

Régionalisation Multivariée Des Conditions Climatiques Locales Dans Le Grand Sud De Madagascar Par ACP Et K-Means

[Multivariate Regionalization Of Local Climate Conditions In The Grand South Of Madagascar Using PCA And K-Means]

Omar Naveroson RABENOMENISOANIRAINY¹, Honoré RAZAFIMANDIMBY², Jean Eugène RANDRIANANTENAINA³, Adolphe Andriamanga RATIARISON⁴

^{1,2,3}Université de Tuléar, Laboratoire de la Dynamique de l'Atmosphère, du Climat et des Océans (DyACO), B.P. 185, Tuléar, Madagascar.

⁴Université d'Antananarivo, Laboratoire de la Dynamique de l'Atmosphère, du Climat et des Océans (DyACO), B.P. 906, Antananarivo, Madagascar.

Auteur correspondant: Omar Naveroson RABENOMENISOANIRAINY, naverosonomar@gmail.com



Résumé : Cette étude propose une régionalisation climatique multivariée du Grand Sud de Madagascar à partir de six variables : précipitations (P), température de l'air (T), évaporation (E), rayonnement solaire (RS), vitesse du vent (V) et humidité spécifique (HUSS). Les données mensuelles couvrent 1978-2024 (564 mois) sur une grille de 25 × 25 pixels. Après masque spatial, 400 pixels sont conservés. Les séries sont harmonisées, agrégées annuellement puis standardisées temporellement pixel par pixel. Les six matrices de 400 × 47 sont concaténées dans l'ordre P | T | E | RS | V | HUSS pour former une matrice de 400 × 282. L'ACP retient 12 composantes expliquant 81,02 % de la variance cumulée. Le nombre de classes est testé de k = 2 à 8. La méthode du coude et la silhouette moyenne convergent vers k = 3 ; pour k = 3, la silhouette moyenne vaut 0,3091. Le K-means distingue alors une zone sèche, une zone intermédiaire et une zone humide. Les profils physiques montrent un gradient cohérent des précipitations et de l'humidité spécifique, du cluster 2 vers le cluster 3. Cette régionalisation constitue une base reproductible pour l'analyse spatiale de la sécheresse dans le Grand Sud de Madagascar.

Mots-clés : sécheresse ; ERA5-Land ; classification spatiale ; humidité spécifique ; anomalies standardisées ; variabilité climatique.

Abstract: Objective: This study proposes a multivariate climate regionalization of the Grand South of Madagascar using six climatic variables: precipitation, air temperature, evaporation, solar radiation, wind speed, and specific humidity. Methods: Monthly data cover the period from 1978 to 2024, representing 564 months on a grid of 25 by 25 pixels. After application of the spatial mask, 400 pixels were retained. The time series were harmonized on a common grid, aggregated annually, and standardized temporally for each pixel. The six annual matrices were then concatenated to form a multivariate matrix containing 400 pixels and 282 variables. Principal component analysis was applied to reduce dimensionality, and 12 components were retained, explaining 81.02 percent of the cumulative variance. Results: The number of climatic groups was evaluated from two to eight using the elbow method and the mean silhouette coefficient. Both criteria supported a three-group solution, for which the mean silhouette coefficient reached 0.3091. K-means clustering identified a dry zone, an intermediate zone, and a humid zone. The physical climatic profiles revealed a coherent gradient in precipitation and specific humidity between the three groups. Conclusion: These results provide a reproducible climatic regionalization and a methodological basis for subsequent spatial analyses of drought in the Grand South of Madagascar.

Keywords: drought; ERA5-Land; spatial classification; specific humidity; standardized anomalies; climate variability.

1. Introduction

Le Grand Sud de Madagascar présente de forts contrastes climatiques et une exposition récurrente à la sécheresse. Les travaux récents confirment la vulnérabilité particulière du sud de Madagascar aux déficits pluviométriques, aux sécheresses prolongées et aux événements chauds et secs, avec des conséquences marquées sur les écosystèmes et les conditions de surface [1], [2]. Ces contrastes ne dépendent pas seulement des précipitations : température, évaporation, rayonnement solaire, vent et humidité atmosphérique contribuent également à différencier les conditions locales. Une approche multivariée est donc pertinente pour construire des zones climatiques homogènes.

L'analyse en composantes principales (ACP) constitue une méthode de référence pour réduire la dimension d'un jeu de données multivarié tout en conservant une part importante de sa variabilité [3]. K-means permet ensuite de regrouper les observations selon leur proximité multivariée [4]. L'objectif est d'identifier des zones climatiques reproductibles sur 1978-2024 en combinant harmonisation, standardisation, ACP et classification non supervisée.

2. Matériels et méthodes

2.1. Zone d'étude et données

La zone étudiée couvre approximativement 43°-49° E et 20°-26° S. Les données mensuelles ERA5-Land couvrent janvier 1978 à décembre 2024. ERA5-Land est une réanalyse terrestre à haute résolution développée pour fournir des champs cohérents sur plusieurs décennies et adaptée aux applications hydrologiques et climatiques [5]. La grille de travail comporte 625 pixels (25 × 25), dont 400 appartiennent au domaine et 225 sont masqués. Les six variables utilisées sont : précipitations (P), température (T), évaporation (E), rayonnement solaire (RS), vent (V) et humidité spécifique (HUSS).

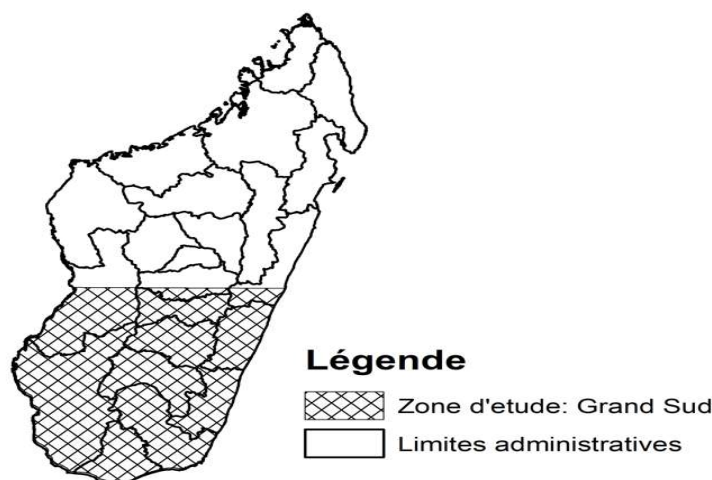


Figure 1. Localisation géographique du Grand Sud de Madagascar.

2.2. Harmonisation et agrégation annuelle

Les six séries ont été harmonisées sur la même grille, le même masque et les mêmes 564 mois. L'orientation spatiale commune est nord en haut, sud en bas, ouest à gauche et est à droite ; aucun retournement supplémentaire n'est appliqué après harmonisation.

Pour une variable v , au pixel i , année y et mois m , la somme et la moyenne annuelles sont :

$$X_{v,i,y}^{(S)} = \sum_{m=1}^{12} X_{v,i,y,m} \quad (1)$$

$$X_{v,i,y}^{(M)} = \frac{1}{12} \sum_{m=1}^{12} X_{v,i,y,m} \quad (2)$$

Comme $X^{(S)} = 12X^{(M)}$, les deux agrégations deviennent équivalentes après standardisation. La branche issue des moyennes annuelles est retenue pour la suite.

2.3. Standardisation et matrice multivariée

La standardisation est réalisée temporellement, pixel par pixel, sur les 47 années :

$$Z_{v,i,y} = \frac{X_{v,i,y} - \bar{X}_{v,i}}{s_{v,i}} \quad (3)$$

$$s_{v,i} = \sqrt{\frac{1}{46} \sum_{y=1}^{47} (X_{v,i,y} - \bar{X}_{v,i})^2} \quad (4)$$

Chaque variable fournit ensuite une matrice $Z_v \in \mathbb{R}^{400 \times 47}$. Les six matrices sont concaténées :

$$X = [Z_P \ Z_T \ Z_E \ Z_{RS} \ Z_V \ Z_{HUSS}] \in \mathbb{R}^{400 \times 282}.$$

Aucune nouvelle standardisation n'est effectuée après cette concaténation.

2.4. Analyse en composantes principales

L'ACP est appliquée directement à la matrice 400×282 . Après centrage, la covariance est calculée par

$$C = \frac{1}{n-1} X_c^T X_c \quad (5)$$

et les axes vérifient $Ca_k = \lambda_k a_k$. La variance expliquée par l'axe k est

$$VE_k = \frac{\lambda_k}{\sum_j \lambda_j} \times 100 \quad (6)$$

Le nombre de composantes retenues est le plus petit q tel que la variance cumulée atteigne au moins 80 %. Ici, 12 composantes sont retenues, pour 81,02 % de variance cumulée.

2.5. K-means et validation du nombre de classes

K-means est appliqué aux scores des 12 composantes principales en utilisant la distance euclidienne au carré, 100 répliques et un maximum de 1000 itérations. Cette méthode de partitionnement vise à minimiser la variabilité intra-classe autour des centroïdes [4]. Afin d'assurer la reproductibilité de la classification, le générateur pseudo-aléatoire de MATLAB est initialisé avec la graine 1 selon l'algorithme Mersenne Twister [`rng(1,'twister')`]. Le critère minimisé est :

$$J = \sum_{r=1}^K \sum_{i \in C_r} \|s_i - \mu_r\|^2 \quad (7)$$

Le nombre de classes est testé pour $k = 2, \dots, 8$ par deux critères complémentaires. Le premier est le WCSS, utilisé pour identifier un changement de pente selon le principe de la méthode du coude. Le second est l'indice de silhouette, qui évalue simultanément la cohésion interne des classes et leur séparation relative [6] :

$$Sil(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \quad (8)$$

Une silhouette positive indique que le pixel est, en moyenne, plus proche de sa propre classe que des autres classes.

2.6. Caractérisation des zones

Après classification, les groupes sont caractérisés par leurs moyennes climatiques physiques sur 1978-2024. P et E sont exprimées en cumuls annuels ; T, RS, V et HUSS en moyennes annuelles. Les noms « sèche », « intermédiaire » et « humide » sont attribués selon les profils obtenus.

3. Résultats

3.1. Structure factorielle

Les trois premières composantes expliquent respectivement 28,01 %, 18,44 % et 9,66 % de la variance, soit 56,10 % cumulés. La variance cumulée atteint 79,65 % à CP11 et 81,02 % à CP12 ; 12 axes sont donc nécessaires pour franchir le seuil de 80 %.

Tableau 1. Variance expliquée par les 12 composantes retenues.

CP	1	2	3	4	5	6	7	8	9	10	11	12
Variance (%)	28,01	18,44	9,66	4,65	4,43	3,53	3,25	2,47	1,93	1,8	1,48	1,37
Cumul (%)	28,01	46,44	56,1	60,76	65,19	68,71	71,97	74,44	76,37	78,17	79,65	81,02

3.2. Validation de k = 3

La WCSS diminue de 28 972,40 pour k = 2 à 21 756,15 pour k = 3, soit une baisse de 24,91 %. Le gain n'est plus que de 16,76 % entre k = 3 et k = 4, puis il diminue progressivement. La courbure discrète est maximale à k = 3 (3570,07), ce qui localise le coude principal à trois classes.

La silhouette moyenne est également maximale pour k = 3 (0,3091), devant k = 2 (0,2980), k = 4 (0,2984) et les autres partitions testées. Pour k = 3, la médiane est 0,3386 et seulement 2,75 % des pixels ont une silhouette négative. La séparation est donc modérée, mais les deux critères convergent vers la même solution.

Tableau 2. Validation du nombre de classes.

k	2	3	4	5	6	7	8
WCSS	28 972,40	21 756,15	18 109,98	16 148,99	14 485,60	13 167,41	12 040,37
Baisse WCSS (%)	—	24,91	16,76	10,83	10,3	9,1	8,56
Silhouette moyenne	0,298	0,3091	0,2984	0,2954	0,2876	0,2786	0,2996

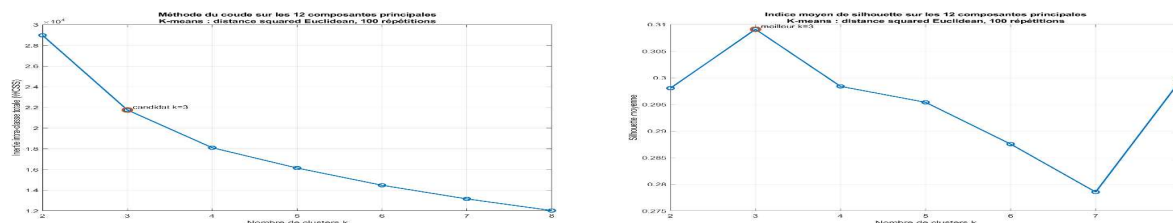


Figure 2. Détermination du nombre de zones par la méthode du coude et l'indice de silhouette.

3.3. Répartition spatiale des zones climatiques

La classification forme trois ensembles bien organisés dans l'espace. La zone sèche se trouve surtout au sud et au sud-ouest. La zone intermédiaire couvre une grande partie de l'ouest, du nord-ouest et du centre. La zone humide se situe surtout à l'est et au nord-est. Les espaces blancs correspondent aux points retirés par le masque spatial. Ils ne forment pas une quatrième zone. Malgré la forme carrée des pixels, les trois ensembles restent globalement continus.

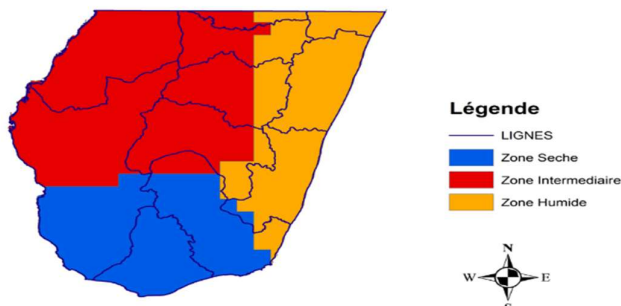


Figure 3. Carte finale des trois zones climatiques régionalisées dans le Grand Sud de Madagascar.

3.4. Classification et signatures climatiques

La partition finale comprend 111 pixels dans le cluster 2, 180 dans le cluster 1 et 109 dans le cluster 3. L'attribution des dénominations climatiques repose sur l'ensemble des caractéristiques physiques présentées au Tableau 3, et non sur une seule variable. Le cluster 2 présente les plus faibles précipitations (121,28 mm/an) et la plus faible humidité spécifique (0,00977 kg/kg), associées à la vitesse du vent la plus élevée (2,04 m/s) : il correspond ainsi à la zone sèche. Le cluster 3 présente au contraire les précipitations les plus fortes (351,30 mm/an), l'humidité spécifique la plus élevée (0,01134 kg/kg), la température moyenne la plus faible (20,78 °C) et les vents les plus faibles (1,08 m/s) : il est interprété comme la zone humide. Le cluster 1 occupe une position intermédiaire pour les précipitations (177,98 mm/an), l'humidité spécifique (0,01013 kg/kg) et le vent (1,86 m/s), tout en présentant les valeurs moyennes les plus élevées de température (23,28 °C) et de rayonnement solaire ; il constitue donc une zone de transition climatique. Les précipitations et l'humidité spécifique définissent notamment un gradient net du cluster 2 vers le cluster 3, ce qui renforce l'interprétation sèche–intermédiaire–humide.

Tableau 3. Principales caractéristiques climatiques physiques des trois zones.

Zone	Pixels	P (mm/an)	T (°C)	E (mm/an)	RS (J/m ²)	V (m/s)	HUSS (kg/kg)
Sèche	111	121,28	23,16	222,13	17413294,52	2,04	0,00977
Intermédiaire	180	177,98	23,28	266,51	17790699,42	1,86	0,01013
Humide	109	351,30	20,78	304,92	15288343,80	1,08	0,01134

3.5. Évolution annuelle des variables par zone

Les séries annuelles montrent que les contrastes entre zones restent globalement persistants entre 1978 et 2024, malgré une variabilité interannuelle parfois marquée. Les interprétations ci-dessous décrivent les signatures relatives des trois zones ; elles ne constituent pas, à elles seules, des tests statistiques de tendance, de causalité ou de synchronisation.

a) Précipitations.

La zone humide conserve les cumuls les plus élevés sur l'ensemble de la période, la zone intermédiaire occupe une position centrale et la zone sèche demeure la moins arrosée (Figure-a). Cette hiérarchie persistante constitue l'un des principaux marqueurs de la régionalisation. Les fluctuations parfois simultanées dans les trois zones suggèrent l'influence d'une variabilité climatique régionale commune, tandis que l'amplitude différente des réponses traduit des conditions locales distinctes.

b) Température.

La zone intermédiaire présente généralement les températures les plus élevées, suivie de près par la zone sèche, alors que la zone humide demeure sensiblement plus fraîche (Figure-b). Ce contraste thermique complète le gradient pluviométrique : la zone la plus humide n'est pas seulement davantage arrosée, elle est également plus fraîche en moyenne. Une hausse apparente des courbes peut être observée visuellement, mais sa significativité devra être vérifiée par un test de tendance avant toute conclusion sur une évolution à long terme.

c) Évaporation.

L'évaporation annuelle est la plus élevée dans la zone humide, intermédiaire dans la zone de transition et la plus faible dans la zone sèche (Figure-c). Ce résultat ne doit pas être interprété comme une contradiction avec l'aridité : l'évaporation réelle dépend à la fois de la demande atmosphérique et de la quantité d'eau effectivement disponible à la surface. Dans les secteurs secs, une disponibilité hydrique limitée peut restreindre le flux d'évaporation, même lorsque les conditions atmosphériques favorisent le dessèchement.

d) Rayonnement solaire.

Le rayonnement solaire est en moyenne le plus élevé dans la zone intermédiaire, suivi de la zone sèche, tandis que la zone humide présente les valeurs les plus faibles (Figure-d). Cette organisation est compatible avec une atténuation plus forte du rayonnement incident dans les secteurs plus humides, possiblement associée à une nébulosité plus fréquente ; cette interprétation reste toutefois à confirmer par des données directes de couverture nuageuse.

e) Vent.

La zone sèche présente les vitesses de vent les plus élevées, la zone intermédiaire des valeurs moyennes et la zone humide les vents les plus faibles (Figure-e). Cette composante aérodynamique différencie nettement les trois signatures climatiques. Dans la zone sèche, des vents plus soutenus peuvent contribuer à renforcer les échanges turbulents et le potentiel de

dessèchement des surfaces, sans permettre à eux seuls de conclure sur l'évaporation réelle.

f) Humidité spécifique.

L'humidité spécifique fournit l'un des contrastes les plus réguliers entre les groupes : la zone humide conserve les valeurs les plus élevées, la zone intermédiaire reste en position médiane et la zone sèche les valeurs les plus faibles (Figure-f). Associé au gradient des précipitations, ce comportement constitue un argument physique majeur pour l'interprétation des trois clusters comme zones sèche, intermédiaire et humide.

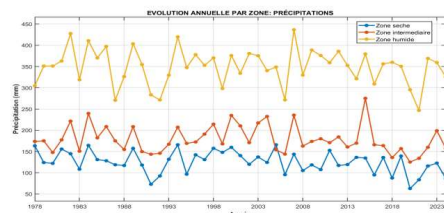


Fig-a)

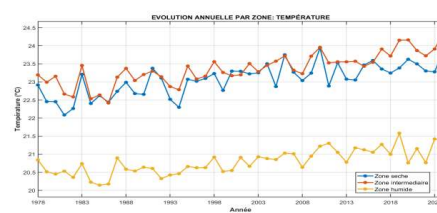


Fig-b)

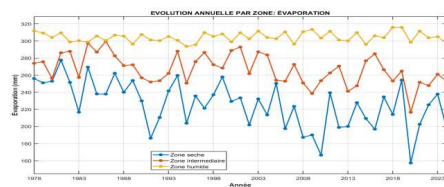


Fig-c)

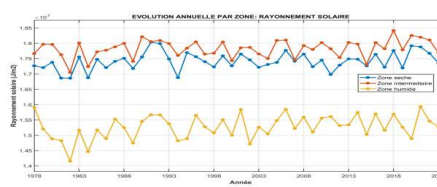


Fig-d)

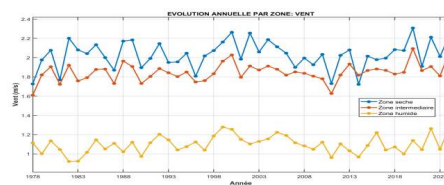


Fig-e)

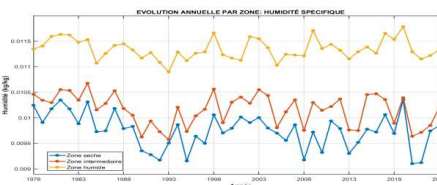


Fig-f)

Figure 4. Évolution annuelle des six variables climatiques dans les trois zones sur la période 1978-2024.

4. Discussion

Les résultats soutiennent une partition en trois zones par deux critères statistiques indépendants. Le coude est marqué à $k = 3$ et la silhouette moyenne y atteint son maximum. La valeur de silhouette (0,3091) ne traduit toutefois pas des groupes parfaitement séparés ; elle suggère des transitions entre régimes, ce qui est cohérent avec la continuité spatiale du climat.

L'utilisation de six variables permet une lecture plus complète qu'une régionalisation uniquement pluviométrique. La cohérence du gradient de P et $HUSS$ renforce l'interprétation sèche-intermédiaire-humide. La zone sèche est moins arrosée, moins humide dans l'atmosphère et plus ventilée ; la zone humide montre le profil inverse ; la zone intermédiaire constitue une transition multivariable. Cette organisation est cohérente avec les travaux décrivant le sud et le sud-ouest de Madagascar comme des secteurs particulièrement arides et soumis à une forte variabilité pluviométrique, ainsi qu'avec les études récentes montrant une vulnérabilité accrue du Sud aux sécheresses et aux extrêmes chauds et secs [1], [2].

La chaîne est reproductible : même grille, même masque, même période, ordre spatial contrôlé, standardisation unique et paramètres ACP/K-means explicités. Elle présente néanmoins une limite importante : la standardisation temporelle pixel par

pixel met l'accent sur la variabilité relative et atténue une partie des contrastes de niveau moyen entre pixels. Une validation indépendante par l'indice d'aridité, SPI/SPEI ou des observations de stations reste donc souhaitable. Pour la validation par l'aridité, la base Global Aridity Index and Potential Evapotranspiration fournit notamment une référence hydroclimatique indépendante fondée sur le rapport précipitations/évapotranspiration de référence [7].

5. Conclusion

La régionalisation multivariée du Grand Sud de Madagascar a été construite à partir de six variables climatiques sur 1978-2024. Après harmonisation, agrégation et standardisation, la matrice finale comprend 400 pixels et 282 dimensions. Douze composantes principales expliquent 81,02 % de la variance cumulée.

Le choix de trois classes est soutenu simultanément par la méthode du coude et par la silhouette moyenne, maximale pour $k = 3$. K-means distingue ainsi une zone sèche (cluster 2), une zone intermédiaire (cluster 1) et une zone humide (cluster 3). Les profils physiques présentent un gradient cohérent de précipitation et d'humidité spécifique.

Cette régionalisation fournit une base reproductible pour l'étude spatiale de la sécheresse. Une validation indépendante par des indicateurs d'aridité et de sécheresse permettra de renforcer l'évaluation de la robustesse des zones climatiques obtenues.

Références

- [1] Randriamahefasoa, T. S. M., & Reason, C. J. C. (2017). Interannual variability of rainfall characteristics over southwestern Madagascar. *Theoretical and Applied Climatology*, 128, 421–437. <https://doi.org/10.1007/s00704-015-1719-0>
- [2] Barimalala, R., Wainwright, C., Kolstad, E. W., & Demissie, T. D. (2024). The 2019–21 drought in southern Madagascar. *Weather and Climate Extremes*, 46, 100723. <https://doi.org/10.1016/j.wace.2024.100723>
- [3] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [4] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- [5] Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13, 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- [6] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [7] Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9, 409. <https://doi.org/10.1038/s41597-022-01493-1>