Atmospheric Blocking Pattern Recognition in Global Climate Model Simulation Data

Grzegorz Muszynski*†, Prabhat†, Jan Balewski†, Karthik Kashinath†, Michael Wehner‡ and Vitaliy Kurlin*
*Department of Computer Science
University of Liverpool, Liverpool, L69 3BX, UK
†National Energy Research Scientific Computing Center
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
‡Computational Research Division
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

***Abstract*—In this paper, we address a problem of atmospheric blocking pattern recognition in global climate model simulation data. Understanding blocking events is a crucial problem to society and natural infrastructure, as they often lead to weather extremes, such as heat waves, heavy precipitation, and the unusually poor air condition. Moreover, it is very challenging to detect these events as there is no physics-based model of blocking dynamic development that could account for their spatiotemporal characteristics. Here, we propose a new two-stage hierarchical pattern recognition method for detection and localisation of atmospheric blocking events in different regions over the globe. For both the detection stage and localisation stage, we train five different architectures of a convolutional neural network (CNN) based classifier and regressor. The results show the general pattern of the atmospheric blocking detection performance increasing significantly for the deep CNN architectures. In contrast, we see the estimation error of event location decreasing significantly in the localisation problem for the shallow CNN architectures. We demonstrate that CNN architectures tend to achieve the highest accuracy for blocking event detection and the lowest estimation error of event localisation in regions of the Northern Hemisphere than in regions of the Southern Hemisphere.**

I. Introduction

Deviations from the normal atmospheric circulation in the mid-latitudes of both hemispheres usually lead to the formation of climate patterns that are known as atmospheric blocks (ABs) [1], [2]. Blocking is a large-scale climate pattern that is often correlated with extreme weather events, such as floods, cold spells, short-lived droughts, heat waves, and extremely poor air conditions in summer and winter seasons [3]–[6] These weather extremes connected with AB phenomena pose a high risk to society and natural infrastructure. For example, the summer heat wave, combined with a drought in 2003 that claimed tens of thousands of lives in Europe, was partially caused by the resilient AB [7]. In 2012, AB in the North Atlantic steered hurricane Sandy into the north-eastern coast of the United States, and its heavy precipitation resulted in severe flooding and brought the high economic cost to society [8]. Therefore, the importance of characterising and understanding of the evolution of atmospheric patterns, including ABs, in a changing climate cannot be overstated. To analyse how these patterns change under different carbon emission scenarios or large-scale climatic variations, ABs have first to be identified.

Objective identification of climate patterns, including AB phenomena, has progressively emerged as one of the most active research areas in climate science over the last decades. However, it remains a challenge for the climate science community. Existing AB pattern recognition methods are engineered heuristics based on human expertise in defining specific atmospheric phenomena. In other words, these conventional methods are built upon hard constraints on subjective thresholds of relevant physical variables, such as surface pressure, temperature, and wind speed. Moreover, AB patterns do have neither a unique theoretical model nor a clear empirical definition that is universally accepted by all climate science community [2]. ABs can take many forms in terms of the shape, the size, and the location of patterns over the globe. For all these reasons, various sets of criteria and thresholds are used to characterise blocking patterns. This non-uniformity in research on ABs usually results in the discrepancies between outputs of different identification methods [9]. What is more, the existing heuristics do not perform as well as visual inspection of climate experts for identifying these patterns.

There are many ongoing efforts to search for alternative identification methods of climate patterns or extreme weather events, such as atmospheric rivers, extra-tropical cyclones, and hurricanes. These methods include an unsupervised segmentation and a discovery of coherent structures in spatiotemporal systems [10], topological methods [11], and machine learning methods along with deep learning techniques [12]. The above mentioned methods by their inherent design circumvent a critical selection of suitable thresholds of different physical variables to characterise atmospheric patterns. In particular, deep-learning-based approaches can learn robust representations of images from raw pixels, outperforming handcrafted features. Furthermore, the recent advances in deep neural networks, especially deep convolutional neural networks (CNNs), have demonstrated a significant improvement in achieving the state-of-the-art results on image recognition tasks [13]–[15], such as object detection and object segmentation. Despite all of these deep learning breakthroughs, the adoption of deep neural networks in climate science research is still unexplored [16], [17]. Although the atmospheric phenomena pattern recognition task in climate data is similar to the task of finding objects

in natural RGB images, there are important differences, as follows:

- Climate model simulation products have more 'channels' of information than RGB images;
- There is lack of publicly available data annotations or labelled data for supervised machine learning methods;
- Climate data do not share the same statistics as RGB images; hence CNN architectures cannot be pre-trained on publicly available databases (e.g., ImageNet [18]).

In this paper, we address the problem of identifying climate patterns, that is formulated as a pattern recognition task. The key contributions of this work are, as follows:

- We develop a new two-stage hierarchical pattern recognition method for the identification of atmospheric blocking events (ABs) in different regions over the globe, as is shown in Figure 2. The first stage is an AB pattern detection problem defined as a binary classification task, and the second stage is an AB pattern localisation problem defined as a regression task. In the detection stage, a CNN-based classifier distinguishes the AB images from the non-AB images. Those images with detected AB events are passed to the localisation stage. In that stage, a CNN-based regressor predicts AB location parameters in the image, i.e. a mass centre (a latitudinal position and a longitudinal position), and a minimum enclosing circular box. For both the detection step and localisation step, we propose and train customised CNN architectures inspired by the Oxford Visual Geometry Group (VGG) architectures [14].
- We investigate five different architectures of the generic CNN-based classifier and regressor respectively designed for each stage of the AB identification method, as shown in Table I and II. We fix other parameters of the CNN architectures and we steadily increase the network depth by stacking more two-dimensional convolutional layers. Furthermore, we observe the general pattern of the AB detection performance increasing significantly for the deep CNN architectures. In contrast, we see the estimation error of event location decreasing significantly in the localisation problem for the shallow CNN architectures.
- We demonstrate that the proposed CNN architectures tend to achieve better AB detection and localisation performances in regions of the Northern Hemisphere than in regions of the Southern Hemisphere. Moreover, we observe that the classification accuracy gradually declines for each CNN architecture over a period of time. We also indicate that the value of estimation error in the regression task strongly depends on geographical location and the associated climatic variability.

In Section II, a brief overview of AB phenomena research and the existing AB pattern detection methods are described. In Section III, we detail climate data and the created benchmark dataset. Section IV describes the proposed AB pattern recognition method and CNN classifier and regressor, including their architectures. In Section VI, we present and discuss all the results. Section VII presents conclusions and future work.

II. Related Work

In this section, we briefly review a related work on atmospheric blocking (AB) patterns and the existing objective identification methods of them.

A. Research on Atmospheric Blocking

An atmospheric blocking pattern is a resilient obstruction of the normal west-to-east atmospheric circulation in the mid-latitudes of both hemispheres [19]. In other words, this pattern temporally redirects the jet stream and can cause one of the most dangerous extreme weather events, such as heat waves, cold spells, and many others [7]. An AB phenomenon has important consequences on the weather of densely populated areas [8], and that is why early blocking studies have been mainly focused on the Northern Hemisphere. Since the middle of the last century, there have been many observational works on the AB phenomena [1], [20]. The first AB climatology analyses by the atmospheric science community were to investigate the geographical distribution of its observational statistics and find a commonly accepted structural definition of this phenomenon [2]. In the second half of the 20th century, many theoretical studies started looking for a physics-based model of AB dynamic development that could account for its spatiotemporal characteristics. However, to date, there is no theoretical model of AB that can constitute all of its observational characteristics [21]. More recently, much attention has been paid to the evaluation of AB occurrence in global climate models. AB patterns still remain a challenge for numerical simulations, although numerous improvements in atmospheric modelling have been made by the climate science community [22].

B. Atmospheric Blocking Pattern Recognition Methods

Turning all broad statements about the qualitative features of AB patterns into an objective identification method is not straightforward. For example, to compare climate simulations with the observational records of AB or non-AB, these patterns have to be first found in climate model products. However, certain features of AB occur in the majority of these patterns, ABs may have various spatial structures or shapes that depend on a specific geographic location and season.

The first attempts at identifying ABs required comprehensive visual inspection of the atmospheric flow patterns and were also limited in their scope [1]. Such simple analysis usually leads to potentially different conclusions about AB patterns. For this reason, a wide range of objective identification methods has been proposed over the years. Many of them are based on 1D analysis techniques called blocking indices [23]–[25]. Some of these 1D blocking indices have been modified into 2D indices of flow fields [9], [26]. A couple of intercomparison studies have been done to quantify if both 1D and 2D identification techniques produce consistent AB characteristics, such as blocking frequency, duration and location [9].

III. Data

A. Climate Dataset

In climate science research, one can distinguish three categories of global datasets: model simulation products, observational data (e.g., satellite images), and reanalysis products (i.e., generated by combining climate models with observations). In this study, we analyze the last category because the reanalysis provides a comprehensive historical record of the Earth's climate and gives a reliable way to monitor how fast it is changing. That is why it fits the problem of studying climate patterns, i.e. atmospheric blocking (AB) phenomena.

In our analysis, we use the ERA-Interim reanalysis product from the European Center for Medium-Range Weather Forecasting. We use five physical variables on a regular grid: temperature, meridional and zonal wind, geopotential height, and potential vorticity variables at eight different pressure levels in millibars (mb): {150mb, 200mb, 250mb, 300mb, 350mb, 400mb, 450mb, 500mb}, 6-hourly timesteps at approximately 80 km spatial resolution (180 pixels $\times$ 360 pixels based on T119 spectral model grid resolution) in the period of January 1, 1980 - December 31, 2016.

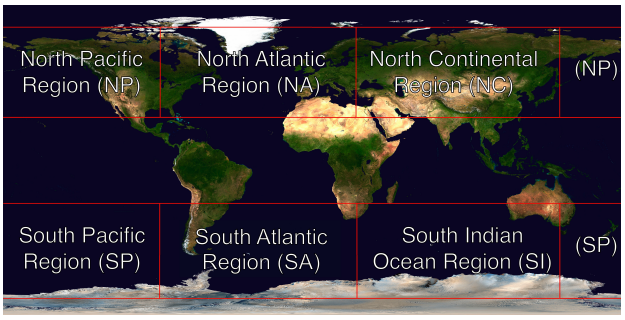


Fig. 1. An example of the world map (globe) with six defined regions of interest: the North Pacific Region (NP); the North Atlantic Region (NA); the North Continental Region (NC); the South Pacific Region (SP); the South Atlantic Region (SA), and the South Indian Ocean Region (SI).

B. Generated Dataset

The publicly available dataset of identified AB patterns is based on a climatological field diagnostic procedure of Schwierz, et al. [26] and is provided by the Institute for Atmospheric and Climate Science at the ETH Zurich, Switzerland. The procedure consists of a few steps: a semi-automatic detection algorithm, and 2D blocking index method that was applied to the ERA-Interim reanalysis data [27]. The output of this procedure yields 2D binary fields (masks) with the value one at pixels that meet the procedure criteria for AB and the value zero at pixels that do not belong (non-AB).

We generated the ground-truth labels (i.e., AB class vs non-AB class) based on the output of the procedure mentioned above. The global image of 180 pixels $\times$ 360 pixels $\times$ 40 channels is divided into six images (regions) of size 60 pixels $\times$ 120 pixels $\times$ 40 channels. Location of regions and extent are based on the studies of [28]. We can distinguish six regions: the North Pacific Region (NP); the North Atlantic Region (NA); the North Continental Region (NC); the South Pacific Region (SP); the South Atlantic Region (SA), and the South Indian Ocean Region (SI), as shown in Figure 1. Each region is roughly centred over a local maximum of AB main frequency occurrence. We use only images of mid-latitudes regions of both hemispheres on the globe. The positive class label is assigned to an image if it contains a compact binary blob of size greater or equal to the average AB blob size. The negative class label is assigned to an image if it does not have a binary blob. The generated labelled dataset has approximately 140K images of both AB and non-AB events, where the number of samples per class is almost balanced. For the localisation (regression) problem, we select the images containing AB phenomena. Then we calculate the centroid or the mass centre of a binary blob in the region and a radius of a minimum bounding circular box.

IV. Methodology

In this section, we describe a hierarchical pattern recognition method for the identification of atmospheric blocking events (ABs). As shown in Figure 2, the hierarchical method consists of two stages:

- In the first stage, AB pattern detection is formulated as a binary classification task. A convolutional neural network (CNN) based classifier distinguishes the AB images from the non-AB images in different regions over the globe. Those images with detected AB events are passed to the second stage.
- In the second stage, AB pattern localisation is defined as a regression problem. A CNN-based regressor predicts AB location parameters in the images, i.e. a mass centre (a latitudinal position and a longitudinal position), and a radius of a minimum enclosing circular box of these events in different regions of the globe.

Both stages employ customised CNN architectures inspired by VGG architectures (VGGs) [14]. Because the VGGs are the most commonly referred ones in the deep learning literature, and one of the state-of-the-art architectures in classification task and localisation task. In both tasks, we investigate five different architectures of the generic CNNs designed for each stage of the method, as outlined in Table I and Table II.

A. Detection of Atmospheric Blocking Events

In this stage, we develop a generic CNN-based classifier to which we refer as the architecture A, as outlined in Table I. The architecture has eight weight layers, including four convolutional layers (conv) and four fully connected layers (FC). Each convolutional layers is followed by a max pool layer (Maxpooling). The width of convs starts from 64 filters in the first layer and then increases by a factor of two after each Maxpooling layer until it reaches 512 filters. In contrast, the width of FC layers starts from 512 channels and then decreases by a factor of two, until it reaches 64 channels. The output FC layer performs a binary classification and therefore contains one value, i.e. an AB label or a non-AB label.

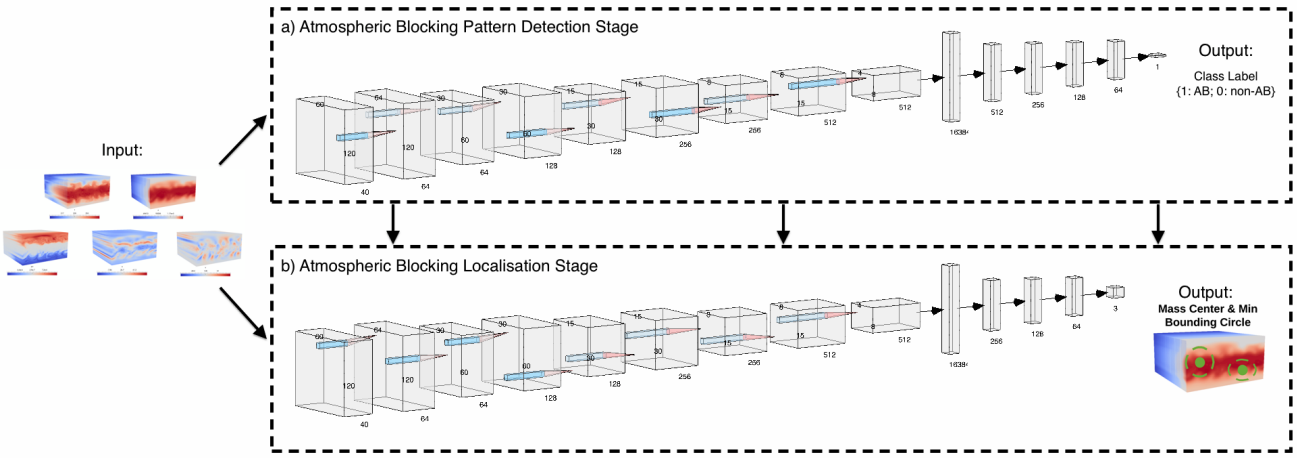


Fig. 2. The big picture of a hierarchical atmospheric blocking pattern recognition method. The method consists of two stages: (a) illustrates an architecture of convolutional neural network (CNN) based classifier that distinguishes atmospheric blocking events (AB label) and otherwise non-blocking events (non-AB label); and (b) illustrates an architecture of CNN-based regressor that predicts three values describing the AB location, i.e. a latitudinal position, a longitudinal position, and a radius of a minimum enclosing circular box.

All five architectures of a CNN-based classifier follow the architecture A and gradually increase the number of conv layers (their depth). The configuration of the FC layers is the same in all architectures. All five architectures are referred by their names (A-E) and all details on the architectures are listed in Table I, one per column.

TABLE I
CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURES ARE SHOWN IN COLUMNS. THE DEPTH OF THE ARCHITECTURES INCREASES FROM THE LEFT (A) TO THE RIGHT (E). THE PARAMETERS OF BOTH TYPES OF LAYERS ARE DENOTED AS FOLLOWS: CONV-(NUMBER OF FILTERS); FC-(NUMBER OF CHANNELS).

CNN Architectures				
A	B	C	D	E
Input: (60 × 120 × 40)				
conv-64	conv-64 conv-64	conv-64 conv-64	conv-64 conv-64	conv-64 conv-64
Maxpooling				
conv-128	conv-128	conv-128 conv-128	conv-128 conv-128	conv-128 conv-128
Maxpooling				
conv-256	conv-256	conv-256	conv-256 conv-256	conv-256 conv-256
Maxpooling				
conv-512	conv-512	conv-512	conv-512	conv-512 conv-512
Maxpooling				
FC-512				
FC-256				
FC-128				
FC-64				
Output: FC-1				

B. Localisation of Atmospheric Blocking Events

In this stage, we develop a generic CNN-based regressor to which we refer as the architecture A, as outlined in Table II. The architecture has totally seven weight layers, including four convolutional layers (conv) and three fully connected layers (FC). Each convolutional layers is followed by a max pool layer (Maxpooling). The width of convs starts from 64 filters in the first layer and then increases by a factor of two after each Maxpooling layer, until it reaches 512 filters. In contrast, the width of FC layers starts from 256 channels and then decreases by a factor of two, until it reaches 64 channels. The output FC layer performs a multi-regression task and therefore contains three values, i.e. a latitudinal position, longitudinal position, and a radius of a minimum enclosing circular box.

All five architectures of a CNN-based regressor follow the architecture A and gradually increase the number of conv layers (their depth). The configuration of the FC layers is the same in all architectures. All five architectures are referred by their names (A-E) and all details on the architectures are listed in Table II, one per column.

TABLE II
CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURES ARE SHOWN IN COLUMNS. THE DEPTH OF THE ARCHITECTURES INCREASES FROM THE LEFT (A) TO THE RIGHT (E). THE PARAMETERS OF BOTH TYPES OF LAYERS ARE DENOTED AS FOLLOWS: CONV-(NUMBER OF FILTERS); FC-(NUMBER OF CHANNELS).

CNN Architectures				
A	B	C	D	E
Input: (60 × 120 × 40)				
conv-64	conv-64 conv-64	conv-64 conv-64	conv-64 conv-64	conv-64 conv-64
Maxpooling				
conv-128	conv-128	conv-128 conv-128	conv-128 conv-128	conv-128 conv-128
Maxpooling				
conv-256	conv-256	conv-256	conv-256 conv-256	conv-256 conv-256
Maxpooling				
conv-512	conv-512	conv-512	conv-512	conv-512 conv-512
Maxpooling				
FC-256				
FC-128				
FC-64				
Output: FC-3				

V. Experimental Settings

A. The Setting of Experiments

We perform 10-fold-cross-validation (CV) on the dataset, where eight folds are for training, the one fold is for validation, and the one fold is for testing. In each CV-round, five convolutional neural networks (CNNs) with different hyperparameters are trained on a training set, evaluated on a validation set, and tested on a testing set. There are two hyperparameters that we search, i.e. dropout rate of neurons in fully connected layers (i.e., $[0.2; 0.6]$) and batch size (i.e., $\{32, 64, 128, 256\}$). The hyperparameters are selected through a random search procedure.

CNNs are trained using Adam optimizer, which is the adaptive learning rate optimization algorithm. For the regression problem, we used the mean squared error loss function, and for the classification task, we used the binary cross-entropy loss function. The rectified linear activation function is used in all convolutional layers and all fully connected layers. The output layer of CNN-based classifier uses the sigmoid activation function and the output layer of CNN-based regressor uses the hyperbolic tangent activation function.

B. Computational Platform

We performed our data processing, CNN architectures training and testing on Cori, a Cray XC40 supercomputing system at the National Energy Research Scientific Computing Center (NERSC). Each of Cori computing node has 32 2.3 GHz Intel Haswell processors or has 68 1.4 GHz Intel KNL processors. In our computations, we used single node TensorFlow backend of Keras. The random search procedure of hyper-parameter optimisation was distributed across Cori compute nodes with tasks fully parallel on 32 cores and 68 cores.

C. Evaluation Metrics

To assess the performance of architectures of a CNN-based classifier and a CNN-based regressor, we use the following metrics: a classification accuracy (ACC) and F1 score for a detection task; Lin's concordance correlation coefficient (CCC) and the mean percentage error (MPE) for a regression problem. The F1 score measures a weighted harmonic mean of the precision and sensitivity of the classifier in case of imbalanced class problem in data of different regions over the globe. The CCC value is the measure of agreement between the true variables and the predicted variables by the regressor. The MPE measures if the regressor systematically underestimates or overestimates the predicted variables. We use McNemar's statistical test at a significance level of 0.05 on the output of classification task [29] and Wilcoxon Signed-Rank test at a significance level of 0.05 for the output regression task [30]. Both tests evaluate which CNN architecture performs the best, i.e. we observe a statically significant increase and decrease in classification accuracy and error of predicting localisation parameters of AB patterns, respectively. In order to facilitate the comparison, CNN architectures have been ranked according to the ACC values for the classification task and the respective MPE values for the regression task.

VI. Results and Discussion

In this section, we present and discuss the obtained results in detecting and localising atmospheric blocking (AB) events in the dataset generated from the ERA-Interim climate model simulation product, as described in Section III-B.

A. Results

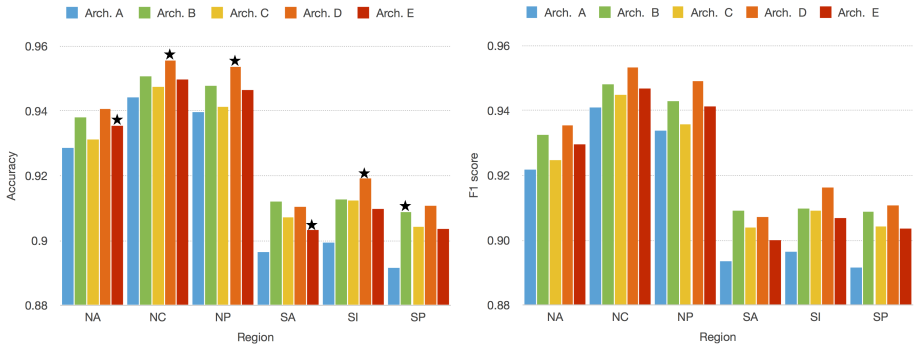
Figure 3 (a) and Figure 3 (b) show a bar chart of classification accuracy measure (ACC) and a bar chart of the mean F1 score of both classes, respectively. The ACC and F1 score values are computed for five architectures outlined in Table I, and for all six regions over the globe shown in Figure 1.

What can be seen in these charts is the sharp decrease in the architectures ACC and F1 score values in the regions of the Southern Hemisphere (i.e., SA, SI, and SP) in comparison to the architectures ACC values and F1 scores in the regions of the Northern Hemisphere (i.e., NA, NC, and NP). Both bar charts reveal that the architecture D of the proposed classifier outperforms the other architectures in regions of both hemispheres, regarding ACC and F1 score values. The only exception to this is the SA region in which values of performance metrics are slightly higher for the architecture B than the architecture D. It can be seen that the architecture A reached the lowest performance for all six regions. Overall, we can observe a statistically significant increase in the ACC values (p-value $\ll 0.05$) for: architecture E in the NA region; architecture D in the NC region; architecture D in the NP region; architecture E in the SA region; architecture D in the SI region, and architecture B in the SP region.

Figure 4 (a) and Figure 4 (b) display the mean values of ACCs and F1 scores (the mean of both classes) for the entire globe over time (i.e., averages of five year periods). If we look at the trends over time, we can see the fluctuation in the ACC and F1 score values for all architectures. However, there is no sign of levelling off; we can observe that for some architectures (i.e., architecture B and architecture A) the scores drop gradually over time after the initial surge. Furthermore, we can see that architecture B, architecture C, architecture D, and architecture E converge moderately at similar ACC and F1 score values in most of the periods.

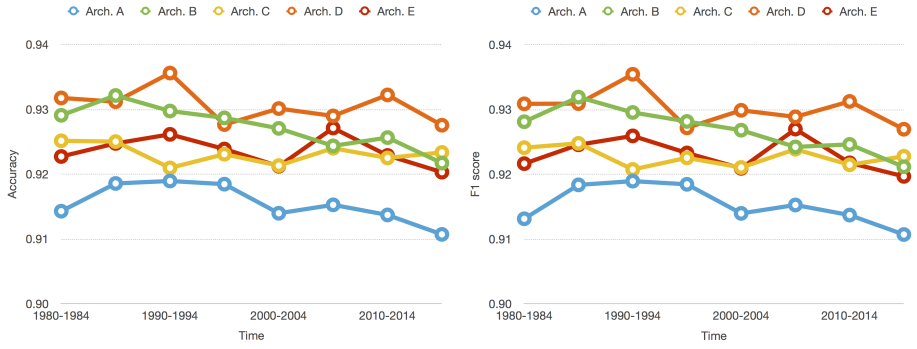
Figure 5 displays a bar chart of Lin's concordance correlation coefficient (CCC) for all architectures in Table II, and all regions over the globe shown in Figure 1. What stands out in this chart is a steep decline in CCC values for the architectures in regions of the Southern Hemisphere (i.e., SA, SI, and SP) in comparison to CCC values for the architectures in regions of the Northern Hemisphere (i.e., NA, NC, and NP). The chart shows the general dominance pattern of CCC values for the architecture A in regions of both hemispheres, except for the SA region. In contrast, it can be seen that CCC values for the architecture E are the lowest for all the regions. In general, we can observe more complex architecture is, more CCC values decrease.

In Figure 6 (a), (b), and (c) show bar charts of mean percentage error (MPE) of three predicted parameters describing the location of AB patterns for all architectures in Table II,



(a) (b)

Fig. 3. Performance of convolutional neural network (CNN) architectures: architecture A, architecture B, architecture C, architecture D, and architecture E; for regions of the North Atlantic Region (NA), the North Continental Region (NC), the North Pacific Region (NP), the South Atlantic Region (SA), the South Indian Ocean Region (SI) and the South Pacific Region (SP). Left bar chart (a) illustrates classification accuracy for each architecture per region and the right chart (b) displays F1 score for each architecture per region. The $\star$ symbol stands for a p-value $\ll 0.05$.



(a) (b)

Fig. 4. Performance of convolutional neural network (CNN) architectures: architecture A, architecture B, architecture C, architecture D, and architecture E; for regions of the North Atlantic Region (NA), the North Continental Region (NC), the North Pacific Region (NP), the South Atlantic Region (SA), the South Indian Ocean Region (SI) and the South Pacific Region (SP). Left bar chart (a) illustrates the mean classification accuracy of each architecture per region over five year periods and the right chart (b) displays the mean F1 score for each architecture per region over five year periods.

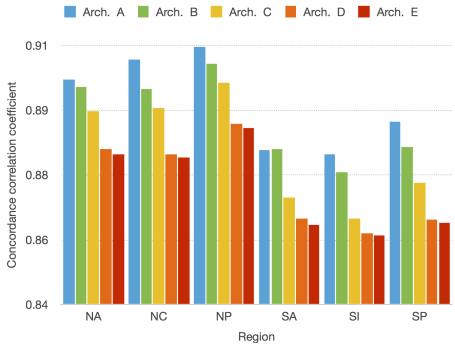


Fig. 5. Performance of convolutional neural network (CNN) architectures: architecture A, architecture B, architecture C, architecture D, and architecture E; for the North Atlantic Region (NA), the North Continental Region (NC), the North Pacific Region (NP), the South Atlantic Region (SA), the South Indian Ocean Region (SI) and the South Pacific Region (SP). The bar chart illustrates the Lin's concordance correlation coefficient for each architecture per region.

and all regions over the globe shown in Figure 1.

Figure 6 (a) displays the MPE of the latitudinal mass centre position of AB events. We can see that all architectures overestimate the latitudinal parameter (positive error) in all the regions. The chart reveals that there has been a steep decrease in the overestimating the parameter by all architectures in the NP region and SI region. It can be observed that the architecture E has the highest MPE in four out of six regions (i.e., NA, SA, SI, and SP). Overall, the MPE of latitudinal parameter decreases statistically significantly (p-value $\ll 0.05$) for: architecture A in the SA region, architecture B in the SP region; architecture C in the NA region, and architecture D in the NC, NP and SI regions.

Figure 6 (b) shows the MPE of the longitudinal mass centre position of AB events. It can be seen that different architectures underestimate the longitudinal parameter in five out of six regions (i.e., NA, NC, NP, SA, and SI). The exception to this is the architecture B in the region SA, which the MPE value is overestimated. What stands out in this chart is that MPE values are positive for the most architectures in one out of six regions, i.e. the SP region, except the architecture A which has a small negative MPE value. The MPE of longitudinal

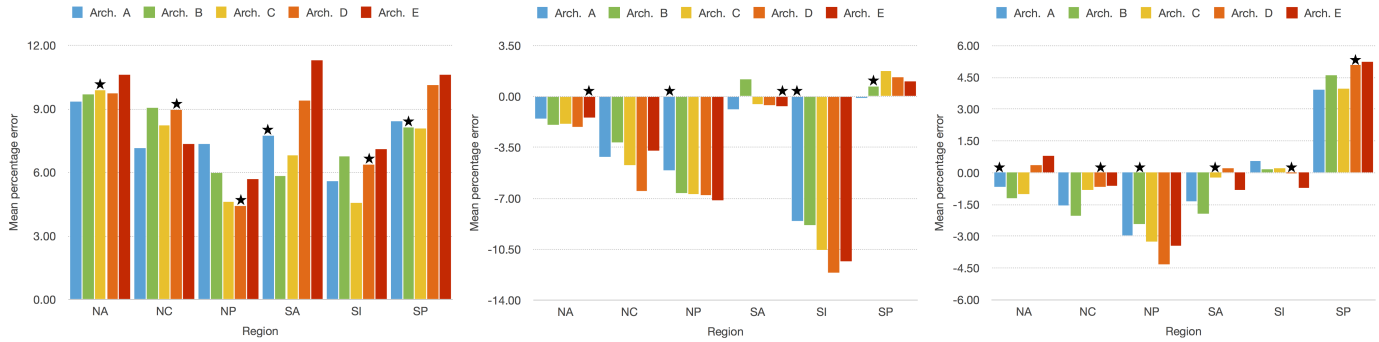


Fig. 6. Performance of convolutional neural network (CNN) architectures: architecture A, architecture B, architecture C, architecture D, and architecture E; for regions of the North Atlantic Region (NA), the North Continental Region (NC), the North Pacific Region (NP), the South Atlantic Region (SA), the South Indian Ocean Region (SI) and the South Pacific Region (SP). Top left bar chart (a) illustrates mean percentage error for each architecture in estimating the latitudinal position of the mass centre of atmospheric blocks (ABs) per region and the top right chart (b) displays mean percentage error for each architecture in estimating the longitudinal position of the mass centre of ABs per region. The bottom chart shows mean percentage error for each architecture in estimating the radius of the mass centre of ABs per region. The $\star$ symbol stands for a p-value $\ll 0.05$.

parameter decreases statistically significantly (p-value $\ll 0.05$) for architecture A in the NP and SI regions; architecture B in the SP region, and architecture E in the NA and SA regions. The NC region is the exception to this in which there is no sign of a significant decrease in the MPE values.

Figure 6 (c) displays the MPE of a radius of a minimum bounding box located in the mass centre of AB events. We can observe low MPE values for architectures in four out of six regions (i.e., NA, NC, SA, and SI). In contrast, it can be seen that the MPE values for all architectures are high in NP and SP regions. In general, the chart reveals that architectures tend to underestimate the radial parameter in five out of six regions. The MPE of the parameter decreases statistically significantly (p-value $\ll 0.05$) for: architecture A in the NA region; architecture B in the NP region; architecture C in the SA region; and architecture D in the NC and SP regions.

B. Discussion

The results that we obtained indicate that it is possible to accurately detect AB events in climate model simulation products by using CNN architectures.

The ACC and F1 score values are the highest for all architectures in three regions of the Northern Hemisphere (N-H). That can be justified by the fact that AB events are much more frequent in the N-H than in the Southern Hemisphere (S-H). Moreover, ABs has important consequences on the weather of densely populated Europe and North America, that is why innumerable observational data improve the ability of climate models to simulate AB patterns in the N-H correctly.

The results of the mean ACC and F1 score values fluctuate over time. This variability in the values can be explained by AB number of occurrence that is sometimes modulated by other periodic climate regimes, for example, El Niño-Southern Oscillation. We observe that the values of ACC or F1 score increase significantly for CNN architectures with a large number of parameters for various regions on the globe.

The values of the CCC are generally higher for all architectures in all regions of the N-H than for models in the Southern Hemisphere. That can also be partially explained by better observational characteristics of AB events and more improvements in the ability of climate models to represent correctly atmospheric dynamics of these events.

The values of the MPE of the latitudinal mass centre position of AB events and the values of the longitudinal mass centre position of these events are usually overestimated and underestimated, respectively. Two reasons can cause that: in the N-H, AB events tend to occur in particular longitudinal and latitudinal localisation, however; in the S-H, these events often occur at slightly lower latitude positions than in the N-H. It can be observed that the values of the MPE decrease statistically significantly for CNN detection models without any particular trend in some geographical regions.

The values of the MPE of a radius of a minimum bounding box located in the mass centre of AB events are relatively low in most of the regions. That can suggest that despite a broad spectrum of spatial location, AB events are probably more similar in size in both hemispheres. Moreover, the MPE values are much higher in the Pacific Ocean regions, which could suggest that the climate model does not correctly represent ABs spatial size due to the vast extent of the ocean.

VII. Conclusions and Future Work

In this paper, we study an identification problem of atmospheric blocking patterns (ABs) in the climate dataset created from a global climate model simulation output. To identify AB events, we propose to apply a hierarchical pattern recognition method that consists of two stages: the detection stage that is formulated as a binary classification task and the localisation stage that is defined as a regression problem. We explore five different convolutional neural network (CNN) architectures and evaluate their performance in six geographical regions over the globe.

We show that the proposed method overall achieves high detection accuracy of AB events and low estimation error of their localisation in different regions. We observe a strong relationship between the CNN architecture depth and the performance in a specific geographical region, i.e. the Northern Hemisphere (N-H) vs the Southern Hemisphere (S-H). That can be explained by the fact that AB events tend to occur much more frequently in the N-H than the S-H.

In the future work, we will explore 3D CNN architectures to study the AB identification problem. We also plan to use customised loss function that can be beneficial in this problem. We will also attempt to reveal more statistical features of AB events as more data will be collected.

ACKNOWLEDGMENT

We thank Marie McGraw (Colorado State University) for sharing her expertise on atmospheric blocking phenomena, and Steinfeld Daniel (ETH Zurich) and Stephan Pfahl (Freie Universität Berlin) for sharing the blocking index dataset. Grzegorz Muszynski and Vitaliy Kurlin would like to acknowledge Intel for supporting the IPCC at the University of Liverpool, UK. Prabhat, Jan Balewski, and Karthik Kashinath were supported by the Intel Big Data Center at NERSC, and Michael Wehner was supported by the Regional and Global Climate Modeling Program of the Office of Biological and Environmental Research in the Department of Energy Office of Science under contract no. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231.

REFERENCES

[1] D. F. Rex, "Blocking action in the middle troposphere and its effect upon regional climate," *Tellus*, vol. 2, no. 4, pp. 275–301, 1950.

[2] S. Tibaldi and F. Molteni, "Atmospheric blocking in observation and models," in *Oxford Research Encyclopedia of Climate Science*, 2018.

[3] S. Pfahl and H. Wernli, "Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the northern hemisphere on (sub-) daily time scales," *Geophysical Research Letters*, vol. 39, no. 12, 2012.

[4] T. Buehler, C. C. Raible, and T. F. Stocker, "The relationship of winter season north atlantic blocking frequencies to extreme cold or dry spells in the era-40," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 63, no. 2, pp. 174–187, 2011.

[5] G. Gangoiti, L. Alonso, M. Navazo, A. Albizuri, G. Perez-Landa, M. Matabuena, V. Valdenebro, M. Maruri, J. A. García, and M. M. Millán, "Regional transport of pollutants over the bay of biscay: analysis of an ozone episode under a blocking anticyclone in west-central europe," *Atmospheric Environment*, vol. 36, no. 8, pp. 1349–1361, 2002.

[6] C.-C. Hong, H.-H. Hsu, N.-H. Lin, and H. Chiu, "Roles of european blocking and tropical-extratropical interaction in the 2010 pakistan flooding," *Geophysical Research Letters*, vol. 38, no. 13, 2011.

[7] R. García-Herrera, J. Díaz, R. M. Trigo, J. Luterbacher, and E. M. Fischer, "A review of the european summer heat wave of 2003," *Critical Reviews in Environmental Science and Technology*, vol. 40, no. 4, pp. 267–306, 2010.

[8] T. M. Hall and A. H. Sobel, "On the impact angle of hurricane sandy's new jersey landfall," *Geophysical Research Letters*, vol. 40, no. 10, pp. 2312–2315, 2013.

[9] M. C. Pinheiro, P. Ullrich, and R. Grotjahn, "Atmospheric blocking and intercomparison of objective detection methods: flow field characteristics," *Climate Dynamics*, pp. 1–28, 2019.

[10] A. Rupe, N. Kumar, V. Epifanov, K. Kashinath, O. Pavlyk, F. Schlimbach, M. Patwary, S. Maidanov, V. Lee, M. Prabhat *et al.*, "Disco: Physics-based unsupervised discovery of coherent structures in spatiotemporal systems," in *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC)*. IEEE, 2019, pp. 75–87.

[11] G. Muszynski, K. Kashinath, V. Kurlin, M. Wehner *et al.*, "Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets," *Geoscientific Model Development*, vol. 12, no. 2, pp. 613–628, 2019.

[12] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins *et al.*, "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv preprint arXiv:1605.01156*, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representation (ICLR*, 2015.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[16] E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, M. Prabhat, and C. Pal, "Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events," in *Advances in Neural Information Processing Systems*, 2017, pp. 3402–3413.

[17] M. Mudigonda, S. Kim, A. Mahesh, S. Kahou, K. Kashinath, D. Williams, V. Michalski, T. O'Brien, and M. Prabhat, "Segmenting and tracking extreme climate events using neural networks," in *Deep Learning for Physical Sciences (DLPS) Workshop, held with NIPS Conference*, 2017.

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[19] T. Glickman, "American meteorological society," *Glossary of Meteorology. 2nd ed. Boston, MA: American Meteorological Society*, 2000.

[20] C.-G. Rossby, "On the dynamics of certain types of blocking waves," *Journal of the Chinese Geophysical Society*, vol. 2, pp. 1–13, 1951.

[21] N. Nakamura and C. S. Huang, "Atmospheric blocking as a traffic jam in the jet stream," *Science*, vol. 361, no. 6397, pp. 42–47, 2018.

[22] P. Davini and F. D'Andrea, "Northern hemisphere atmospheric blocking representation in global climate models: Twenty years of improvements?" *Journal of Climate*, vol. 29, no. 24, pp. 8823–8840, 2016.

[23] R. M. Dole and N. D. Gordon, "Persistent anomalies of the extratropical northern hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics," *Monthly Weather Review*, vol. 111, no. 8, pp. 1567–1586, 1983.

[24] S. Tibaldi and F. Molteni, "On the operational predictability of blocking," *Tellus A*, vol. 42, no. 3, pp. 343–365, 1990.

[25] A. A. Scaife, T. Woollings, J. Knight, G. Martin, and T. Hinton, "Atmospheric blocking and mean biases in climate models," *Journal of Climate*, vol. 23, no. 23, pp. 6143–6152, 2010.

[26] C. Schwierz, M. Croci-Maspoli, and H. Davies, "Perspicacious indicators of atmospheric blocking," *Geophysical research letters*, vol. 31, no. 6, 2004.

[27] M. Sprenger, G. Fragkoulidis, H. Binder, M. Croci-Maspoli, P. Graf, C. M. Grams, P. Knippertz, E. Madonna, S. Schemm, B. Škerlak *et al.*, "Global climatologies of eulerian and lagrangian flow features based on era-interim," *Bulletin of the American Meteorological Society*, vol. 98, no. 8, pp. 1739–1748, 2017.

[28] J. M. Wiedenmann, A. R. Lupo, I. I. Mokhov, and E. A. Tikhonova, "The climatology of blocking anticyclones for the northern and southern hemispheres: Block intensity as a diagnostic," *Journal of Climate*, vol. 15, no. 23, pp. 3459–3473, 2002.

[29] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[30] F. Wilcoxon, "Probability tables for individual comparisons by ranking methods," *Biometrics*, vol. 3, no. 3, pp. 119–122, 1947.