beta_1 \times \text{pression} + \beta_2 \times \text{SST} + \beta_3 \times \text{ventU} + \beta_4 \times \text{ventV} + \beta_5 \times \text{courantU} + \beta_6 \times \text{courantV} + \beta_7 \times \text{salinité} \quad (1.3)$$

Où :

- $\beta_1, \beta_2, \dots, \beta_7$  sont les coefficients estimés pour chaque variable indépendante,
- $\alpha$  est l'ordonnée à l'origine

Le tableau II donne les coefficients estimés pour chaque variable incluse dans le modèle de régression linéaire multiple, ainsi que leurs statistiques associées, à savoir :

- **Valeur estimée** : indique l'effet de chaque variable indépendante sur la hauteur de vague.
- **Erreur standard (SE)** : mesure la variabilité de l'estimation du coefficient.
- **t-statistique** : permet de tester si un coefficient est significativement différent de zéro.
- **p-value** : indique le niveau de significativité statistique de chaque variable.

Une p-value inférieure à 5% est généralement considérée comme significative.

Un t-statistique élevé indique une relation statistiquement significative avec la variable dépendante.

Tableau II : Valeurs des coefficients

	Estimation	Erreur standard	t-Stat	p-Value
$\alpha$	<b>-27.399</b>	23.543	-1.1638	0.24563
$\beta_1$	<b>0.00070552</b>	0.00011828	5.9649	8.3926e-09
$\beta_2$	<b>0.024321</b>	0.023995	1.0136	0.31177
$\beta_3$	<b>-0.097849</b>	0.02466	-3.9679	9.4964e-05
$\beta_4$	<b>0.18381</b>	0.024521	7.4962	1.1621e-12
$\beta_5$	<b>3.4968</b>	1.4664	2.3846	0.017849
$\beta_6$	<b>-3.9985</b>	1.2753	-3.1354	0.0019228
$\beta_7$	<b>-1.414</b>	0.26743	-5.2873	2.7232e-07

## VII. Analyse des coefficients

### a. Variables significatives au seuil de 5% :

Les variables **pression**, **vent zonal**, **vent méridien**, **salinité**, **courant méridional**, **courant zonal** et **salinité** sont statistiquement significatives, comme le montrent leurs p-values inférieures à 0.05. Le vent méridional influence indirectement la formation des vagues en transportant de l'énergie et en interagissant avec d'autres variables, telles que le vent zonal et les courants marins [3]. Il est associé à des régimes climatiques comme les Alizés et les Moussons, qui modifient les conditions océaniques [4]. Ces variables contribuent de manière importante à la prédiction de la hauteur des vagues.

### b. Variables non significatives mais incluses :

Bien que la variable SST ne soit pas statistiquement significatif (p-value = 0.31177), il est conservé pour assurer une interprétation physique cohérente des phénomènes étudiés [5] à cause de la valeur de VIF (inférieure à 10).

## VIII. Validation croisée

La validation croisée est une méthode d'évaluation de la performance des modèles et pour éviter le surapprentissage (overfitting). Dans le cadre de cette méthode, les données sont généralement divisées en deux ensembles :

- **Ensemble d'entraînement** : Cet ensemble représente 70 % des données et est utilisé pour ajuster le modèle. Il permet d'apprendre les relations et les patterns présents dans les données.
- **Ensemble de test** : Composé des 30 % restants des données, cet ensemble est utilisé pour évaluer la performance du modèle sur des données non vues. Cela permet de mesurer la capacité de généralisation du modèle et d'obtenir une estimation plus précise de sa performance dans des situations réelles.

Cette approche garantit que le modèle est testé sur des données qu'il n'a jamais rencontrées auparavant, offrant ainsi une évaluation rigoureuse de sa robustesse et de son efficacité [6] [7] [8] [9].

## IX. Analyse des résidus

L'évaluation des résidus du modèle a été réalisée afin de vérifier les hypothèses nécessaires pour la validité statistique de la régression. Deux analyses graphiques ont été utilisées : un histogramme des résidus et un graphique des résidus en fonction des valeurs prédites.

L'histogramme des résidus (Fig. 2, gauche) montre une distribution centrée autour de zéro, avec une symétrie apparente. Les résidus se situent dans l'intervalle  $[-0.2, 0.3]$ .

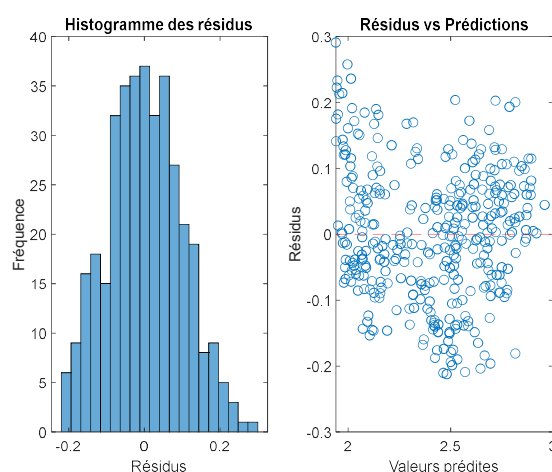


Figure 2 : Représentation des résidus

### Test de normalité des résidus :

Le test de Kolmogorov-Smirnov (K-S) appliqué aux résidus du modèle a permis d'évaluer leur conformité à une distribution normale.

Les résultats obtenus sont :

- Hypothèse nulle ( $H_0$ ) : Les résidus ne suivent pas une distribution normale.
- Statistique du test (T-stat): T-stat=1
- P-value :  $p=0$



Ces résultats indiquent un rejet de  $H_0$  au seuil de signification standard ( $\alpha = 0.05$ ). La valeur extrêmement faible de p-value ( $p \ll 0.05$ ) indique que les résidus suivent une loi normale.

L'analyse des résidus montre que les hypothèses fondamentales pour la validité du modèle de régression sont globalement satisfaisantes. Le modèle peut donc être considéré comme fiable pour les données utilisées.

#### X. Performance du modèle : indicateurs de qualité de l'ajustement

Pour évaluer la qualité du modèle de régression linéaire multiple, plusieurs indicateurs statistiques ont été calculés :  $R^2$ ,  $R^2$  ajusté et RMSE. Avec un  $R^2$  de **0.8916**, le modèle explique **89.16 %** de la variance des hauteurs de vagues, indiquant un excellent ajustement. Le  $R^2$  ajusté, à **0.8886**, confirme la pertinence des variables explicatives tout en évitant les surajustements.

Le **RMSE (Root Mean Square Error)** de notre modèle, évalué à **0.1044 m**, représente la racine carrée de la moyenne des carrés des écarts entre les valeurs observées ( $y_i$ ) et les valeurs prédites ( $\hat{y}_i$ ) [10]. Il est calculé à l'aide de la formule suivante :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1.4)$$

où  $n$  est le nombre total d'observations.

$y_i$  : les valeurs observées

$\hat{y}_i$  : les valeurs prédites

Pour évaluer la performance relative du modèle, le pourcentage d'erreur basé sur le **RMSE** peut être calculé à l'aide de la formule suivante :

$$\text{Pourcentage d'erreur} = \left( \frac{RMSE}{H_{max} - H_{min}} \right) \times 100 \quad (1.5)$$

$H_{max}$  : Hauteur maximale de la vague

$H_{min}$  : Hauteur minimale de la vague

Le pourcentage d'erreur, calculé à partir du RMSE, est d'environ **8.7%**. Ce pourcentage indique que les prévisions du modèle sont relativement précises (inférieure à 10%). En d'autres termes, les erreurs de prévision sont faibles en proportion des valeurs réelles, ce qui renforce la fiabilité du modèle dans ses estimations.

#### XI. Présentation des figures issues du modèle

La figure 3 représente la courbe issue du modèle ainsi que celle des valeurs observées.

- **Courbes prédictives (en bleu)** : Ces courbes montrent les valeurs estimées par le modèle, facilitant l'identification des tendances générales.
- **Données observées (en rouge)** : Elles représentent les valeurs réelles mesurées à chaque intervalle de temps, offrant une comparaison directe avec les prévisions.

- **Points de test observés (en vert) :** Ces points correspondent aux données réelles non utilisées pour l'ajustement du modèle, permettant d'évaluer sa capacité à généraliser ses prédictions.
- **Points de test prédits (en noir) :** Ils illustrent les prévisions du modèle sur le jeu de données de test, fournissant un aperçu de sa précision sur des observations indépendantes.

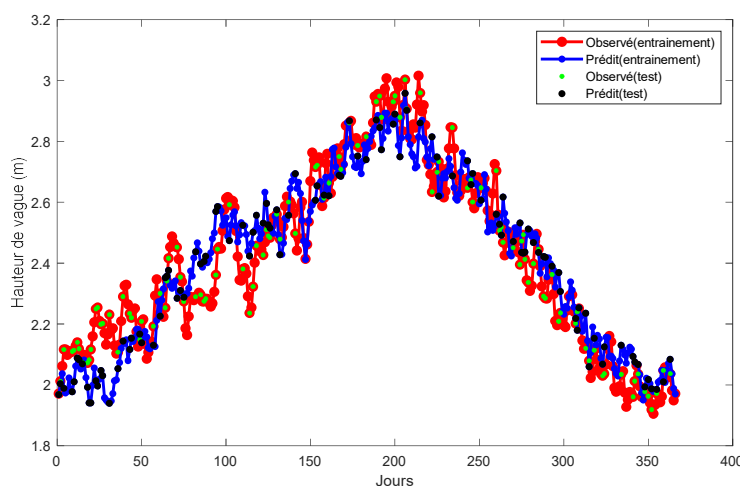


Figure 3 : Représentation graphique du modèle et de la valeur observée

Cette représentation permet de comparer directement les prévisions du modèle et les valeurs réelles, tout en offrant un aperçu clair de la capacité du modèle à ajuster et à prédire avec précision.

## CONCLUSION :

Dans cette étude, nous avons construit un modèle de régression linéaire multiple pour analyser les variations dans les moyennes climatologiques des hauteurs de vagues au Sud et Sud-Est de Madagascar. L'analyse préliminaire nous a permis d'éliminer la variable SSH.

Les variables retenues nous ont donné un modèle de régression linéaire multiple. Avec un coefficient de détermination ( $R^2$ ) de 0.8916, il démontre une capacité explicative élevée, indiquant que 89.16 % de la variance des hauteurs de vagues est expliquée par ces variables. Le  $R^2$  ajusté, évalué à 0.8886, renforce la pertinence des variables sélectionnées tout en minimisant les risques de surajustement. En outre, le RMSE (Root Mean Square Error) est de 0.1044 m, ce qui traduit un faible écart type des résidus. Le pourcentage d'erreur associé est estimé à 8.7 %, ce qui témoigne de la précision du modèle. Les résultats soulignent l'importance des interactions entre les paramètres atmosphériques et océaniques dans la variabilité des hauteurs de vagues.

En guise de perspective, l'intégration de modèles non linéaires ou de techniques d'apprentissage automatique pourrait enrichir les prédictions en capturant des relations plus complexes.

## REFERENCES

- [1] C. Granger, «Investigating causal relations by econometric models and cross-spectral methods,» *Econometrica*, p. 424–438, 1969.
- [2] M. P. Jacob et K. Eda , «Qu'est-ce que la multicollinéarité ?,» 21 novembre 2023. [En ligne]. Available: <https://www.ibm.com/fr-fr/topics/multicollinearity#f15>.
- [3] J. Smith, L. Johnson et R. Wang, «Oceanic Wind Dynamics and Wave Formation,» *Journal of Marine Science*, 2020.
- [4] A. Dupont, «Climate Regimes and Ocean Conditions,» *Climate Research Letters*, 2019.
- [5] P. Martin et C. Lefèvre, «Statistical Analysis of Wind Effects on Waves,» *Oceanography Review*, 2021.
- [6] R. Kohavi, «A study of cross-validation and bootstrap for accuracy estimation and model selection,» *International Joint Conference on Artificial Intelligence*, 1995.
- [7] S. Arlot et A. Celisse, «A survey of cross-validation procedures,» *Statistics Surveys*, pp. 40-79, 2010.
- [8] T. Hastie, R. Tibshirani et J. Friedman, «The Elements of Statistical Learning: Data Mining, Inference, and Prediction,» *Springer*, 2009.
- [9] C. M. Bishop, «Pattern Recognition and Machine Learning,» *Springer*, 2006.
- [10] C. David , J. W. Matthijs et J. Giuseppe , «The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,» *PeerJ computer Science*, pp. 1-24, 2021.