

Prévision de sécheresse en utilisant une architecture neuronale hybride intégrant des séries temporelles et des données statiques

Julian Agudelo^{1,2}, Vincent Guigue¹, Cristina Manfredotti¹, Hadrien Piot²

¹ AgroParisTech, UMR MIA Paris-Saclay.

² Agrial, Service Agronomique.

j.agudelo@agrial.com

Résumé

Une prévision fiable est essentielle aux systèmes d'alerte précoce et à la gestion adaptative des sécheresses. La plupart des approches d'apprentissage profond existantes se limitent souvent à des régions homogènes et à des données à modalité unique. Cet article propose une architecture neuronale hybride intégrant des séries temporelles et des données statiques, atteignant des performances de l'état de l'art sur l'ensemble de données publiques DroughtED. Nos résultats illustrent le potentiel de la conception de modèles neuronaux pour le traitement de données hétérogènes dans les tâches liées au climat et présentent une prédiction fiable des catégories de l'USDM, une métrique de sécheresse mise au point par des experts.

Mots-clés

Prévision de la sécheresse, séries temporelles, données hétérogènes, réseaux neuronaux, LSTM, mécanisme d'attention, architecture hybride.

Abstract

Reliable forecasting is critical for early warning systems and adaptive drought management. Most previous deep learning approaches focus solely on homogeneous regions and rely on single-modality data. This paper presents a hybrid neural architecture that integrates time series and static data, achieving state-of-the-art performance on the DroughtED dataset. Our results illustrate the potential of designing neural models for the treatment of heterogeneous data in climate related tasks and present reliable prediction of USDM categories, an expert-informed drought metric.

Keywords

Drought Forecasting, time Series, heterogeneous data, neural networks, LSTM, attention mechanism, hybrid architecture.

1 Introduction

La sécheresse est un phénomène naturel caractérisé par une période prolongée de précipitations inférieures à la moyenne, ce qui entraîne des déséquilibres hydrologiques importants ayant un impact négatif sur les ressources terrestres [22]. Les sécheresses sont souvent classées en

différentes catégories : météorologiques, agricoles, hydrologiques et socio-économiques [12, 7, 28]. Cette classification reflète leurs impacts variables sur les différents systèmes. À mesure que le réchauffement climatique s'intensifie, la fréquence, la durée et la gravité des événements de sécheresse augmentent, exacerbant les vulnérabilités. Compte tenu des multiples facettes des sécheresses et de leurs effets à travers diverses échelles spatiales et temporelles, une prévision fiable est essentielle pour les systèmes d'alerte précoce et la gestion adaptative des ressources. Des prévisions efficaces aident à réduire les impacts sur l'approvisionnement en eau, l'agriculture, les écosystèmes et les communautés [1].

Plusieurs indices sont couramment utilisés pour évaluer les conditions de sécheresse. Parmi les exemples les plus fréquemment cités, figurent l'Indice de Sévérité de la Sécheresse de Palmer (PDSI) [21], l'Indice Standardisé de Précipitation (SPI) [16], et l'Indice Standardisé de Précipitation et d'Évapotranspiration (SPEI) [26]. Ces indices dépendent des variables météorologiques pour quantifier les écarts par rapport aux normes climatiques. En revanche, les catégories du U.S. Drought Monitor (USDM) offrent une évaluation complète en intégrant des données hydrologiques, climatiques et météorologiques avec des connaissances expertes, permettant ainsi de saisir une vue d'ensemble des impacts de la sécheresse [24]. Nous utilisons les catégories de l'USDM comme variable cible pour cette étude en raison de leur nature globale et de leur intégration dans DroughtED [17].

Divers auteurs ont utilisé l'apprentissage automatique pour prédire des indices de sécheresse [19]. Traditionnellement, l'extraction manuelle de caractéristiques a été utilisée pour alimenter des algorithmes classiques d'apprentissage automatique [9]. Ces dernières années, une évolution significative a été observée vers les approches d'apprentissage profond, qui utilisent l'apprentissage par représentation pour extraire automatiquement des caractéristiques [31]. Un nombre considérable d'études antérieures utilisant des techniques d'apprentissage profond pour la prévision de la sécheresse se sont concentrées sur des régions homogènes [6]. D'autres ont principalement utilisé des données de structure uniforme, comme des images [5] ou des séries temporelles seules [14, 27].

Cet article utilise le jeu de données DroughtED [17] et présente une nouvelle approche de modélisation pour la prévision de la sécheresse avec des données de structures hétérogènes. Les contributions de cette recherche sont triples. (1) Nous introduisons une architecture neuronale intégrant des séries temporelles et des données statiques à travers des Réseaux de neurones à action directe (FFNN), des réseaux récurrents à mémoire court et long terme (LSTM), des embeddings catégoriels et un mécanisme d'attention. Nous comparons notre modèle aux modèles de référence du jeu de données. (2) Nous réalisons une étude d'ablation pour évaluer la contribution de chaque composant dans le modèle proposé. (3) Nous appliquons des techniques de visualisation sur les états latents pour effectuer une introspection du modèle.

2 L'ensemble de données

DroughtED est un vaste ensemble de données conçu pour la prévision des conditions de sécheresse aux États-Unis, en intégrant des caractéristiques spatiales et temporelles [17]. Il comprend des séries temporelles météorologiques, des caractéristiques physiques des sols et des informations historiques sur l'intensité de la sécheresse au niveau des comtés. Les données météorologiques proviennent du projet NASA Prediction Of Worldwide Energy Resources (NASA POWER) [30], les propriétés du sol sont issues de la Harmonized World Soil Database [18] et les évaluations de l'intensité de la sécheresse sont extraites de l'USDM [24].

Les données relatives à la sécheresse sont des indicateurs ordinaux mesurés localement. Ces indicateurs sont réduits à des valeurs moyennes continues, puis agrégés au niveau des comtés. Les valeurs cibles sont 6 valeurs continues à l'échelle du comté, correspondant à 6 semaines consécutives. Nous avons des données $\mathbf{Y} \in \mathbb{R}^{N \times 6}$ où N correspond au nombre de paires (comté, pas de temps) notées (c, t) dans la suite. Pour prédire ces cibles, nous disposons de descripteurs statiques $\mathbf{S} \in \mathbb{R}^{C \times f}$, où f est le nombre de caractéristiques décrivant les propriétés physiques du sol, et C le nombre de comtés. Il convient de noter que les descripteurs seront divisés en caractéristiques catégorielles $\mathbf{s}_d \in \mathbb{R}^{f_d}$ et numériques $\mathbf{s}_n \in \mathbb{R}^{f_n}$ dans la suite du document. Les données météorologiques sont représentées sous forme de séries temporelles multivariées et regroupées dans un tenseur $\mathbf{X} \in \mathbb{R}^{C \times P \times M}$: pour chaque comté, il y a $M = 20$ mesures différentes sur plusieurs années correspondant à P jours.

Nous utilisons tous les descripteurs disponibles dans l'ensemble de données [17] : chaque cible locale $\mathbf{y} \in \mathbb{R}^6$ (6 semaines suivant t dans le comté c) est associée à des descripteurs statiques \mathbf{s}_d et \mathbf{s}_n , ainsi qu'à une série temporelle multivariée $\mathbf{x} \in \mathbb{R}^{T \times M'}$. La période extraite de \mathbf{X} correspond à $T = 180$ jours avant le timestamp t pour le comté c , tandis que nous prenons les M mesures disponibles ainsi que les M mesures correspondant à l'année précédente sur les mêmes jours afin de permettre au modèle de construire des caractéristiques comparatives,

donc $M' = 2M$.

Les comtés sont indexés à l'aide de l'identifiant FIPS [20]. Nous utilisons les sous-ensembles d'entraînement, validation et test disponibles en Kaggle¹.

3 Le modèle hybride proposé

Malgré la flexibilité des bibliothèques pour la construction de réseaux neuronaux, le traitement des données avec des structures hétérogènes reste un problème ouvert dans l'apprentissage profond [10, 13]. Nous proposons un modèle neuronal hybride qui combine quatre modules : des LSTM, des FFNN, des couches d'embeddings et un mécanisme d'attention. La figure 1 présente le schéma de l'architecture proposée.

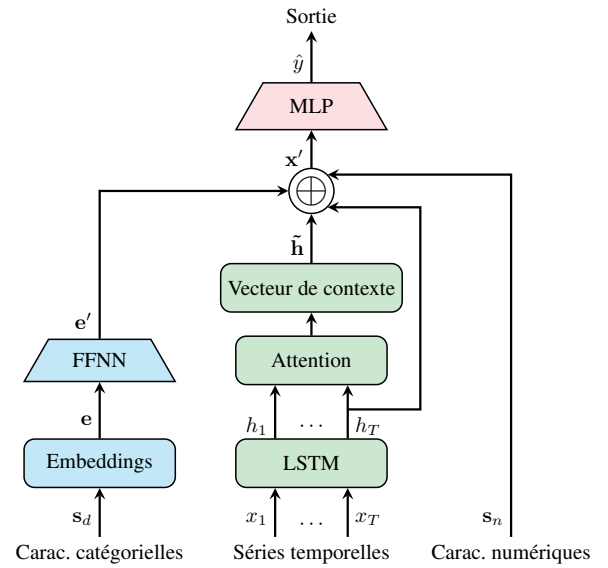


FIGURE 1 – Schéma de l'architecture proposée.

Pendant la propagation vers l'avant, les caractéristiques catégorielles \mathbf{s}_d passent par la couche d'embeddings E , qui les transforme en vecteurs denses $\mathbf{e} \in \mathbb{R}^z$. Ensuite, les représentations denses des caractéristiques catégorielles sont concaténées et passées à travers un réseau de neurones à action directe (FFNN) \mathcal{F} qui réduit leur dimensionnalité et permet d'obtenir un vecteur $\mathbf{e}' = \mathcal{F}(E(\mathbf{s}_d)) \in \mathbb{R}^{z'}$, où $z' < z$.

En parallèle, la série temporelle multivariée \mathbf{x} est introduite dans le LSTM, produisant des états cachés à chaque pas de temps t : $\{\mathbf{h}_t\}_{t=1:T} \in \mathbb{R}^{h \times T}$. Les états cachés sont ensuite traités par le mécanisme d'attention (détaillé dans l'Annexe A) pour produire un vecteur de contexte $\hat{\mathbf{h}} \in \mathbb{R}^h$. Les caractéristiques numériques $\mathbf{s}_n \in \mathbb{R}^{f_n}$ restent inchangées.

La représentation finale est obtenue en concaténant le vecteur de contexte $\hat{\mathbf{h}}$, le dernier état caché \mathbf{h}_T , les caractéristiques continues \mathbf{s}_n et la représentation latente des caractéristiques catégorielles \mathbf{e}' .

1. <https://www.kaggle.com/datasets/cdminix/us-drought-meteorological-data>

TABLE 1 – Résultats hebdomadaires sur l’ensemble de test pour le modèle hybride proposé (HM) ainsi que pour le modèle LSTM de référence [17].

	Semaine 1		Semaine 2		Semaine 3		Semaine 4		Semaine 5		Semaine 6	
Modèle	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁
LSTM	0.150	81.6	0.229	71.6	0.286	64.5	0.347	57.4	0.394	54.2	0.432	49.6
HM	0.126	82.2	0.169	74.7	0.209	68.6	0.244	64.0	0.269	58.6	0.294	51.0

TABLE 2 – Résultats de l’étude d’ablation pour le modèle hybride proposé (HM).

Composants présents					
Caractéristiques statiques	Series temporelles	Mécanisme d’attention	MAE	RMSE	F1
✓	✓	✓	0.217	0.377	66.3
	✓	✓	0.267	0.419	56.2
	✓		0.271	0.420	56.6
✓	✓		0.280	0.427	57.1
✓			0.755	0.920	21.2

Le vecteur résultant est $\mathbf{x}' = [\tilde{\mathbf{h}}, \mathbf{h}_T, \mathbf{e}', \mathbf{s}_n] \in \mathbb{R}^{2h+z'+f_n}$. Ce vecteur concaténé passe ensuite par un perceptron multicouche (MLP) \mathcal{M} , qui produit la prédiction $\hat{y} = \mathcal{M}(\mathbf{x}') \in \mathbb{R}^6$.

A notre connaissance, aucune recherche préalable ne s’est appuyée sur une architecture neuronale combinant ces éléments pour prédire les catégories de sécheresse de l’USDM.

4 Expériences et résultats

Étant donné le cadre décrit, nous avons mené des expériences pour étudier les questions suivantes. a) Comment le modèle hybride proposé se compare-t-il au meilleur modèle de référence établi [17] en termes de performance prédictive et de capacité de généralisation? b) Quelle est l’influence relative de chaque composant architectural sur la performance globale du modèle? c) Le modèle proposé maintient-il des performances supérieures dans le cadre d’un entraînement indépendant à la localisation par rapport à un entraînement spécifique à la localisation? d) De quelle manière le mécanisme d’attention et les embeddings contribuent-ils à la construction des représentations apprises par le modèle?

L’optimisation des hyperparamètres a été réalisée à l’aide de l’optimisation bayésienne via Optuna [2], en utilisant l’optimiseur AdamW [15] et un taux d’apprentissage cyclique [23]. Nous avons répliqué le modèle LSTM de référence en suivant le code source et les directives publiquement disponibles [17]. Nous présentons les résultats issus de notre implémentation afin d’assurer la cohérence, d’autant plus que nos résultats s’alignent étroitement sur ceux précédemment rapportés. Le modèle Transformeur de référence a été exclu de l’analyse en raison des performances supérieures du LSTM, rendant son inclusion non pertinente pour cette étude. Cette décision est en accord avec les conclusions des auteurs de DroughtED [17] et est

également soutenue par des résultats précédents [29], qui mettent en évidence les limites des approches basées sur les Transformeurs dans la prévision de séries temporelles.

4.1 Performance prédictive et généralisation

Après avoir entraîné l’architecture avec l’ensemble optimal d’hyperparamètres, nous avons observé que le modèle hybride proposé (HM) présente de meilleures performances en termes de MAE et de score macro F_1 au fil des semaines, comme l’indique la table 1. Sur l’ensemble de test, le modèle montre des améliorations relatives de 30 % pour le MAE, 9% pour le F_1 et 7% pour le score ROC-AUC multi-classes par rapport au LSTM de référence (Appendice C). Afin d’estimer plus précisément le risque espéré, nous avons réalisé une validation croisée à 5 blocs (Appendice D) et effectué un test t apparié pour évaluer si le HM surpasse significativement le LSTM. Les résultats montrent une amélioration significative du MAE, du RMSE et du F_1 , avec des valeurs- p de 0.03, 0.04 et 0.02, respectivement.

4.2 Étude d’ablation

Afin d’évaluer la contribution individuelle des différents composants du modèle dans le contexte de la prévision de sécheresse, nous avons mené une étude d’ablation en analysant l’impact de trois variables clés (Table 2).

Les variables statiques, décrivant les caractéristiques physiques du sol, encapsulent des informations contextuelles sur la capacité de résilience face au stress hydrique. Lorsqu’elles sont utilisées seules, ces variables conduisent à des performances limitées. Bien que cruciales pour situer le contexte de la prévision, les variables statiques ne suffisent pas à capturer la dynamique inhérente à l’apparition des sécheresses.

Les séries temporelles, quant à elles, jouent un rôle central en reflétant l’évolution des conditions météorologiques et hydrologiques. Leur utilisation exclusive améliore signifi-

cativement les prévisions. Ce comportement est conforme aux résultats précédents [17, 8].

Les meilleures performances ont été obtenues lorsque le mécanisme d'attention a été intégré. Nous attribuons ce comportement au fait que ce mécanisme permet au modèle de focaliser son analyse sur les périodes clés de la fenêtre temporelle, en attribuant une pondération plus élevée aux observations les plus informatives.

TABLE 3 – Résultats de la première semaine de prévision pour le modèle HM, comparant l'entraînement sur les données de chaque comté à celui sur l'ensemble des comtés. N désigne le nombre d'observation dans les ensembles.

Entraînement (N)	Évaluation	Semaine 1	
		MAE	F_1
Iowa (196)	Iowa	0.101	86.7
Montana (148)	Montana	0.203	52.3
Oklahoma (26)	Oklahoma	0.156	75.8
Tout (97099)	Iowa	0.086	89.3
	Montana	0.144	59.5
	Oklahoma	0.096	83.1

4.3 Entraînement indépendant vs spécifique à la localisation

Comme dans l'expérience menée dans l'évaluation initiale de l'ensemble de données pour le modèle LSTM de référence [17], nous avons repris les trois comtés obtenus de manière aléatoire (Iowa, Montana et Oklahoma) et avons entraîné le modèle hybride (HM) sur chaque comté individuellement ainsi que sur l'ensemble des données d'entraînement (Appendice F). Le modèle entraîné sur les données de tous les comtés —indépendant à la localisation— a montré une amélioration relative moyenne de 9,3%. Cela indique que, pour le modèle HM, l'entraînement indépendant à la localisation surpasse l'entraînement spécifique à la localisation. En comparaison, le modèle LSTM de référence a montré une amélioration relative moyenne de 4,6% [17].

4.4 Introspection du modèle

Nous menons une analyse qualitative des représentations intermédiaires du modèle. Pour ce faire, nous appliquons la technique de réduction de dimension t-SNE (avec une perplexité de 100 et 1000 itérations) [25] et étudions la manière dont les observations se regroupent en fonction des différentes variables catégorielles. Nos résultats révèlent que les embeddings se structurent de manière cohérente avec les catégories de la variable «Disponibilité des nutriments» (voir Figure 2). Nous estimons ce résultat favorable, car il appuie les conclusions de travaux antérieurs ayant mis en évidence l'impact significatif des sécheresses sur la présence et l'accessibilité des nutriments dans le sol [11, 4]. Nos observations suggèrent que le modèle capture

efficacement les relations sémantiques sous-jacentes entre les catégories des variables catégorielles, renforçant ainsi la pertinence de ses prédictions.

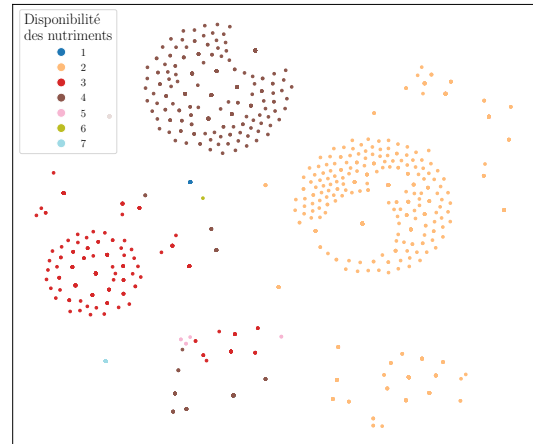


FIGURE 2 – Visualisation du t-SNE colorée par «Disponibilité des nutriments».

En ce qui concerne les poids d'attention, nous avons fait passer l'ensemble de test à travers le modèle, puis nous avons tracé le poids d'attention moyen pour chaque jour et son intervalle de confiance à 95% (Figure 3). L'attention se concentre principalement sur les 10 premiers jours, tout en accordant une importance supplémentaire aux 30 derniers jours de la fenêtre d'observation.

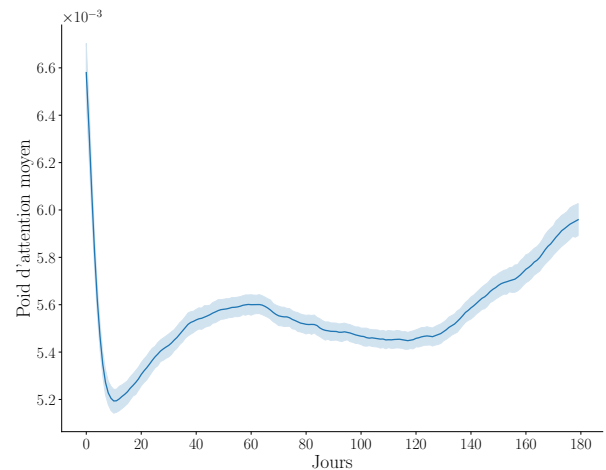


FIGURE 3 – Poids moyens d'attention sur l'ensemble de test.

5 Conclusions et perspectives

Dans cet article nous présentons une architecture neuronale hybride intégrant des séries temporelles et des données

statiques. Cette approche permet d'exploiter efficacement des données hétérogènes pour la prévision des conditions de sécheresse. Nous validons son efficacité par une évaluation empirique sur l'ensemble de données DroughtED, en obtenant des performances de l'état de l'art dans la prévision des catégories de sécheresse USDM.

Nos résultats ont montré que l'intégration des descripteurs statiques, des séries temporelles multivariées et du mécanisme d'attention améliore la précision des prédictions et la capacité de généralisation du modèle par rapport aux performances des modèles de référence [17]. L'étude d'ablation a confirmé l'importance de chaque composant architectural, en particulier des séries temporelles et du mécanisme d'attention, qui permettent au modèle de capturer les dynamiques complexes produisant des événements de sécheresse.

Les améliorations futures de l'architecture proposée viseront en priorité à affiner le mécanisme d'attention par le biais de deux stratégies : (1) calibrer l'acuité des poids d'attention via un paramètre de température appris dans la fonction softmax et (2) tirer parti des annotations d'experts pour superviser l'apprentissage de l'attention à l'aide d'une fonction de perte auxiliaire.

Au-delà des améliorations de performance obtenues pour la tâche abordée dans cet article, la conception modulaire de notre architecture est dotée d'une structure modulaire. Ce cadre flexible peut être étendu à la prévision d'autres événements météorologiques extrêmes et à d'autres tâches nécessitant l'intégration de données statiques et temporelles.

Remerciements

Nous tenons à remercier Agrial pour avoir mis à notre disposition les installations et le matériel informatique nécessaires à la conduite de cette recherche.

Références

- [1] Pouya Aghelpour, Hadigheh Bahrami-Pichaghchi, and Ozgur Kisi. Comparison of three different bio-inspired algorithms to improve ability of neuro fuzzy approach in prediction of agricultural drought, based on three different indexes. *Computers and Electronics in Agriculture*, 170 :105279, 2020.
- [2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna : A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- [3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.
- [4] Deepesh R. Bista, Scott A. Heckathorn, Dileepa M. Jayawardena, Sasmita Mishra, and Jennifer K. Boldt. Effects of drought on nutrient uptake and the levels of nutrient-uptake proteins in roots of drought-sensitive and -tolerant grasses. *Plants*, 7(2), 2018.
- [5] Shilpa Chaudhari, Vandana Sardar, and Prosenjit Ghosh. Drought classification and prediction with satellite image-based indices using variants of deep learning models. *International Journal of Information Technology*, 15(7) :3463–3472, Oct 2023.
- [6] Abhirup Dikshit, Biswajeet Pradhan, and M. Santosh. Artificial neural networks in drought prediction in the 21st century—a scientometric analysis. *Applied Soft Computing*, 114 :108080, 2022.
- [7] John A. Dracup, Kil Seong Lee, and Edwin G. Paulson Jr. On the definition of droughts. *Water Resources Research*, 16(2) :297–302, 1980.
- [8] Paulina Dukat, Ewa Bednorz, Klaudia Ziemblińska, and Marek Urbaniak. Trends in drought occurrence and severity at mid-latitude european stations (1951–2015) estimated using standardized precipitation (spi) and precipitation and evapotranspiration (spei) indices. *Meteorology and Atmospheric Physics*, 134(1) :20, Jan 2022.
- [9] Nikhil Gaikwad, Gaurav Chavan, Hemant Palivela, and Preeja Ravishankar Babu. Study of various feature extraction and selection techniques for drought prediction in precision agriculture. *Advances in Intelligent Systems and Computing*, 340 :611–619, 01 2015.
- [10] Wenzhong Guo, Jianwen Wang, and Shiping Wang. Deep multimodal representation learning : A survey. *IEEE Access*, 7 :63373–63394, 2019.
- [11] Mingzhu He and Feike A. Dijkstra. Drought effect on plant nitrogen and phosphorus : a meta-analysis. *New Phytologist*, 204(4) :924–931, 2014.
- [12] W. J. Humphreys. How droughts occur. *Bulletin of the American Meteorological Society*, 12(1) :18–22, 1931.
- [13] Simon Kamm, Sushma Sri Veekati, Timo Müller, Nasser Jazdi, and Michael Weyrich. A survey on machine learning based analysis of heterogeneous data in industrial automation. *Computers in Industry*, 149 :103930, 2023.
- [14] Christossy Lalika, Aziz Ul Haq Mujahid, Mturi James, and Makarius C.S. Lalika. Machine learning algorithms for the prediction of drought conditions in the wami river sub-catchment, tanzania. *Journal of Hydrology : Regional Studies*, 53 :101794, 2024.
- [15] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.
- [16] T. B. Mckee, Nolan J. Doesken, and John R. Kleist. The relationship of drought frequency and duration to time scales. In *8th Conference on Applied Climatology*, 1993.
- [17] Christoph D Minixhofer, Mark Swan, Calum McMee-kin, and Pavlos Andreadis. Droughted : A dataset

- and methodology for drought forecasting spanning multiple climate zones. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.
- [18] F.O. Nachtergaele, H. van Velthuizen, L. Verelst, N.H. Batjes, J.A. Dijkshoorn, V.W.P. van Engelen, G. Fischer, A. Jones, L. Montanarella, M. Petri, S. Prieler, E. Teixeira, D. Wilberg, and X. Shi. *Harmonized World Soil Database (version 1.0)*. FAO, Italy, 2008. Food and Agric Organization of the UN (FAO); International Inst. for Applied Systems Analysis (IIASA); ISRIC - World Soil Information; Inst of Soil Science-Chinese Acad of Sciences (ISS-CAS); EC-Joint Research Centre (JRC).
- [19] Neeta Nandgude, T. P. Singh, Sachin Nandgude, and Mukesh Tiwari. Drought prediction : A comprehensive review of different drought prediction models and adopted technologies. *Sustainability*, 15(15), 2023.
- [20] National Institute of Standards and Technology NIST. Federal information processing standards publication : counties and equivalent entities of the united states, its possession, and associated areas. *reports of the National Technical Information Service*, 1990.
- [21] Wayne C. Palmer. Keeping track of crop moisture conditions, nationwide : The new crop moisture index. *Weatherwise*, 21(4) :156–161, 1968.
- [22] A. Reichhuber, M. Svoboda, C. King-Okumu, A. Mirzabaev, S.M. Vicente-Serrano, R. Srinivasan, K. Ehler, X. Jia, A. Karnib, R. Lal, B. Mislimshoeva, N.H. Ravindranath, A. López Santos, L. Schipper, R. Stefanski, A. Vuković, and H. Zhang. Multiscale approaches for the assessment and monitoring of social and ecological resilience to drought. a report of the science-policy interface. Technical report, United Nations Convention to Combat Desertification, Bonn, Germany., 2023.
- [23] Leslie N. Smith. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 464–472, 2017.
- [24] Mark Svoboda, Doug LeComte, Mike Hayes, Richard Heim, Karin Gleason, Jim Angel, Brad Rippey, Rich Tinker, Mike Palecki, David Stooksbury, David Miskus, and Scott Stephens. The drought monitor. *Bulletin of the American Meteorological Society*, 83(8) :1181 – 1190, 2002.
- [25] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86) :2579–2605, 2008.
- [26] Sergio M. Vicente-Serrano, Santiago Beguería, and Juan I. López-Moreno. A multiscalar drought index sensitive to global warming : The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7) :1696 – 1718, 2010.
- [27] S Vijaya Shetty, M S Bharath, R Chandan, V Divya, and S Sushma Nayak. Prediction of drought - a machine learning approach using time series data. In *2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)*, pages 1–6, 2023.
- [28] Donald A. Wilhite, Mark D. Svoboda, and Michael J. Hayes. Understanding the complex impacts of drought : A key to enhancing drought mitigation and preparedness. *Water Resources Management*, 21(5) :763–774, May 2007.
- [29] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, AAAI’23/IAAI’23/EAAI’23*. AAAI Press, 2023.
- [30] Taiping Zhang, William S. Chandler, James M. Hoell, David Westberg, Charles H. Whitlock, and Paul W. Stackhouse. A global perspective on renewable energy resources : Nasa’s prediction of worldwide energy resources (power) project. In D. Yogi Goswami and Yuwen Zhao, editors, *Proceedings of ISES World Congress 2007 (Vol. I – Vol. V)*, pages 2636–2640, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [31] Alice Zheng and Amanda Casari. *Feature Engineering for Machine Learning : Principles and techniques for Data scientists*. O’Reilly Media, 2018.

A Détail sur le mécanisme d’attention

Le modèle proposé étant conçu pour être un cadre flexible, une implémentation simple du mécanisme d’attention est utilisée. Cette implémentation ressemble à l’attention de Bahdanau [3]. Alors que leur mise en œuvre vise à calculer les scores d’alignement entre les séquences d’un encodeur-décodeur, l’approche proposée se concentre sur la pondération différentielle des états cachés de la LSTM.

Etant donné l’ensemble des états cachés $H = [h_1, h_2, \dots, h_T]$ à la sortie de la LSTM, nous calculons les scores à l’aide d’une couche linéaire :

$$s_t = Wh_t + b \quad (1)$$

où W est le vecteur des poids et b le biais.

Nous calculons ensuite les poids d’attention en faisant passer les scores par une fonction softmax :

$$\alpha_t = \text{softmax}(s_t) = \frac{e^{s_t}}{\sum_{i=1}^T e^{s_i}} \quad (2)$$

Le vecteur de contexte $\tilde{\mathbf{h}}$ est calculé comme la somme pondérée des états cachés du LSTM, en utilisant les poids d’attention.

$$\tilde{\mathbf{h}} = \sum_{t=1}^T \alpha_t h_t \quad (3)$$

B Hyperparamètres sélectionnés pour chaque modèle

TABLE 4 – Hyperparamètres pour les modèles de référence [17] et pour le modèle hybride proposé (HM). Le cas échéant, la notation reste cohérente avec celle utilisée dans le corps de l’article.

Hyperparamètre	Notation	LSTM	Transformer	Modèle Hybride (HM)
LSTM ou Transformer Nombre de couches		2	4	2
LSTM Taille de l’état cachée	h	512	512	490
Taille initiale de l’embedding	z	-	256	27
Taille réduite de l’embedding (après FFNN)	z'	-	-	6
Nombre de couches du MLP final		-	-	2
Taille de la couche cachée du FFNN		-	4096	-
Nombre de têtes d’attention		-	2	-
Taille des lots		128	128	128
Probabilité de dropout		0.1	0.1	0.1
Probabilité de dropout des embeddings		-	-	0.4
Weight Decay (Pénalisation des poids)		0.01	0.01	0.01
Taux d’apprentissage		7e-5	7e-5	7e-5
Nombre d’époques		7	7	9

C Résultats sur l’ensemble de test

TABLE 5 – Résultats sur l’ensemble de test pour la LSTM de référence et le modèle proposé.

Modèle	MAE	RMSE	F ₁	ROC-AUC
LSTM	0.306	0.478	61.9	80.6
HM	0.218	0.378	67.3	85.9

D 5-Folds CV

TABLE 6 – Résultats détaillés de la validation croisée à 5-fold pour la LSTM de référence et le modèle proposé.

Fold	LSTM			HM		
	MAE	RMSE	F1	MAE	RMSE	F1
1	0.347	0.553	58.34	0.244	0.433	60.22
2	0.365	0.570	42.79	0.302	0.519	59.67
3	0.272	0.444	66.22	0.254	0.404	75.22
4	0.332	0.548	44.82	0.266	0.433	59.84
5	0.310	0.504	63.88	0.299	0.502	71.06

TABLE 7 – Moyenne et écart-type pour chaque métrique sur les résultats de la validation croisée à 5-fold.

Modèle	MAE ($\bar{x} \pm \sigma$)	RMSE ($\bar{x} \pm \sigma$)	F ₁ ($\bar{x} \pm \sigma$)
LSTM	0.325 \pm 0.036	0.524 \pm 0.051	55.2 \pm 0.108
HM	0.273 \pm 0.026	0.458 \pm 0.050	65.2 \pm 0.074

E Détail de l’étude d’ablation

TABLE 8 – Résultats hebdomadaires de l’étude d’ablation sur l’ensemble de test.

Modèle	Semaine 1		Semaine 2		Semaine 3		Semaine 4		Semaine 5		Semaine 6	
	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁
HM	0.126	82.2	0.169	74.7	0.209	68.6	0.244	64.0	0.269	58.6	0.294	51.0
TS+Att	0.134	65.9	0.189	62.3	0.250	56.3	0.307	51.2	0.360	51.6	0.361	50.8
TS	0.136	61.9	0.192	62.3	0.253	56.3	0.312	56.4	0.364	51.6	0.368	50.8
SF+TS	0.144	73.9	0.203	62.3	0.262	56.3	0.320	51.2	0.374	49.3	0.375	50.8
SF	0.779	20.4	0.746	22.7	0.752	25.3	0.713	18.9	0.754	19.5	0.787	17.0

F Entraînement indépendant vs spécifique à la localisation

TABLE 9 – Résultats hebdomadaires du HM sur les données d’entraînement locales vs nationales. Les comtés choisis pour l’expérience sont Iowa (IA), Montana (MT) et Oklahoma (OK).

Entraînement	Éval.	Semaine 1		Semaine 2		Semaine 3		Semaine 4		Semaine 5		Semaine 6	
		MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁	MAE	F ₁
IA	IA	0.101	86.7	0.179	67.7	0.214	69.5	0.287	63.5	0.298	60.9	0.272	59.4
MT	MT	0.203	52.3	0.314	49.1	0.339	50.9	0.341	38.6	0.377	37.0	0.407	35.9
OK	OK	0.156	75.8	0.230	59.7	0.269	56.9	0.327	61.6	0.352	59.1	0.387	57.8
Tout	IA	0.086	89.3	0.122	79.2	0.151	78.4	0.189	71.9	0.214	73.9	0.235	66.4
	MT	0.144	59.5	0.168	52.9	0.178	50.0	0.209	45.5	0.237	44.4	0.265	38.6
	OK	0.096	83.1	0.160	75.8	0.196	77.9	0.209	77.5	0.260	73.5	0.298	66.3

TABLE 10 – Résultats sur l’ensemble de test en utilisant les données d’entraînement locales vs nationales.

Entraînement	Évaluation	MAE	RMSE	F ₁
IA	IA	0.201	0.383	73.8
MT	MT	0.301	0.354	46.7
OK	OK	0.278	0.402	63.1
Tout	IA	0.166	0.315	76.6
	MT	0.200	0.320	48.2
	OK	0.218	0.378	67.3