Geosciences

AIMS Geosciences, 10(3): 524–539.
DOI: 10.3934/geosci.2024027
Received: 28 April 2024
Revised: 02 July 2024
Accepted: 15 July 2024
Published: 19 July 2024

http://www.aimspress.com/journal/geosciences

Letter

Fuzzy time series clustering using autoencoders neural network

Chibuike Chiedozie Ibebuchi[1,2,*]

[1] Department of Geography, Kent State University, Kent, Ohio, USA
[2] ClimRISE Lab, Kent State University, Kent, Ohio, USA

* **Correspondence:** Email: cibebuch@kent.edu.

Abstract: This study presents a novel approach that employs autoencoders (AE)—an artificial neural network—for the nonlinear transformation of time series to a compact latent space for efficient fuzzy clustering. The method was tested on atmospheric sea level pressure (SLP) data towards fuzzy clustering of atmospheric circulation types (CTs). CTs are a group of dates with a similar recurrent SLP spatial pattern. The analysis aimed to explore the effectiveness of AE in producing and improving the characterization of known CTs (i.e., recurrent SLP patterns) derived from traditional linear models like principal component analysis (PCA). After applying both PCA and AE for the linear and nonlinear transformation of the SLP time series, respectively, followed by a fuzzy clustering of the daily SLP time series from each technique, the resulting CTs generated by each method were compared to assess consistency. The findings reveal consistency between the SLP spatial patterns from the two methods, with 58% of the patterns showing congruence matches greater than 0.94. However, when examining the correctly classified dates (i.e., the true positives) using a threshold of 0.8 for the congruence coefficient between the spatial composite map representing the CT and the dates grouped under the CT, AE outperformed PCA with an average improvement of 29.2%. Hence, given AE's flexibility and capacity to model complex nonlinear relationships, this study suggests that AE could be a potent tool for enhancing fuzzy time series clustering, given its capability to facilitate the correct identification of dates when a given CT occurred and assigning the dates to the associated CT.

Keywords: fuzzy classification; autoencoders; principal component analysis; sea level pressure; artificial neural network

1. Introduction

The improved classification of atmospheric circulation types (CTs) is among the core interests of climate research. Understanding CTs provides insight into regional climate phenomena, as well as broader atmospheric dynamics, which are critical for weather forecasting, climatic modeling, and environmental risk assessments [1]. However, the complex, nonlinear nature of atmospheric systems [2,3] poses significant challenges to accurate CT classification. A variety of CT classification methods have been adopted over time, aiming to decipher complex atmospheric systems [4–7]. These techniques include cluster analysis [8] and rotated principal component analysis (PCA) applied to temporally decomposed climate data [6], among others. While each of these techniques has its strengths and limitations, cluster analysis and PCA, for example, are more adept at handling large datasets and reducing dimensionality. Still, they may oversimplify complex and nonlinear relationships [9]. Therefore, the performance of these traditional techniques in classifying physically consistent CTs might be constrained by the nonlinear and complex nature of atmospheric systems [10].

Traditionally, rotated T-mode (the variable is the time series, and the observations are grid boxes or locations where the climate field was measured) PCA has been extensively employed for clustering time series climate data for subsequent classification of CTs [6,11,12]. T-mode PCA functions by reducing the dimensionality of the time series, focusing on temporal patterns that explain the most variance in the climate data [6]. The subsequent rotation of the PCs to a simple solution enhances the identification time series that covary in terms of having a similar spatial pattern over time. This is achieved by assigning larger loading magnitudes to such time steps that covary, enabling the classification of the time steps under a given retained PC (principal component). However, being a linear method, rotated PCA might be inadequate to model complex, nonlinear relationships in the data. The inadequacies of PCA to fully capture nonlinear interactions have led to the exploration of alternative methods. Fuzzy rotated T-mode PCA has been proposed as an alternative method, aiming to provide a more holistic approach to data variability by incorporating degrees of membership and some level of nonlinearity to CTs rather than definitive assignments [6]. However, while fuzzy rotated T-mode PCA is more flexible than traditional T-mode PCA, it might still inherit some limitations from its linear nature. Therefore, a more comprehensive method capable of capturing nonlinear relationships is needed. One such promising technique is the application of artificial neural networks (ANNs).

Several studies have successfully applied ANN, such as convolutional neural networks and self-organizing maps to classify atmospheric circulation patterns [13–16]. Autoencoders (AE), on the other hand, are a specific type of ANN designed to learn efficient data coding in an unsupervised manner [17]. While AE has been applied in climate science for several other purposes such as detecting anomalous climate events [18,19] and dimensionality reduction [20], among others [21], its application in time series clustering (of climate data) has not gained wide applications. Even so, AE has been applied to examine synoptic behaviors [22], and convolutional AE has been used in clustering the states of the polar vortex [23]; in other fields, it has been applied for clustering time series [24–28].

Since rotated T-mode PCA is one of the most extensively used and established traditional techniques for time series clustering (or decomposition) of climate data leading to the classification of CTs [6], this study will examine the consistency of CTs from AE and from rotated T-mode PCA. AE applied to cluster time series data can be expected to offer a more flexible and robust framework that

can model complex nonlinear relationships, potentially overcoming the limitations posed by rotated T-mode PCA. The nonlinear dimensionality reduction capability of AE could prove particularly useful for CT classification in regions with complex climates like the southern region of Africa. It is also noteworthy that self-organizing maps excel at topological representation of the data and clear visualization of clusters in a 2D space. However, considering that the primary goal of this study is to capture essential linear and nonlinear temporal patterns in the data and then use the reduced-dimensional representation for clustering, the AE is considered more suitable. The latent space of the AE can serve a similar role to the PC loadings from rotated T-mode PCA but with the added ability to capture nonlinearities.

Given that rotated T-mode PCA has been applied to classify physically interpretable CTs in the southern region of Africa [6,29,30], this study uses the already classified CTs in [6] as a reference to analyze the effectiveness of AE for the time series clustering of sea level pressure (SLP) data to classify CTs in the southern region of Africa. This endeavor is poised to contribute significantly to the current state of knowledge in data science by providing an innovative approach to time series clustering.

2. Materials and methods

Daily SLP data from 1950 to 2020 was obtained from ERA5 reanalysis [31]. The horizontal resolution of the data is 0.25° longitude and latitude. The spatial extent for clustering the SLP time series leading to classifying the CTs in the southern region of Africa is 5.25°–55.25°E and 0°–50.25°S, which is based on matching degrees of freedom and prior knowledge of the spatial extent of circulation processes in the region [6].

AE, a type of artificial neural network, operates by compressing input data into a latent space representation and then reconstructing the original data from this representation [17]. This process is unsupervised and aims to learn efficient data representations through an encoder-decoder architecture. The encoder-decoder architecture consists of two main components: the encoder, which compresses the input data into a latent representation, and the decoder, which reconstructs the data from this representation. This framework is central to AE, enabling it to capture meaningful features and patterns in the input data while minimizing reconstruction error. In this study, AE was applied to encode the time series of gridded SLP data. The encoded temporal patterns were further classified to identify dates with coherent features, where "coherent" denotes dates with the alignment and consistency of temporal features (i.e., similar SLP spatial patterns) across the dataset, critical for precise classification.

The logic for the time series clustering follows the same as the rotated T-mode PCA, where the output is PC loadings (time series) with each day in the analysis period having a weight (or loading magnitude), which indicates the amplitude (or relative contribution) of the circulation pattern detected by the rotated PC on that day [6]. The AE in the same paradigm is applied to achieve a compact representation of the temporal SLP patterns in the region, synonymous with the retained PCs, where each temporal output pattern in the latent space in this study is referred to as a *Node*. Each day under a Node will also have a weight that designates the amplitude (or relative contribution) of its circulation pattern to the overall atmospheric circulation of the day in question. The Nodes similar to the retained PC represent a class, and the daily SLP patterns under each Node (similar to the retained PC) are the potential variables to be classified under the class in question. The larger the weight of a day under a

given class, the higher the tendency to classify that day under the class; the lower the weight of a day under a given class, the higher the tendency to discard the day under that class using a fixed hyperplane width threshold. A hyperplane acts as a decision boundary for the encoded values (or rotated PCA loadings); days categorized under each Node are given weights that signify their relative contributions to the day's circulation pattern. The hyperplane helps in distinguishing between significant and negligible contributions by setting a threshold (defined around a zero interval). Days with weights below this threshold are considered to have minimal impact and are hence discarded as noise. This method enhances the focus on more influential data points, improving the overall classification accuracy.

The methodological approach of rotated PCA for time series clustering is already well documented in [6], where singular value decomposition was applied to T-mode SLP data to derive PCs. To enhance the physical interpretability of the PCs, the PC loadings (temporal patterns) were rotated obliquely using the Promax routine [6]. Figure 1 highlights the approach introduced in this study in using AE to achieve the same purpose toward CT classification.

The target variable to reduce its dimensionality and extract the most crucial nonlinear temporal patterns is the daily time series of the SLP data from 1950 to 2020 comprising 25,933 variables. Hence, the aim is to reduce the 25,933 daily SLP time series (measured at 40,198 grid points) to a lower dimension K. k_i for example, represents one of the classes or CTs (or nodes) in the latent compact space and has a dimensionality of 1×25933. Each of the 25,933 days has a corresponding weight that is used to determine if the day should or should not be classified under k_i. In other words, the dates classified under k_i are synonymous with the dates when the recurrent atmospheric circulation pattern associated with k_i occurred in the historical data.

The SLP data was normalized to a [0,1] range using the MinMaxScaler function from the Scikit-learn library [32]. This step ensures that all data have the same scale and enhances convergence, which is essential for the training process of neural networks.

Next, the AE was defined using the Keras library in Python [33]. The encoder takes in high-dimensional input data and compresses it into a lower-dimensional space. The decoder then takes this compressed representation and attempts to reconstruct the original high-dimensional data. The number of input and output neurons equated to the dimension (i.e., 25,933 daily time series) of the SLP data. The hidden layer, a component of the encoder and decoder, and the associating optimal epoch number were selected iteratively, seeking to balance the representation of the complexity of the data and the risk of overfitting. To optimize the hyperparameters, the data was split into 70% for training, 10% for validating, and 20% for testing. Different neuron configurations were tested (i.e., 2, 4, 8, 16, 32, 64, 128, 256), with a maximum of 50 epochs and early stopping if the validation loss does not decrease after 5 consecutive epochs. Given that the minimum validation loss values after 64 neurons (and 15 epochs) are insignificant compared to using 128 neurons and above, 64 neurons were used, balancing the need for dimensionality reduction with the desire to retain key features in the data and avoiding overfitting with a more complex model. Therefore, the model underwent training multiple times to optimize performance, with the results presented derived from the simulation that exhibited the best performance. This approach ensures that the model does not merely fit the training data but generalizes well to new, unseen data.

The rectified linear unit (ReLU) activation function was utilized in the encoder. ReLU is widely used in neural network models due to its properties of mitigating the vanishing gradient problem and

introducing nonlinearity into the model [34]. The decoder employed the sigmoid activation function, which ensures that the output values are within the [0,1] range, aligning with the normalized data.

The AE was compiled using the Adam optimizer and mean squared error (MSE) as the loss function [35]. Adam is an optimization algorithm that adjusts the learning rate adaptively, which typically results in faster convergence. MSE quantifies the difference between the original and reconstructed data, driving the autoencoder to learn a compressed representation that maintains the key features of the input.

The trained autoencoder was subsequently used to encode the original SLP data into a lower-dimensional representation. This compressed representation captures the most important temporal patterns of the SLP data necessary for distinguishing different CTs. The encoded vectors are daily time series from 1950 to 2020.

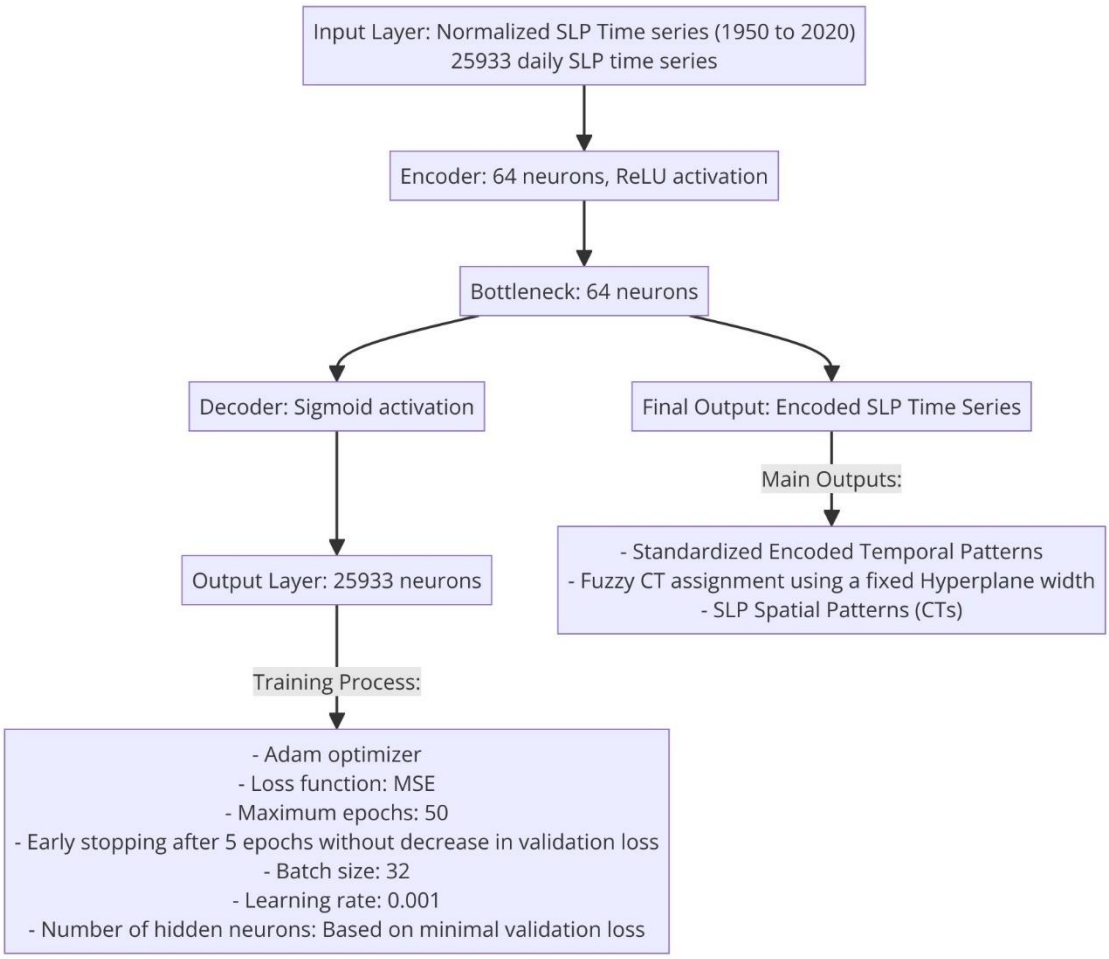

Figure 1. Flowchart illustrating the process of applying autoencoder neural networks for the classification of atmospheric circulation types.

To represent the asymmetry in climate patterns, the encoded vectors were z-score standardized and subsequently normalized between −1 and +1 to resemble rotated correlation-based PCA loadings. Days under a given Node with amplitude within the zero interval do not notably contribute to the circulation variability of the day. Hence, following [6], dates where the Node's value exceeded ±0.3

thresholds were selected and assigned to the positive phase and negative phase, defining two asymmetric CTs from a Node. This approach led to a subset of dates for each Node that were interpreted as periods when the circulation pattern presented by the Node was particularly active. The method of assigning the CTs also results in a fuzzy solution, since a day can be assigned to more than one Node or CT insofar as the CT was active on that day [6]. This fuzzy assignment method offers a more holistic and realistic portrayal of overlapping climate patterns, enhancing the model's applicability to real-world climatic phenomena.

The spatial patterns (i.e., composite maps) for the positive and negative phases of each CT derived from a given Node were computed using the weighted mean approach, which was documented to be optimal in representing fuzzy CTs [6]. This implied that for a given Node, all dates with values greater than 0.3 for the positive phase or lower than −0.3 for the negative phase were retained, the values of the other dates within the hyperplane were set to 0, and the non-zero values served as the weight when the encoded time series was projected onto the original gridded SLP data to derive spatial patterns (Equation 1).

$$\text{Weighted mean} = \left(\sum_{i=1}^{n} Z_i \, b_{iT}\right) / \sum_{i=1}^{n} b_{iT} \tag{1}$$

b_{iT} = Standardized/normalized encoded values. Encoded values with vector elements greater than |T| are retained and the vector elements with magnitudes lower than |T| are set to zero.
n = number of observations in the SLP time series, Z_i is the SLP data matrix.

$$weight = \begin{cases} 0 \text{ if } |b| < |\mathrm{T}| \\ b \text{ if } |b| \geq |\mathrm{T}| \end{cases}$$

The patterns derived from AE were matched to the patterns derived from the fuzzy correlation-based rotated T-mode PCA classified in [6] using a congruence coefficient (Equation 2) to document the distance between the vectors [36].

Finally, to assess the best-performed method in time series clustering of the SLP data, the SLP field of the dates/days assigned to the CTs derived from each method was matched to the corresponding composite map representing the CT. A congruence match of at least 0.8 was used to subjectively define "true positives" or the number of dates correctly classified under a given CT. This analysis is instrumental in comparing PCA and AE based on the percentage of true positives.

$$g(X,Y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \, \sum y_i^2}} \tag{2}$$

where X has elements x_i and Y has elements y_i. $g(X,Y)$ is the congruence coefficient between vectors X and Y.

3. Results

The architecture of the compiled AE resulted in six non-zero vectors or latent representations, which will be referred to as Nodes, for brevity. These Nodes represent the most salient features or temporal patterns in the SLP data. This makes the subsequent process of classifying CTs less

computationally intensive and possibly more accurate, as it is based on the most significant temporal patterns identified by the AE. Also, the six Nodes implied 12 CTs, as two asymmetric CTs (positive and negative phase) are derived from each Node.

The AE-based CTs were matched to the CTs derived from PCA using the congruence coefficient, supported by visual inspection of the SLP spatial patterns representing each CT. Figure 2 shows the SLP spatial patterns of the 12 CTs from AE and PCA. *N1p* is the positive phase of CT1 from autoencoders, while *N1n* is the negative phase of CT1 from autoencoders. Similarly, *P1p* is the corresponding positive phase of the CT1 classified from the rotated PCA and *P1n* is the negative phase of the CT classified from PCA. The 12 AE CTs were reproduced within 9 rotated PCs, i.e., among 9 PCs, 6 resembled the encoded patterns.

Interestingly, Figure 2 shows that the patterns of each CT and the asymmetric nature of the CTs from a single PC/Node are generally consistent. For example, for CT1, the positive phase is associated with a high-pressure anomaly on the south coast of southern Africa from both PCA and AE. Similarly, for the negative phase, from both AE and PCA, the southern hemisphere mid-latitude cyclone tracks northward. Very close similarities between the spatial pattern and asymmetry of the CTs are observable for most of the other CTs classified from both methods (Figure 2). The general consistency between the CTs classified independently from the two methods is quite promising—it shows that ANN-based CTs can produce patterns consistent with the traditional rotated T-mode PCA, widely applied in climate science for deriving (potentially) physically interpretable climate modes of variability. Moreover, as mentioned earlier, previous studies have documented the physical interpretability of the PCA-based CTs, used in this work in Figure 2 [29,30]. Hence, reproducing the CTs with AE is indicative that AE-based CTs can be interpreted physically.

Despite the close similarity between AE and PCA-based CTs, some differences in the corresponding spatial patterns are evident (Figure 2). For example, under the negative phase of CT2, the strength of the subtropical high-pressure system, south of Madagascar, is notably stronger under P2n than N2n (Figure 2).

Quantitatively from Table 1, the negative phase of CT1 has the highest congruence match of 0.99 between the AE and PCA-based CTs. The lowest congruence match is for the negative phase of CT6 (Table 1). The asymmetric patterns from CT6 and CT4 have relatively the lowest congruence matches, while CT1, CT3, and CT5 patterns have congruence matches greater than 0.94 (Table 1). The differences between the AE-based patterns and PCA-based patterns observed in Figure 2 could likely be due to their respective architectures and capabilities in modeling linear and nonlinear relationships. While PCA identifies a subset of dates that exhibit linear correlations, the AE, capable of capturing both linear and nonlinear correlations, locates a subset of dates accordingly. This distinction in handling linear and nonlinear relationships may account for the observed dissimilarities in the patterns produced by PCA and AE.

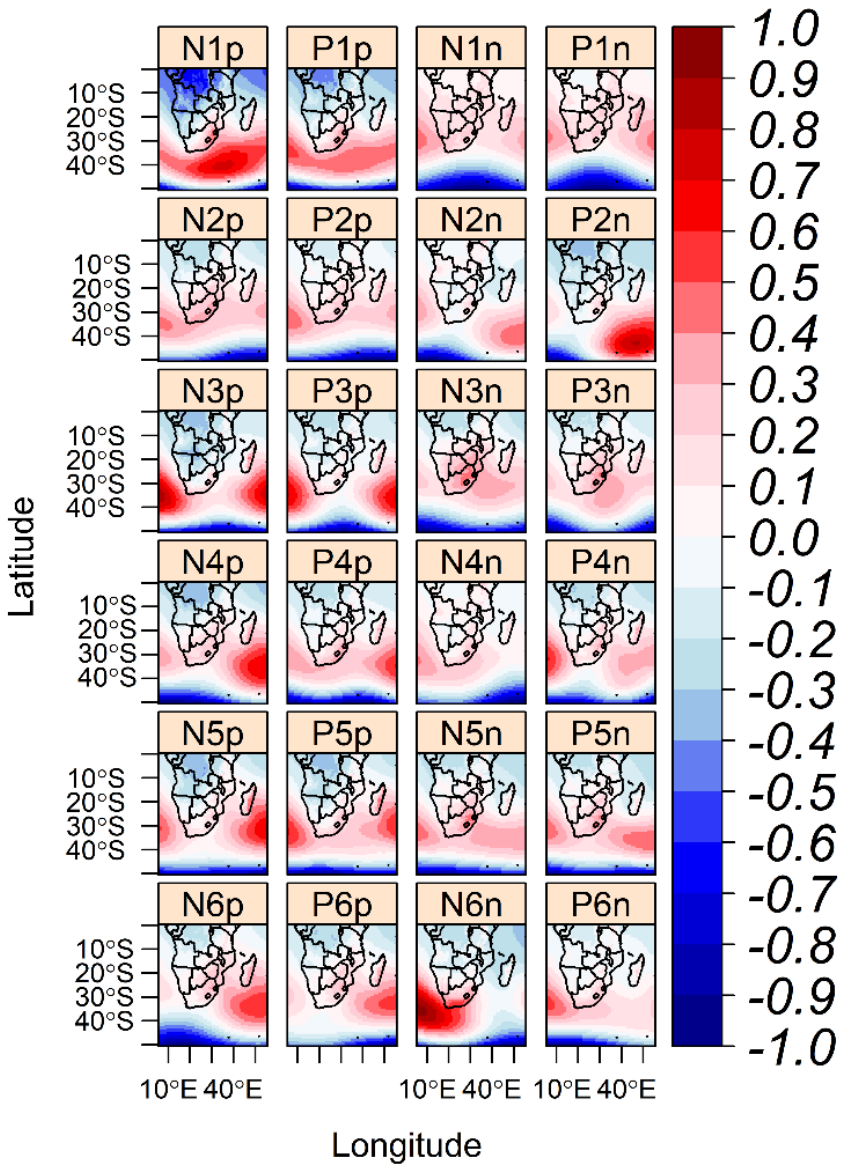

Figure 2. Circulation types from autoencoders (i.e., the "Ns") and the counterparts from rotated PCA (i.e., the "Ps"). The maps were calculated using the weighted mean approach. The weights are values above or below the ±0.3 hyperplane threshold. "p" indicates the positive phase and "n" indicates the negative phase of the asymmetric CTs derived from a given node or PC. Color is standardized SLP.

The time series of Node 1 and PC1 are shown in Figure 3. The inter-annual variability of the time series from the respective techniques is similar, such that for the individual months, the congruence matches are generally greater than 0.90. For other Nodes and corresponding PCs, the variability of the associating time series was close but not as close as that in Node 1 and PC1. The differences in the time series also imply the potential of assigning non-consistent dates to a respective CT, from either of the two methods. Therefore, the subtle differences in the spatial patterns in Figure 2 might be due to assigning some non-representative dates to a given CT or missing some representative dates in each CT.

Table 1. Congruence matches between the AE-based CTs and the corresponding PCA-based CTs in Figure 2. The value in the Table is the congruence coefficient.

Circulation type	Congruence coefficient
CT1 (positive)	0.94
CT1 (negative)	0.99
CT2 (positive)	0.95
CT2 (negative)	0.73
CT3 (positive)	0.95
CT3 (negative)	0.94
CT4 (positive)	0.88
CT4 (negative)	0.66
CT5 (positive)	0.97
CT5 (negative)	0.98
CT6 (positive)	0.81
CT6 (negative)	0.65

Though the PCA-based CTs in Figure 2 have been widely used in previous physical studies of the climate of southern Africa, the CTs might still have some limitations in their representation of the actual circulation variability due to the linear method used to create the CTs.

The performance of each method in classifying CTs was further assessed based on the ability to accurately assign dates to a given CT. To determine this, the standardized SLP fields of the respective days assigned to a specific CT were matched with the corresponding spatial map (as shown in Figure 2). This criterion was established to ensure that the identified CTs and any subsequent analysis derived from them are representative of the actual circulation variability and are not merely artifacts of the applied method. Hence, a congruence match between the SLP field of each day and the corresponding map (or CT) the day is assigned to, greater than or equal to 0.8, is used to define “true positives”. Indeed, it is reasonable to anticipate that multiple circulation patterns could co-exist at any given time. Thus, a single CT may not accurately encapsulate a substantial proportion of the variability present, especially if the CT represents rarer patterns with weaker signals [6]. Despite this, if a CT displays noteworthy characteristics on a particular day, it can still be considered representative and assigned to that day's CT. Also, at this point, it is vital to state that all days in the analysis period were classified under the CT(s) that occurred on the day. In Figure 4, when comparing the performance of the two different methods, the percentage of true positives from ANN-based CTs was used as the baseline. The percentage values represent improvements relative to each other. When the ANN-based CTs result in fewer true positives than the PCA-based CTs, the difference is expressed as a negative percentage change. When some values are negative, it suggests that for those specific CTs, the ANN-based method identified a lower number of dates with a congruence match greater than 0.8 compared with the PCA-based method.

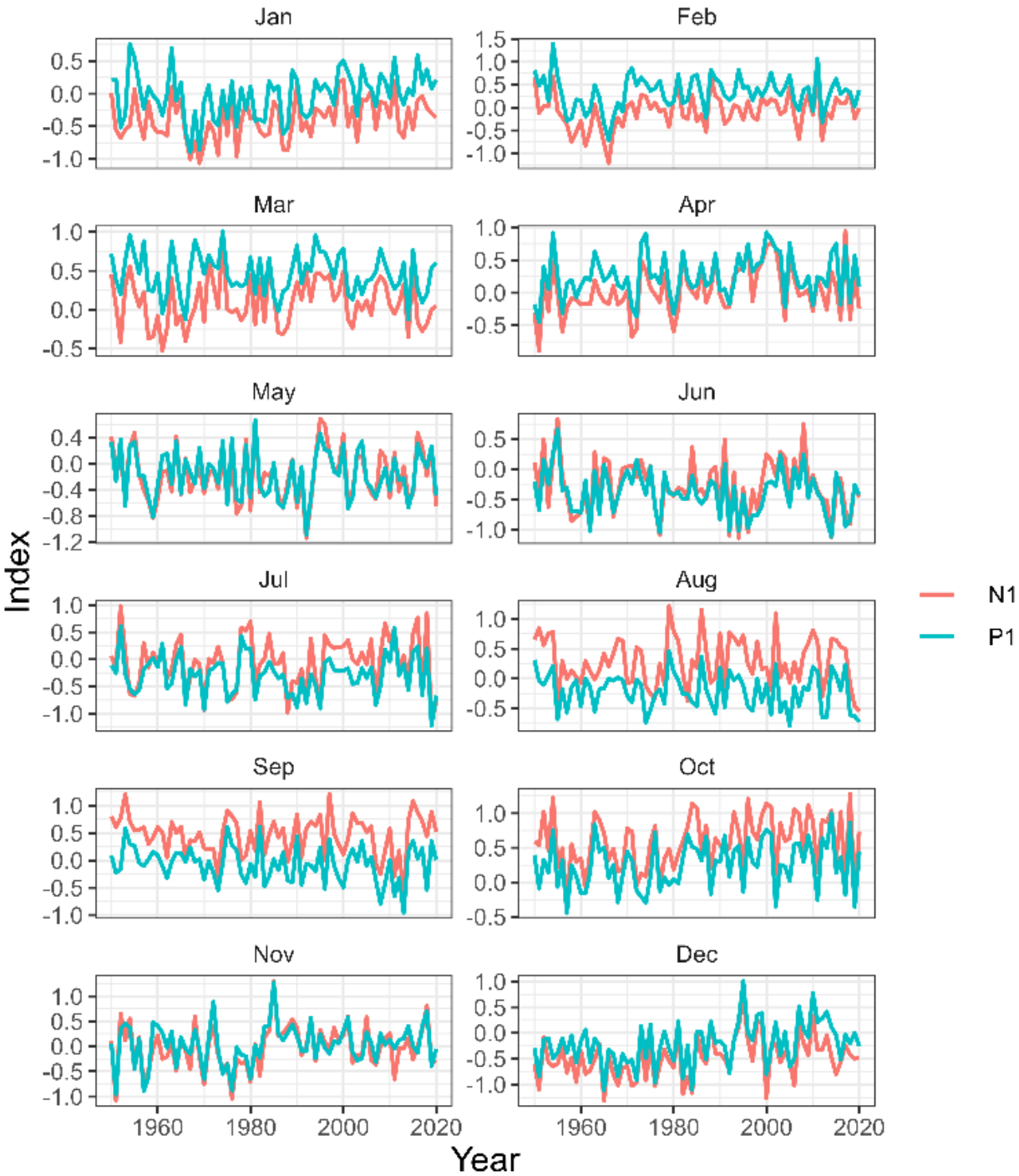

Figure 3 Time series of Node 1 and PC1 from 1950 to 2020. Time series was computed as the annual mean of the PC1 loadings/Node 1 values in each calendar month.

From Figure 4, the ANN-based method outperformed the PC-based method in correctly assigning dates to the positive phases of CT1, CT2, CT4, CT5, and CT6 and the negative phases of CT4 and CT5 (Figure 4). For the others, the PCA-based method performed better. Moreover, except for the negative phase of CT2, the AE patterns outperformed the PCA-based patterns by a large margin (Figure 4). The average percentage of correctly classified dates under each of the 12 CTs is 35.8% for

the ANN CTs and 27.7% for the PCA-based CTs, representing approximately a 29.2% improvement. Moreover, the ANN-based CTs outperformed the PCA-based CTs in terms of the percentage of dates not incorrectly classified under a given CT.

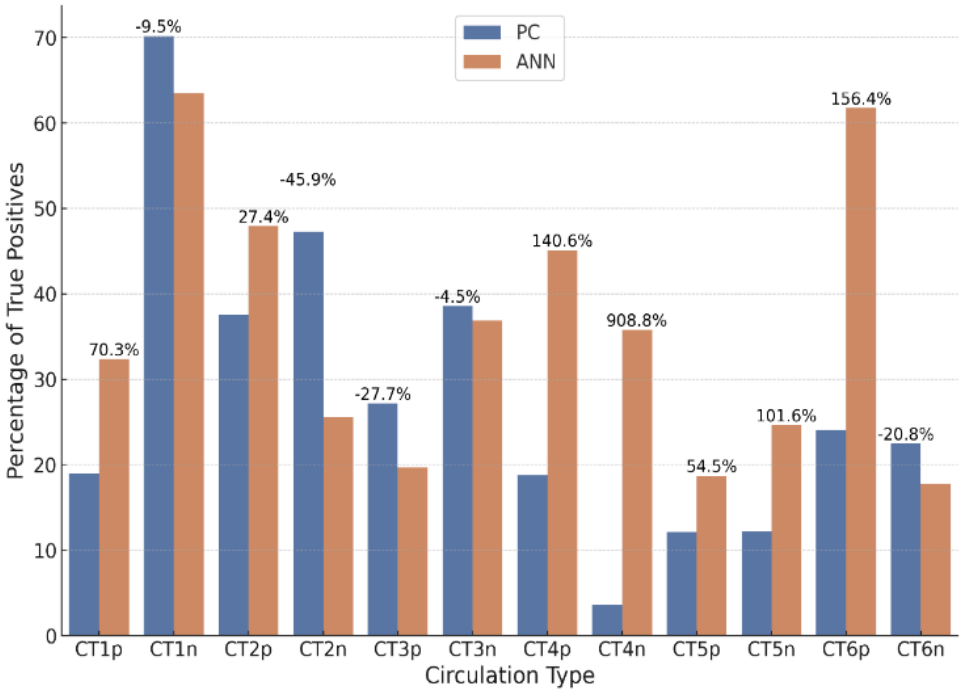

Figure 4 Comparison of the percentage of true positives for the ANN-based CTs and the PCA-based CTs. The y-axis represents the percentages of true positives. A "true positive" is defined as a date for which there is a congruence match greater than 0.8. These percentages are calculated based on the total number of days assigned to each CT. The values in the charts are percentage increases or decreases. Negative values imply a decrease.

4. Discussion and conclusions

While artificial neural networks (ANNs) offer numerous advantages in climate science research, a common challenge lies in their interpretability [37–39]. However, in recent periods, the physical interpretability of ANN outputs has increased [40]. For example, Pierdicca and Paolanti [41] highlighted the application of interpretable ANNs for the interpretation of geomatic data. Mamalakis et al. [42] reported that ANNs are applicable in model fine-tuning. Labe and Barnes [43] detected climate signals using explainable ANN.

In a field like climate science where physical interpretations are of paramount importance, leveraging ANN algorithms to yield interpretable results is a critical task. This study contributes to resolving this challenge by utilizing autoencoders (AE), a type of ANN, to temporally decompose climate data, resulting in physically consistent atmospheric circulation patterns. Given the prevalence of PCA in climate science for deriving potentially physically interpretable modes of climate variability [6], comparing patterns derived from PCA and ANN was deemed essential. The findings of this study show that the AE-derived CTs generally align consistently with the PCA-based CTs. This outcome

highlights the potential of AE as a valuable tool for climate data time decomposition and subsequent CT identification, supporting previous research affirming the ability of ANN, especially convolutional neural networks, to enhance physically consistent classifications of the atmospheric system [13–15].

A major limitation of this study is the possibility of a few emergent patterns, whereby the ANN-based patterns might not be sufficiently comparable to any PCA-based patterns. This might be the case of CT4 and CT6 with the lowest congruence matches (Figures 2 and Table 1).

Furthermore, the results show that the average percentage of correctly assigned days to given CTs was higher by 29.2% for ANN-based CTs than PCA-based ones. This finding suggests that ANNs, with their capacity to model complex linear and nonlinear systems, offer promise in enhancing the accurate classification of CTs representative of the actual circulation variability. Therefore, AE and other ANN techniques should be further explored and refined in future studies for broader applicability and enhanced interpretation in climate science.

While the dense layers of AE were applied in this study due to its ability to capture nonlinear relationships and encode crucial temporal patterns in time series data, providing a strong foundation for identifying coherent temporal features crucial for time series classification, other ANN models such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks offer distinct advantages [44–47]. CNNs extract spatial features and LSTMs can capture temporal dependencies. The potential of integrating CNNs and LSTMs to exploit their spatial and temporal processing capabilities, respectively, could further enhance the model's ability to handle the multidimensional nature of climate data. This will be investigated in subsequent studies.

Funding statement

Dr. Ibebuchi is funded as a postdoctoral researcher at Kent State University.

Code availability

Python was used for coding the methods as described in the methodology section.

Data availability

The ERA5 data is available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview. Code availability: The typical Python code for encoding gridded climate data is available at https://github.com/cibebuchi/Projects/blob/main/NAO%20AutoEncoder.

Use of AI tools declaration

The author declares he/she has not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The author declares no conflict of interest.

References

1. Philipp A, Della-Marta PM, Jacobeit J, et al. (2007) Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. *J Clim* 20: 4065–4095. https://doi.org/10.1175/JCLI4175.1
2. Pasini A, Lorè M, Ameli F (2006) Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system. *Ecol Modell* 191: 58–67. https://doi.org/10.1016/j.ecolmodel.2005.08.012
3. Mihailović DT, Mimić G, Arsenić I (2014) Climate predictions: The chaos and complexity in climate models. *Adv Meteorol* 2014: 878249. https://doi.org/10.1155/2014/878249
4. Esteban P, Martin-Vide J, Mases M (2006) Daily atmospheric circulation catalogue for Western Europe using multivariate techniques. *Int J Climatol* 26: 1501–1515. https://doi.org/10.1002/joc.1391
5. Philipp A, Bartholy J, Beck C, et al. (2010) Cost733cat–A database of weather and circulation type classifications. *Phys Chem Earth* 35: 360–373. https://doi.org/10.1016/j.pce.2009.12.010
6. Ibebuchi CC, Richman MB (2023) Circulation typing with fuzzy rotated T-mode principal component analysis: methodological considerations. *Theor Appl Climatol*, 495–523. https://doi.org/10.1007/s00704-023-04474-5
7. Huth R, Beck C, Philipp A, et al. (2008) Classifications of atmospheric circulation patterns: recent advances and applications. *Ann N Y Acad Sci* 1146: 105–152. https://doi.org/10.1196/annals.1446.019
8. Deligiorgi D, Philippopoulos K, Kouroupetroglou G (2014) An assessment of self-organizing maps and k-means clustering approaches for atmospheric circulation classification. *Recent Adv Environ Sci Geosci*, 17.
9. James G, Witten D, Hastie T, et al. (2013) *An introduction to statistical learning*. New York: springer. 112: 3–7.
10. Hannachi A, Jolliffe IT, Stephenson DB (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int J Climatol* 27: 1119–1152.
11. Compagnucci RH, Richman MB (2008) Can principal component analysis provide atmospheric circulation or teleconnection patterns? *Int J Climatol J R Meteorol Soc* 28: 703–726.
12. Huth R (2000) A circulation classification scheme applicable in GCM studies. *Theor Appl Climatol* 67: 1–18. https://doi.org/10.1007/s007040070012
13. Chattopadhyay A, Hassanzadeh P, Pasha S (2020) Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Sci Rep* 10: 1317. https://doi.org/10.1038/s41598-020-57897-9
14. Davenport FV, Diffenbaugh NS (2021) Using machine learning to analyze physical causes of climate change: A case study of US Midwest extreme precipitation. *Geophys Res Lett* 48: e2021GL093787. https://doi.org/10.1029/2021GL093787

15. Weyn JA, Durran DR, Caruana R (2020) Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J Adv Model Earth Sy* 12: e2020MS002109. https://doi.org/10.1029/2020MS002109
16. Lee CC, Sheridan SC, Dusek GP, et al. (2023) Atmospheric Pattern–Based Predictions of S2S Sea Level Anomalies for Two Selected US Locations. *Artif Intell Earth Syst* 2: 220057. https://doi.org/10.1175/AIES-D-22-0057.1
17. Hinton GE, Salakhutdinov RR (2006) Reducing the Dimensionality of Data with Neural Networks. *Science* 313: 504–507. https://doi.org/10.1126/science.1127647
18. Murakami H, Delworth TL, Cooke WF, et al. (2022) Increasing frequency of anomalous precipitation events in Japan detected by a deep learning autoencoder. *Earths Future* 10: e2021EF002481. https://doi.org/10.1029/2021EF002481
19. Ibebuchi CC, Abu IO, Nyamekye C, et al. (2024) Utilizing Machine Learning to Examine the Spatiotemporal Changes in Africa's Partial Atmospheric Layer Thickness. *Sustainability* 16: 256. https://doi.org/10.3390/su16010256
20. Myrzaliyeva M (2022) INVESTIGATING THE IMPACT OF CLIMATE CHANGE ON WEATHER REGIMES USING DIMENSIONALITY REDUCTION WITH DEEP AUTOENCODERS. Available from: https://cris.vub.be/ws/portalfiles/portal/94889169/MA_ACS_Myrzaliyeva_Madina_S3_2122_final.pdf.
21. Kurihana T, Franke J, Foster I, et al. (2022) Insight into cloud processes from unsupervised classification with a rotationally invariant autoencoder. *arXiv preprint arXiv*, 2211.00860. https://doi.org/10.48550/arXiv.2211.00860
22. Huang Z, Tan X, Wu X, et al. (2023) Long-Term Changes, Synoptic Behaviors, and Future Projections of Large-Scale Anomalous Precipitation Events in China Detected by a Deep Learning Autoencoder. *J Clim* 36: 4133–4149. https://doi.org/10.1175/JCLI-D-22-0737.1
23. Krinitskiy MA, Zyulyaeva YA, Gulev SK (2019) Clustering of polar vortex states using convolutional autoencoders. Available from: https://ceur-ws.org/Vol-2426/paper8.pdf.
24. Richard G, Grossin B, Germaine G, et al. (2002) Autoencoder-based time series clustering with energy applications. *arXiv preprint arXiv*, 2002.03624. https://doi.org/10.48550/arXiv.2002.03624
25. Tavakoli N, Siami-Namini S, Adl Khanghah M, et al. (2020) An autoencoder-based deep learning approach for clustering time series data. *SN Appl Sci* 2: 937. https://doi.org/10.1007/s42452-020-2584-8
26. Kalinicheva E, Sublime J, Trocan M (2020) Unsupervised satellite image time series clustering using object-based approaches and 3D convolutional autoencoder. *Remote Sens* 12: 1816. https://doi.org/10.3390/rs12111816
27. Harush S, Meidan Y, Shabtai A (2021) DeepStream: autoencoder-based stream temporal clustering and anomaly detection. *Comput Secur* 106: 102276. https://doi.org/10.1016/j.cose.2021.102276
28. Noering FKD, Schroeder Y, Jonas K, et al. (2021) Pattern discovery in time series using autoencoder in comparison to nonlearning approaches. *Integr Comput-Aid E* 28: 237–256. https://doi.org/10.3233/ICA-210650

29. Ibebuchi CC (2021) On the relationship between circulation patterns, the southern annular mode, and rainfall variability in Western Cape. *Atmosphere* 12: 753. https://doi.org/10.3390/atmos12060753
30. Ibebuchi CC (2021) Circulation pattern controls of wet days and dry days in Free State, South Africa. *Meteorol Atmos Phys* 133: 1469–1480. https://doi.org/10.1007/s00703-021-00822-0
31. Hersbach H, Bell B, Berrisford P, et al. (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146: 1999–2049. https://doi.org/10.1002/qj.3803
32. Pedregosa F, Varoquaux G, Gramgort A, et al. (2011). Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12: 2825–2830.
33. Chollet F (2015) Keras. Available from: https://keras.io.
34. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings. 5: 315–323.
35. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv*, 1412.6980.
36. Lorenzo-Seva U, Ten Berge JMF (2006) Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology* 2: 57–64. https://doi.org/10.1027/1614-2241.2.2.57
37. Gardner MW, Dorling SR (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ* 32: 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0
38. Campozano L, Mendoza D, Mosquera G, et al. (2020) Wavelet analyses of neural networks-based river discharge decomposition. *Hydrol Processes* 34: 2302–2312. https://doi.org/10.1002/hyp.13726
39. Castelvecchi D (2016) Can we open the black box of AI? *Nat News* 538: 20–23.
40. Toms BA, Barnes EA, Ebert-Uphoff I (2020) Physically interpretable neural networks for the geosciences: Applications to earth system variability. *J Adv Model Earth Sy* 12: e2019MS002002. https://doi.org/10.1029/2019MS002002
41. Pierdicca R, Paolanti M (2022) GeoAI: a review of artificial intelligence approaches for the interpretation of complex geomatics data. *Geosci Instrum Methods Data Syst* 11: 195–218. https://doi.org/10.5194/gi-11-195-2022
42. Mamalakis A, Ebert-Uphoff I, Barnes EA (2020) Explainable artificial intelligence in meteorology and climate science: Model fine-tuning, calibrating trust and learning new science, *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. Cham: Springer International Publishing. 315–339. https://doi.org/10.1007/978-3-031-04083-2_16
43. Labe ZM, Barnes EA (2021) Detecting climate signals using explainable AI with single-forcing large ensembles. *J Adv Model Earth Sy* 13: e2021MS002464. https://doi.org/10.1029/2021MS002464
44. Karim F, Majumdar S, Darabi H, et al. (2017) LSTM fully convolutional networks for time series classification. *IEEE Access* 6: 1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939
45. Sadouk L (2019) CNN approaches for time series classification. *Time series analysis-data, methods, and applications*, 5: 57–78.

46. Zhao B, Lu H, Chen S, et al. (2017) Convolutional neural networks for time series classification. *J Syst Eng Electron* 28: 162–169. https://doi.org/10.21629/JSEE.2017.01.18
47. Ibebuchi CC, Richman MB (2024) Deep learning with autoencoders and LSTM for ENSO forecasting. *Clim Dyn*.

AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)