Computers & Geosciences 166 (2022) 105176

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Polar night jet characterization through artificial intelligence

María Rodríguez-Montes [a,*], Blanca Ayarzagüena [b], María Guijarro [a]

[a] Facultad de Informática. Universidad Complutense de Madrid, C/Profesor José García Santesmases, 9, 28040, Madrid, Spain
[b] Facultad de Ciencias Físicas. Universidad Complutense de Madrid, Plaza Ciencias, 1, 28040, Madrid, Spain

ARTICLE INFO

Keywords:
Stratosphere
Polar night jet
Region growing
Machine learning
Climate change
Decision trees

ABSTRACT

The stratospheric polar vortex is a cyclonic circulation that forms over the winter pole, whose edge is characterized by a strong westerly jet (also called polar night jet, PNJ). The PNJ plays a key role in processes such as the distribution of atmospheric constituents in the polar stratosphere or the wave propagation. Further, variations in PNJ can also affect the troposphere, being behind the occurrence of extreme events near the Earth's surface. Thus, it is important to correctly characterize the mean state of the PNJ and its variability. Already existing algorithms, although working, may present several issues. The simplest ones, those based on zonal mean wind, can miss important information. In contrast, the 2-dimensional ones usually involve multiple calculations with several fields, some of them not always included in typical datasets.

In this study, we describe a new artificial intelligence technique to characterize the PNJ. The algorithm only requires data of zonal wind that is classified each time step with a decision trees algorithm with 95.5% accuracy, trained with images processed by a climate science researcher. The classifier is applied to JRA-55 reanalysis data and the output of simulations of three climate models and is found to perform reasonably well when validated against traditional zonal-mean methods. Indeed, it provides more information about the PNJ, as it offers in one step the PNJ region, averaged magnitudes and even identify if the PNJ is under perturbed conditions. We have explored two examples of potential applications of the classifier such as the study of the influence of climate change on the PNJ and the variability of the PNJ on monthly and daily scales. In both cases, our algorithm has produced coherent results with those produced with previous studies, but with more detail obtained at a single step.

1. Introduction

The stratospheric polar vortex is a cyclonic circulation that forms every winter in the stratosphere over the winter pole. The polar vortex can also be identified by a region of low geopotential height and high potential vorticity (PV, a measure of rotation and stratification of the air parcel) (Waugh and Polvani, 2010). The edge of the polar vortex is characterized by sharp PV gradients and a core of strong westerly jet, which roughly corresponds to what is known as polar night jet (PNJ). Indeed, steeper PV gradients are linked to a stronger jet. The origin of the polar vortex and PNJ is the large temperature difference between the pole and lower latitudes during winter, due to the lack of sunlight over the winter pole (Waugh and Polvani, 2010). Consequently, it forms in fall, intensifies in mid-winter and disappears during spring, when the sunlight returns to the pole. The PNJ also acts as an air-barrier in such a way that polar air is isolated from air at lower latitudes.

The stratospheric polar vortex and PNJ are not only controlled by radiative processes, but dynamical effects also play a relevant role in its characterization, particularly in the Northern Hemisphere. Large-scale (Rossby) waves propagate upward from the troposphere and dissipate in the stratosphere, depositing energy and easterly momentum and so, perturbing and weakening the westerly winds (Waugh and Polvani, 2010). The Rossby wave breaking takes place in the PNJ area, colocated with steep PV gradients, since PV gradients are involved in the restoring mechanism for the wave propagation. Thus, the location of the PNJ is of extreme relevance for the vortex, and by extension, for the polar stratosphere dynamics. The wave activity is substantially higher in the Northern Hemisphere than in the Southern Hemisphere, explaining the large differences in climatological intensity of the polar vortex between both hemispheres and hence, their different behavior (Andrews et al., 1987). Bearing in mind these differences, from now on we will focus on the Northern Hemisphere.

The wave activity is not constant during winter, but it undergoes a large variability that results in high variability in the polar vortex too (e.g. Waugh and Polvani, 2010). This has relevant consequences

* Corresponding author.
E-mail addresses: marodr51@ucm.es (M. Rodríguez-Montes), bayarzag@ucm.es (B. Ayarzagüena), mguijarro@ucm.es (M. Guijarro).
URLs: http://stream-ucm.es/Blanca_Ayarzaguena.html (B. Ayarzagüena), https://www.ucm.es/directorio?id=25324 (M. Guijarro).

https://doi.org/10.1016/j.cageo.2022.105176
Received 27 October 2021; Received in revised form 8 April 2022; Accepted 10 June 2022
Available online 18 June 2022

not only in other parts of the stratosphere but also in other atmospheric layers such as the troposphere. For instance, an extraordinary amplification of certain Rossby waves can lead to extreme polar vortex events called sudden stratospheric warmings (SSWs), which consist of a reversal of meridional temperature gradient, as well as a replacement of a westerly by an easterly jet (Labitzke, 1981; McInturff, 1978; Pomerantz, 1963; Baldwin et al., 2021). In these cases, apart from a weakening of the polar vortex, the vortex either shifts out of the pole or splits into two pieces of similar size (Baldwin et al., 2021). SSWs influence near-surface weather, increasing the likelihood of cold air outbreaks over North America and Eurasia, and modifying jet streams, temperatures, rainfall and storm tracks (e.g Karpechko et al., 2018; Baldwin et al., 2021; Lu et al., 2021).

The stratospheric polar vortex and PNJ are also influenced by anthropogenic changes in greenhouse gases concentrations (GHGs). However, there is still high uncertainty in the stratospheric response to these changes across studies due to the different involved processes (Wu et al., 2019; Ayarzagüena et al., 2020; Rao and Garfinkel, 2021). On one hand, increasing GHGs leads to a radiative cooling in the stratosphere (Fels et al., 1980). This is particularly intense in the winter pole given the absence of sunlight, and so, would lead to a stronger polar vortex in the future. On the other hand, tropospheric wave activity entering into the stratosphere is also affected by GHGs changes in such a way that it is enhanced under anthropogenic climate change conditions (McLandress and Shepherd, 2009). This would result in a weaker and more disturbed polar vortex and consequently, an increase in SSWs frequency (Bell et al., 2010). The combination of these opposing effects leads to a weak amplitude of the stratospheric response to increasing GHGs (Mitchell et al., 2012). In addition, the overestimation or underestimation of any of these processes in individual climate models might result in a change of the vortex of opposite sign, explaining the lack of consensus across models in this topic. This large spread in future projections of the stratospheric polar vortex does not only impact the projections of the rest of the stratosphere, i.e. other stratospheric regions and pressure levels. Given the mentioned coupling between the stratosphere and other atmospheric layers, it also affects the future predictions of near-surface circulation (Simpson et al., 2018).

The characterization of the Northern Hemisphere polar vortex and in particular, the PNJ in different datasets and climate scenarios is, therefore, of high relevance for the whole climate community. Some methods already exist in the literature, described hereafter. The standard method for characterizing the intensity of the PNJ is to average the zonal wind speed (u) along the 60 degrees north (60° N) at 10 hPa, used in many studies such as Charlton and Polvani (2007), Hardiman et al. (2019) or Ayarzagüena et al. (2020). This location is close to the core of the PNJ and matches well with where higher changes of the winds were identified in observations (Waugh and Polvani, 2010). This method has several advantages, particularly its simplicity. It only requires zonal mean zonal wind data and can be applied straight-forward to several datasets such as observations and the output of model simulations. However, the precise location of the PNJ is important too and might change in time or be different in climate models with respect to observations. For instance, Rossby wave sources can change in location during different SSWs (e.g. Shi et al., 2020), leading to variations in the location of the polar vortex, PNJ or their features. Conversely, the location of the polar vortex and PNJ also modulates the propagation of Rossby and gravity waves into the stratosphere, which in turn affects the polar vortex. However, all these changes in the latitude of the PNJ would not be fully captured by using the 60° N standard algorithm. Indeed, it may provide misleading results in some cases. For instance, the polar vortex can sometimes be slightly shifted out of the pole and so, the PNJ would separate from being circumpolar and its core would move away of 60° N. However, all this does not always imply that the PNJ intensity has drastically changed, but the standard method would probably identify it as a weak PNJ event.

Some of the disadvantages of the zonal mean perspective can be avoided with the two-dimensional (2D) analysis of the vortex edge. There are different 2D techniques. Some of them are related to the distribution of chemical tracers considering that the vortex air is an isolated air mass (e.g McDonald and Smith, 2013; Krützmann et al., 2008). Others are based on Lagrangian methods following tracers trajectories and transport barriers (Serra et al., 2017; Smith and McDonald, 2014). Another group of 2D diagnostics of vortex characteristics and edge are derived from the PV field (e.g Waugh and Randel, 1999; Waugh, 1997). The vortex shape and variability can be characterized based on vortex moment analysis (Waugh and Randel, 1999; Mitchell et al., 2011) and the edge is identified by seeking for steep PV gradients (Nash et al., 1996). Although all these techniques provide a more accurate description of the vortex edge, the procedure is much more complex than the 60° N standard methodology. In most cases, they involve several computations such as coordinates changes, 2D integrations, trajectory calculations, which imply a high computational time, particularly in large databases. In addition, some of the required variables such as tracers concentrations or PV are not a common output of climate models or reanalysis data and should be first derived from other fields. In this sense, more recently, the 2D momentum method has been modified to be applied to a more common variable such as geopotential height at 10hPa (Seviour et al., 2013; Rao et al., 2021). However, this method still requires other long computations. Further, a time mean value of the zonal mean geopotential height at a fixed latitude (60° N) is needed to identify the vortex area and assimilate to the vortex edge. However, as happens with the standard method, the location of this edge might be different in different scenarios or datasets, and these changes are not included in the diagnostics, leading to inaccurate results in polar vortex characterization.

In this paper, we investigate the application of machine learning techniques for the full characterization of the PNJ as a way to simplify this characterization. Machine learning techniques only recently have started to be applied at larger scales, instead of just very specific studies in climate science. Lately, interest has begun to increase in applying these techniques to climate change analysis (Sebestyén et al., 2021).

Convolutional neural networks (CNNs) are some of the most common algorithms in the field of climate machine learning, and have been applied in tropospheric studies. A study has been able to predict the monthly mean temperature on Earth with an accuracy of 97% for one climate model (CRU TS 4.01, Ise and Oba (2019)), by using data over 30 years to predict the subsequent 10 years with a CNN. Another algorithm using CNNs is presented by Chattopadhyay et al. (2020), where an automatic labeling algorithm based on K-means clustering was designed in order to obtain the training dataset, leading to 94% accuracy in predicting North-American weather patterns. However, for these studies, there is a lack of expert-classified data. Therefore, training and testing is usually performed with autolabeling and historical data. This is due to the complexity and high number of datasets, which would escalate the amount of human effort for classification.

In stratospheric climate science, some machine learning algorithm studies have focused on stratospheric ozone depletion (Nowack et al., 2018; Sedona et al., 2020; Keeble et al., 2021). A study has been made on polar stratospheric clouds, which play a key role in this topic. Techniques like random forest and support vector machines were found to be suitable to classify the composition of polar stratospheric clouds, training with infrared spectra data (Sedona et al., 2020). Another study has predicted stratospheric column ozone, using Ridge regression, Lasso regression, random forests and extra trees techniques to train and classify the output of climate model simulations under different future climate scenarios, using greenhouse gases emissions as classification features (Keeble et al., 2021). Regarding geometric identification of climate structures in the stratosphere, computer vision techniques have been applied by Lawrence and Manney (2018) in order to characterize the stratospheric polar vortex based on PV reanalysis data, and, more specifically, to detect onset dates for split vortex events.

Their algorithm involves spatial overlap and some adaptive–predictive techniques, while our method will focus on training a machine learning classifier. The stratosphere is an area with a lot of potential to be explored by means of machine learning techniques and this study is going to focus on that area.

The aim of this work is to present a new technique to detect and characterize the PNJ (intensity and location) by using artificial intelligence (AI) that can overcome some of the issues of the existing methodologies and spans several climate models and reanalysis data. The intention is that, by merely using a single simple field such as the zonal wind (u), more information about the PNJ can be obtained than the intensity provided by applying the 60° N standard methodology described in the previous paragraph. This will be also done without imposing assumptions based on observations such as the 60° N latitude and with simple fields (u) that are even common output of model simulations. This also results in less complexity and computational time than 2D moment analysis. More specifically, the method we will use is the development of a machine learning classifier that seeks for the grid points that form an area of high u values. A machine learning classifier is a black-box model to the end-user (Kotsiantis et al., 2006). In this case, the input variables to the black-box model will be climate variables (u or at the most, geopotential height too) given on a grid and over time, obtained from available climate simulations. The outputs of this function will be the PNJ region and extension, averaged values over the region, and whether the polar night jet on each time step is under extreme event conditions or not. Speaking in terms of machine learning, the input variables to the function are called features, and the outputs are called class labels. The goal of the machine learning process is to model the outputs or class labels, given a distribution of features curated, initially, by a human, i.e. the training process (Muller et al., 2021). The resulting machine learning model, once trained, will be able to function in an unsupervised way by assigning labels based on features. The final product of this process, and this study, is a trained classifier, and all the end-user needs to do in order to obtain the PNJ extension and characterization features is to prepare the required inputs, as climate variables of the models to be classified. Apart from the mere description of the technique, we compare its performance with that of the standard method to first validate our algorithm and secondly, highlight the advantages of our technique. Further, we give evidence of the usefulness of our method by presenting different applications such as the analysis of the PNJ mean state and its variability in different datasets and under different climate conditions.

This article is divided into the following sections. The details of the different datasets together with the explanation of climate images used in the study are indicated in Section 2. Section 3 describes the methodology and artificial intelligence techniques applied to our scientific challenge. Section 4 describes the results obtained. Finally, Section 5 details the main conclusions.

2. Data

In this study, datasets of two different kinds, i.e. reanalysis and climate model simulations, are used.

2.1. Reanalysis

Japanese 55-year reanalysis (JRA-55) data (Kobayashi et al., 2015) are considered in this study as observations. Though treated as observations, reanalyses are a combination of short-range weather forecasts with observational data. They provide a global dataset that is homogeneous in time for the latest several decades (Kalnay et al., 1996). The JRA-55 dataset includes operational and delayed observations, digitized observations, as well as processed satellite data, in order to produce an homogeneous dataset that represents recent decades most realistically. The horizontal resolution of this dataset is 2.5° lon. x 2.5° lat. and our period of study extends from 1958 to 2015.

Table 1
List of climate models used in the analysis, indicating the model resolution and the number of years included in each simulation.

Model name	Model resolution	piControl	abrupt4xCO2
CanESM5	T63L47 top 1 hPa	451 years	151 years
CESM2-WACCM	1.25°× 0.94° L70, top 150 km	499 years	150 years
IPSL-CM6A-LR	N96, top 80 km	1200 years	900 years

2.2. Climate models

We also employ the output of simulations performed with state-of-the-art climate models participating in the recent Coupled Model Intercomparison Project, Phase 6 (CMIP6) (Eyring et al., 2016). These models provide a detailed representation of the climate system, including as many of Earth system processes as possible. In our analysis, we will focus on the three models shown in Table 1. These three models have their model top in the stratopause or above, which means that they can have a good representation of at least the lower and middle stratosphere. This is important for our analysis as it is focused on the middle stratosphere. Other recent studies focusing their analysis in the mid-stratosphere are (Martineau et al., 2018; Domeisen et al., 2019; Chun et al., 2019; Vargin et al., 2020). The choice of these specific models was done following results of previous work (e.g. Liu et al., 2019; Ayarzagüena et al., 2020; Rao and Garfinkel, 2021) that analyzed the PNJ state based on the standard method. For instance, the CESM2-WACCM (Danabasoglu, 2019; Gettelman et al., 2019) was selected because of the model's similarities with observational data, especially when simulating frequency of SSWs and mean PNJ intensity (Liu et al., 2019). CanESM5 (Swart et al., 2019a,b) and IPSL-CM6A-LR (Boucher et al., 2018) were also chosen because of the different PNJ state under present climate conditions (i.e. too strong for CanESM5 and too weak for IPSL-CM6A-LR with respect to observations) and the opposite response of the polar vortex to an extreme increase in CO2 concentrations according to Ayarzagüena et al. (2020). This previous knowledge will allow for a validation of our algorithm and a better comparison of results with those derived from the standard method for PNJ characterization.

In terms of simulations, we use the output of only one ensemble member of the piControl and abrupt4xCO2 runs of the mentioned models. The piControl run consists of an evaluation simulation with prescribed preindustrial CO2 concentrations (Taylor et al., 2012). It extends several centuries, allowing the characterization of the internal stratospheric variability. The abrupt4xCO2 simulation is identical to piControl one, except for the CO2 concentrations that are instantaneously quadrupled from pre-industrial levels at the beginning of the simulation, and held fixed during the rest of the simulation. This simulation was firstly performed for climate sensitivity and feedback analysis to an abrupt and extreme increase in CO2 concentrations (Taylor et al., 2012).

2.3. Climate images

We will analyze monthly mean zonal wind (u) data at 10 hPa in winter months (December–January–February) of the previously described databases in all sections except for Section 4.6 where daily data of u and geopotential height at the same level are used. The reason for choosing these data is that the PNJ is characterized by high values of u in the extratropics of the winter hemisphere and the 10hPa level corresponds to the middle stratosphere (National Weather Service, 2021), where the PNJ is typically studied. Indeed, the standard method for characterizing the PNJ is based on the zonal mean zonal wind speed at 10 hPa (e.g Charlton and Polvani, 2007; Hardiman et al., 2019), as previously indicated in Section 1. The datasets described in Sections 2.1 and 2.2 are originally available in netCDF (Network Common Data Form)

format, which is a self-explained binary format for atmospheric data. For the purpose of this paper, they are converted into 2D images in a latitude–longitude grid.

A sample image is shown in Fig. 1a, which shows, in the x axis, the longitude coordinates, and on the y axis, the latitude coordinates. In this image, we can see that the Northern Hemisphere contains a region of high zonal wind speeds, and it is where we will try to locate the PNJ. The training and testing methods described in Section 3 are applied to a pool of 200 images in total, containing: 50 JRA-55 images, 50 CanESM5 piControl images, 50 CESM2-WACCM piControl images, and 50 IPSL-CM6A-LR piControl images, obtained from the first 50 winter months in each of the experiments. Since each dataset has a different horizontal resolution, each of these images has been mapped to the CESM2-WACCM grid (1.25° x 0.94°), which consists of a square grid of 192 latitude points and 288 longitude points. Final classification will be performed on the full datasets.

3. Methodology

This section describes the methodology based on artificial intelligence algorithms applied in this work.

3.1. Region growing

First, as a previous step to expert classification, a 8-connected region growing algorithm (Efford, 2000) was applied on each of the images representing the monthly mean u data at 10 hPa. This algorithm consists on comparing one pixel to its eight neighboring pixels, and appending those pixels to the region if they meet some condition. Its steps are described in detail below.

Algorithm 1 8-connected region growing algorithm

1. Choose a seed pixel. In this case, the seed pixel is the grid point with maximum u.
2. Check the 8 neighboring pixels and add them to the region if they are inside the growth boundary;
3. Repeat the previous step for every one of the added pixels, until no more can be added to the region.

In region growing algorithms, it is either 4-connected or 8-connected regions that are typically used in image segmentation literature (He et al., 2019). The reason for choosing one or the other depends on the specific use case's shape and wanted speed, as described in Wu et al. (2008). In our case, an 8-connected region makes the final shape slightly closer to a continuous boundary, and also provides a faster region growth than four neighbors.

The growth boundary is established in an empirical way as 75% of the maximum zonal wind speed. The region growing algorithm is done to facilitate the work in Section 3.2, and is only used as a starting point to build the ground truth. It is not used in any way by the final classifier. The region growing results obtained for one sample month are shown in Fig. 1b with the corresponding contour. It is observed that the segmentation algorithm has selected the region of highest u without discontinuities.

3.2. Ground truth: expert labeling

As a second step before classifier training, expert modification of the images was performed on the results provided by the region growing method in order to add or delete any area that was not consistent with the expert's vision. This work was performed by a climate expert, a researcher who manually inspects each of the available region growing results and modifies them according to their experience. The expert selected these regions on 50x4 images equally distributed in JRA-55, CanESM5 piControl CESM2-WACCM piControl, and IPSL-CM6A-LR piControl, as previously mentioned. More importantly, the expert also discarded any images which, in their view, contained zonal wind regions that are not consistent with the PNJ structure. An example of discarded image is shown in Fig. 1d, since the maximum u and region detected by the region growing algorithm is located at very low latitudes, which is not compatible to an actual PNJ state.

Fig. 1b shows an example of the results of the region growing algorithm and the final expert-modified PNJ field, which are identified over the zonal wind data field. In this case, the expert has only minimally modified the region produced by the region growing method. The expert contour on Fig. 1b for this image and equivalent ones are the ones provided to the classification algorithm described later in this paper.

The images that the climate expert labeled make up the ground truth dataset, which is considered the most objective dataset in the scope of this study (Muller et al., 2021). A ground truth dataset consists of correctly labeled data, which is a necessary requirement for training machine learning models (Zhou et al., 2018). It is against this ground truth data that we will measure how accurate the classification algorithms are.

3.3. Classification

A binary classification, with the goal of detecting the PNJ region, is performed for each point in the longitude–latitude grid in order to obtain which of the grid points belongs to the PNJ region (labeled as 1), and which one does not (labeled as 0). The classifier will be specifically trained to label each grid point based on the ground truth developed in Section 3.2, and will work independently of any fixed growth boundary, which is not possible when only applying region growing techniques (Section 3.1).

In order to achieve an efficient and effective machine learning algorithm, two predictors as features of the classifier are selected. First, the normalized u at each grid point, which we will refer to as x_1 in this section, and secondly, the difference in latitude of each grid point with respect to the latitude of maximum u, which will be referred to as x_2. Relative features have been selected because of possible variability between climate databases and over time.

Regarding the type of classifier, we decided to try four: first, the Gaussian Naive Bayes classifier because of its simplicity and fast performance; then, we tried random forest and decision trees classifiers because they are among those used in recent stratospheric climate science machine learning literature, as mentioned in Section 1; lastly, we tested the k-Nearest Neighbor classifier, since it is one of the most accurate statistical classifiers and is commonly used in image segmentation (Bieniecki and Grabowski, 2004; Rajini et al., 2011). All these algorithms are available in the scikit-learn Python library (Pedregosa et al., 2011). Each of them is described below.

Deep learning algorithms were not applied because of the big training dataset size that would be needed, ranging from thousands to millions of images (Sharma et al., 2020).

The Gaussian Naive Bayes classifier (Salmi and Rustam, 2019) is based on the Bayes theorem and assumes strong independence between the predictors. The probability distribution for each feature is assumed to be Gaussian. The algorithm picks the output y with maximum probability for all possible values of the output (in our case, 0 and 1) and all possible values of predictors x_1 and x_2: $y = argmax_y P(y)P(x_1|y)P(x_2|y)$. By calculating the maximum conditional probability out of $P(x_1|1)P(x_2|1)$ and $P(x_1|0)P(x_2|0)$, we would obtain whether the grid point is or is not part of the PNJ region.

Decision trees (Daniya et al., 2020) are a rule based classification approach which split the dataset into branches, until smaller datasets belonging to each class (PNJ or no PNJ) have been isolated from each other. Scikit-learn Python library uses the CART (Classification and regression trees algorithm) by default for decision trees. This algorithm splits the dataset subsequently into two child nodes. For example,

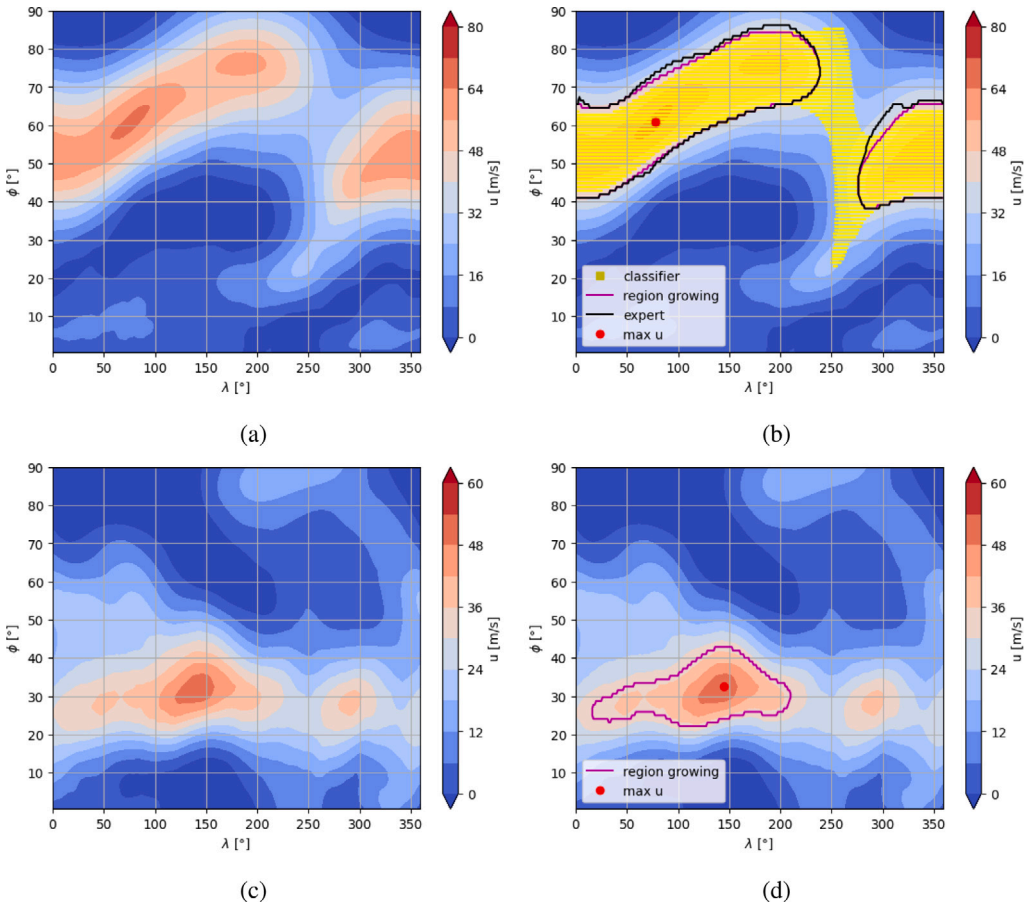

Fig. 1. **(a)** Zonal wind u at 10 hPa for one sample winter month of the CESM2-WACCM piControl simulation. Contour interval 8 m/s. **(b)** Same as (a) but including the point of maximum zonal wind speed in red, the polar night jet (PNJ) contour approximated by the region growing algorithm in magenta, the PNJ expert hand-drawn contour for the ground truth dataset in black, and the PNJ region as detected by a decision trees classifier in yellow. **(c)** Zonal wind u at 10 hPa for one sample winter month of the JRA-55 simulation. Contour interval 8 m/s. **(d)** Same as (a) but including the point of maximum zonal wind speed in red, and the polar night jet (PNJ) contour approximated by the region growing algorithm in magenta. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hypothetically, it would split the dataset into two categories, $x_1 \leq value$ and $x_1 > value$, and continue splitting the tree from there. The CART algorithm uses the Gini index as a quality measure of the purity of each potential split. The equation for the Gini index in this case would be $Gini = 1 - p_0^2 - p_1^2$, being p_1 the probability of picking a data node from the PNJ region, and p_0 the probability of picking a data node which does not belong to the PNJ. The split into child nodes would be made according to the smallest Gini index. In our use case, after several tests, we have set the maximum depth of the tree to 5.

Random forests (Fawagreh et al., 2014) consist of averaging the predictions of several decision trees. If the majority of decision trees that form the forest agree that a grid point is PNJ, the data point will be labeled as such, and vice versa. Random forests use the bootstrap aggregation method in order to reduce overfitting. This method provides each of the decision trees in the ensemble equal vote toward the final labeling result, as well as drawing random samples of the dataset for training individual trees. In our use case, after several tests, we have set the maximum depth of the trees to 5, and the number of decision trees to 10.

The k-Nearest Neighbors algorithm (kNN) (Alfeilat et al., 2019) classifies a test sample based on what the majority of k closest data points have been classified as. In this case, the distance to determine those k neighbors is the Euclidean distance, one of the most commonly used ones for kNN, and k has been set, after some tests, to 3. The Euclidean distance is calculated as the linear distance two points a and b with their respective coordinates: $p = [x_1, x_2]$: $d = \sqrt{(p_a - p_b)^2}$. Regarding our use case, if 2 out of the 3 closest points are classified as PNJ, the study point will also be tagged as PNJ. One of the disadvantages of the kNN algorithm is its high computational time (Trisal and Kaul, 2019).

4. Results

4.1. Classification

The classification algorithms described in Section 3.3 have been applied to the ground truth dataset. The accuracy with respect to the ground truth and the elapsed training and classification times have been checked. In order to do this, the amount of images available is divided into hold-out subsets consisting of a 70/30 train-test ratio (Awwalu and Ogwueleka, 2019). The accuracy results and elapsed time for this analysis are shown in Table 2. The elapsed time corresponds to running the algorithms in the environment described in the Code availability section. The accuracy has been obtained by comparing predicted images with ground truth images.

Given the results in Table 2, the decision tree classifier was selected going forward as a compromise between accuracy and training and testing computational complexity. Decision trees classifiers also have the advantage of being easier to interpret than random forests (Meng et al., 2020). The full expert dataset will from now on be used to retrain the algorithm and classify the rest of the images in the databases.

The classification algorithm results for one sample month are shown in Fig. 1b. The figure shows, in yellow, the PNJ region identified by our classifier, on one of the images fed to the classifier as training data,

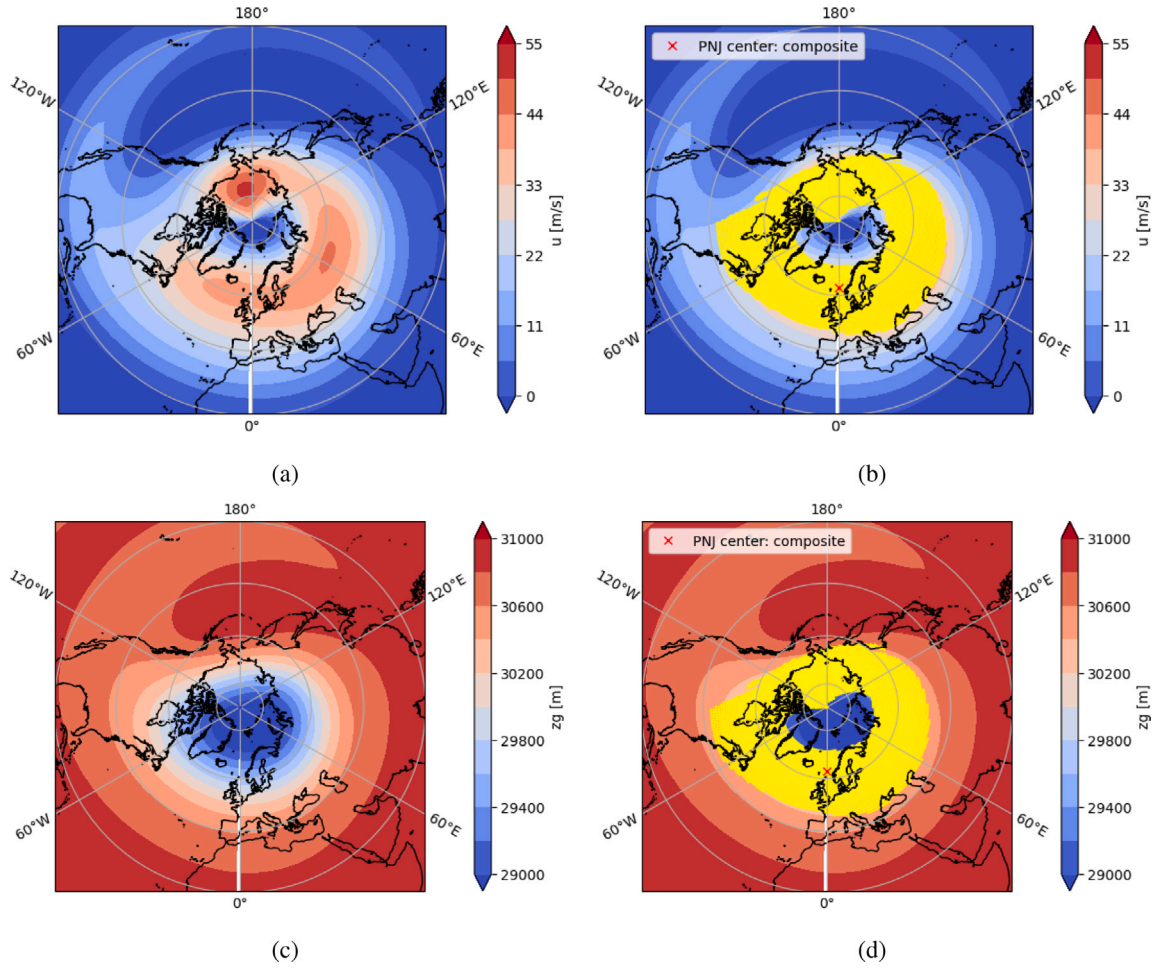

Fig. 2. **(a)** Composite map of u in JRA-55 dataset for PNJ events. Contour interval 5.5 m/s. North-polar stereographic projection. **(b)** Same as (a) but showing, in yellow, the PNJ region as detected by a decision trees classifier, and in red, the point representing the weighted averaged latitude and longitude of the yellow region. **(c)** Composite map of geopotential height z_g in JRA-55 dataset for PNJ events. Contour interval 200 m. North-polar stereographic projection. **(d)** Same as (b) but plotted over the composite map of z_g in JRA-55 dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison of classification algorithms for piControl CESM2-WACCM 50-image dataset: hold-outs method, including accuracy, training time and prediction time.

Results	Accuracy [%]	Training time [s]	Prediction time [s]
Naïve Bayes	94.1	0.95	0.3
Decision tree	95.5	8.3	0.1
Random Forest	95.5	42.5	1.7
kNN	94.8	129.9	56.2

as part of the ground truth dataset. The classified region is close to the expert contour, which is over the high u region on the Northern Hemisphere, the PNJ, visually validating the high accuracy of the tested classifiers.

After training and testing, the classifier is applied to each remaining image in the climate databases. As an additional validation of the train-test process, we have compared the identified PNJ region with the u and geopotential field. To do so, we have used averaged fields along the time period of study because of two reasons (Fig. 2). First, time-averaged fields summarize information and can be used instead of single month fields. In addition, these fields are smoother than that of a single sample month. In these figures we observe that the PNJ is located over the strongest u, which also coincides with the edge of a big geopotential (z_g) drop, confirming the high skill of our algorithm to identify the PNJ.

4.2. Characterization of climatological PNJ

Once classification of each latitude–longitude grid point has been finished, the climatology of several variables has been obtained to characterize the PNJ in each dataset. In this first part, we analyze JRA-55 and piControl simulations. Firstly, we define three magnitudes averaged over the PNJ region: intensity (u_c) and latitude (ϕ_c). Intensity consists of u weighted with the latitude cosine at each PNJ point. Latitude is weighted with u at each PNJ point. Secondly, these results are averaged over the time contained in each climate dataset, thus obtaining their climatology, named $\overline{u_c}$ and $\overline{\phi_c}$. Obtaining these two magnitudes implies that the PNJ is perfectly described in extension, intensity, and latitudinal position, and this will be useful for comparisons between climate models and with the observational dataset. Apart from the relevance of the characterization of the PNJ itself, these magnitudes can also provide us a relevant information about the climatological state of the polar vortex. As previously mentioned, the intensity of the PNJ is intimately related to that of the vortex. ϕ_c denotes the location of the edge of the vortex and so, gives a hint of the vortex extension and its position either pole-centered or shifted out of the pole, when considering together with the PNJ intensity.

The characterization results are shown in Table 3. They show that the obtained PNJ latitude is very close to 60° N in all datasets, suggesting once again the good performance of our algorithm and the consistency of using the standard 60° N method to calculate the strongest u at the edge of the polar vortex. Across datasets, only small

Table 3
Characterization magnitudes of the polar night jet (PNJ): climatology of PNJ intensity ($\overline{u_c}$), climatology of weighted latitude ($\overline{\phi_c}$), and respective standard deviations. Values obtained from classifying with the decision trees model, for JRA-55 and piControl simulations of CESM2-WACCM, CanESM5 and IPSL-CM6A-LR models.

Model	Simulation	$\overline{u_c}$ [m/s]	std(u_c) [m/s]	$\overline{\phi_c}$ [°]	std(ϕ_c) [°]
JRA-55	JRA-55	38.59	10.95	62.86	6.75
CESM2-WACCM	piControl	34.72	6.32	61.60	6.49
CanESM5	piControl	39.01	7.16	63.00	6.31
IPSL-CM6A-LR	piControl	34.88	7.24	59.54	6.20

differences in the PNJ intensity are observed. JRA-55 and CanESM5 have the strongest $\overline{u_c}$ compared to CESM2-WACCM and IPSL-CM6A-LR models. In terms of PNJ intensity and latitude, CanESM5 is the closest climate model to observations, and both higher intensity and latitude are obtained in these two cases as compared to the remaining climate models.

4.3. Characterization of monthly PNJ variability in climate databases

As indicated in Section 1, the boreal polar vortex and so, the PNJ shows a large variability, ranging from a very strong polar vortex to extreme weak polar vortex events. In the latter cases, the PNJ might not exist or is highly perturbed and so, our algorithm has discarded those images from the analysis in Section 4.2. We have divided these events into two types, with conditions set manually after iterating.

The first kind of events take place when the average obtained latitude is below 35° N, taken as the subtropical range limit (Ping et al., 2001). These events will be referred to from now on as type 1 events, and are in some way similar to displaced polar vortex events as the jet is out of its typical location at 10 hPa. Thresholding techniques like this one are common in machine learning both in image segmentation preprocessing and also for classification in multiple steps (e.g. Sheeren et al., 2009). Composites of u and z_g for these events are shown in Fig. 3, where the good performance of our algorithm to detect the strong winds delimiting the polar vortex (region of low values of geopotential height at high latitudes) is confirmed, even in special cases. Further, during type 1 events, the intensity of the PNJ is weaker than the average of that of the climatological PNJ of Section 4.2. Consistently, the polar vortex is also weak and displaced out of the pole (compare panels c and d of Figs. 2 and 3. The displacement of the polar vortex towards Eurasia is typically associated with an intensification of the Aleutian high and an enhancement of wave activity, that would also weaken the polar vortex (Labitzke, 1977). The frequency of these extreme events is evidently low (roughly 1%), being higher for the JRA-55 dataset (3%) than for climate models (Table 4). Please note that these events cannot be directly identified as SSWs as we are using monthly mean data. Nevertheless, type 1 events correspond to months when a persistent SSW has taken place in its surroundings.

We have also observed a second type of weak events taking place over 35° N latitudes, which also show a very small classified area (type 2 events). These events have initially and tentatively been tagged as such when the intensity of the PNJ is as low as $u_c < \overline{u_c} - 2std(u_c)$, and the total number of PNJ grid points is below 10% of the total amount of image grid points. These events are described in more detail in Section 4.6.

4.4. Comparison of results with standard characterization method

Results obtained from the 60° N characterization method have been included in Table 5 for comparison with the classifier results. In general, $\overline{u}_c$ results for PNJ events are lower, and the standard deviation is higher with respect to the classification results shown in Tables 3 and 6. These differences are statistically significant at the 95% confidence level. The reason is that in many cases the core of the PNJ (the strongest

Table 4
Frequency of polar night jet (PNJ) and type 1 events for JRA-55 and climate models CESM2-WACCM, CanESM5 and IPSL-CM6A-LR for piControl simulation. Averaged value of zonal wind speed u is also provided.

Dataset		PNJ		Type 1	
Model	Simulation	f [%]	$\overline{u_c}$ [m/s]	f [%]	$\overline{u_c}$ [m/s]
JRA-55	JRA-55	96.55	38.59	3.45	17.57
CESM2-WACCM	piControl	98.86	34.72	0.73	20.95
CanESM5	piControl	99.33	39.01	0.30	17.17
IPSL-CM6A-LR	piControl	98.06	34.88	1.39	15.86

Table 5
Climatology of zonal wind speed (u) obtained with 60° N parallel method ($[u_{60^\circ N}]$). Calculations are performed for each climate database JRA-55, models CESM2-WACCM, CanESM5 and IPSL-CM6A-LR for piControl simulation.

Model	Simulation	$\overline{[u_{60^\circ N}]}$ [m/s]	std($[u_{60^\circ N}]$) [m/s]
JRA-55	JRA-55	31.11	13.50
CESM2-WACCM	piControl	28.09	8.81
CanESM5	piControl	32.68	11.58
IPSL-CM6A-LR	piControl	25.92	10.99

u) might be displaced from the parallel and so, the standard method is not capturing, when taking the climatology at parallel 60° N. Similarly, by averaging at 60° N, we might be including points which do not correspond to the PNJ region, since the fact that the PNJ is characterized by a nearly zonal westerly circulation is an approximation. The most intense polar vortex corresponds, once again, to the JRA-55 dataset and CanESM5 climate model, versus CESM2-WACCM and IPSL-CM6A-LR.

Fig. 4 shows all events obtained with standard and classification methods for JRA-55, marked with circle and "x" markers, respectively. Furthermore, type 1 events as tagged by the classifier have been labeled in red, versus blue color for PNJ events. We observe that the classifier captures a higher $[u]$ than the standard method in cases of PNJ latitude south of 60°N. Thus, the aforementioned fact that PNJ intensity is lower by applying the 60°N characterization method is explained. This is due to the maximum intensity of the PNJ being located in a different latitude than 60°N, information that can be seen on the x-axis of Fig. 4. This result is particularly true and theoretically expected for red points, but it is also remarkable for events of PNJ latitude not that far from 60°N. Indeed, while the classifier results of $[u]$ overlap or are very similar to those of the standard method for episodes of PNJ latitude between 57.5°N and 67.5°N, there are events of PNJ latitude around 50-55 °N that would have a very weak intensity based on the standard method but would show a PNJ intensity similar to 60°N events from the classifier. This means that our classifier method is able to offer more information when the PNJ moves away from 60°N, such as precise location and a more accurate value of the zonal wind speed, while preserving the ability to label type 1 events. This valuable information is lost when using the standard characterization method.

4.5. Carbon coupling: analyzing the future

Under increasing CO2 concentrations, the wind field at 10 hPa in extratropics (including 60° N) might be affected by the changes in tropical wind structures at lower levels (e.g. Bunzel and Schmidt, 2013; Oberländer et al., 2013). These changes would not represent thus the changes in PNJ. Other studies such as Zhang et al. (2016) and Zhang et al. (2018) have documented a shift of the polar vortex towards Eurasia under climate change conditions. In all these cases, the analysis of future PNJ changes from a zonal mean perspective, and more specifically based on zonal mean u at 60° N and 10 hPa, might be misleading. Thus, the same characterization exercise with our AI technique has been applied to the abrupt4xCO2 simulations, in order to quantify the effect of carbon emissions on the PNJ intensity and

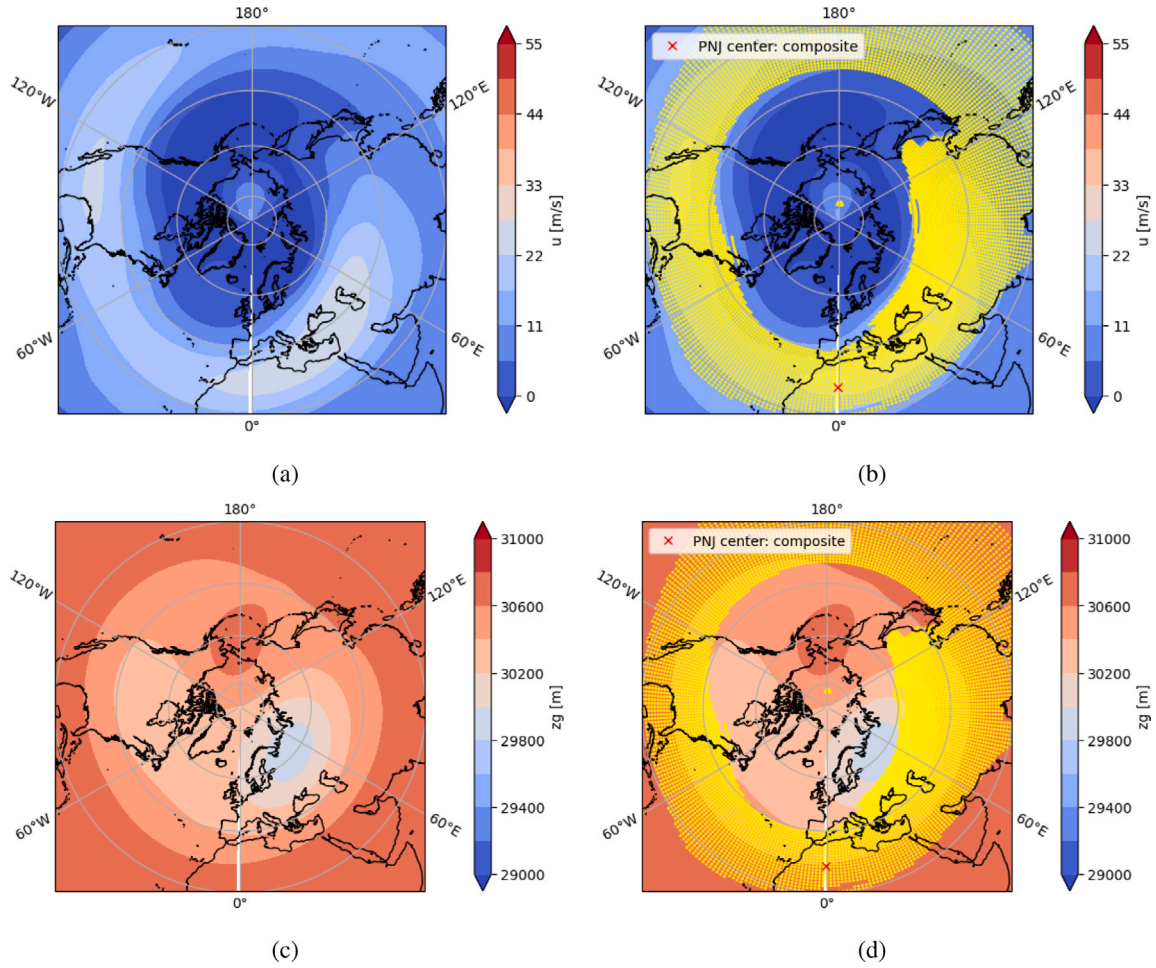

Fig. 3. **(a)** Composite map of zonal wind speed u in JRA-55 dataset for type 1 events. Contour interval 5.5 m/s. North-polar stereographic projection. **(b)** Same as (a) but showing, in yellow, the PNJ region as detected by a decision trees classifier, and in red, the point representing the weighted averaged latitude and longitude of the yellow region. **(c)** Composite map of geopotential height z_g in JRA-55 dataset for type 1 events. Contour interval 200 m. North-polar stereographic projection. **(d)** Same as (b) but drawn over the composite map of z_g in JRA-55 dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

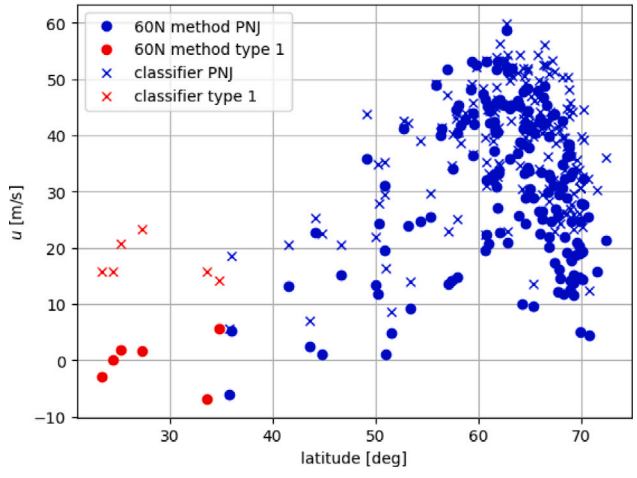

Fig. 4. Zonal wind speed from standard method $[u_{60^\circ N}]$, plotted with circle markers, and the classifier results for u_c, plotted with "x" markers, versus the averaged latitude (grid increment 10 degrees) obtained with the classifier, for each monthly occurrence in the JRA-55 dataset. The plot also shows a color division between PNJ and type 1 events, marked blue and red respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

latitude. A Student t-test has been performed on the results to check for statistically significant changes, on variables u_c and ϕ_c. It has been concluded that the changes in these variables between Tables 3 and 6 are significant at the 95% confidence level.

Results point to the lack of consensus about climate models with respect to climate change that was expressed in Section 1. The PNJ in CESM2-WACCM model in abrupt4xCO2 becomes weaker than for piControl (Table 7), and presents a higher frequency of type 1 events than in piControl. In contrast, in CanESM5 and IPSL-CM6A-LR the PNJ becomes more intense with respect to piControl, and a lower frequency of type 1 events (i.e. less disturbed) by extreme CO2 concentrations. These inconsistencies of PNJ response to climate change for CESM2-WACCM versus CanESM5 and IPSL-CM6A-LR are coherent with what is described by Ayarzagüena et al. (2020), which presents the mean SSW frequency comparison between piControl and abrupt4xCO2 simulations for the climate models in this study, among others.

Regarding the location of the PNJ core, latitude in Table 6 is consistently lower than its respective piControl values. Within abrupt4xCO2 datasets, latitude is still the highest for CanESM5, which is also the most intense PNJ, maintaining the piControl tendency in this regard. This might imply a southward shift of the PNJ that would be consistent with the vortex movement out of the pole described by Zhang et al. (2016). Interestingly, the trend is detected in models showing either a weakening and a strengthening of the PNJ.

Thus, conclusions derived from our method on the impact of increasing CO2 concentrations on the PNJ agree with those obtained with

Table 6
Same as Table 3 but for JRA-55, models CESM2-WACCM, CanESM5 and IPSL-CM6A-LR for abrupt4xCO2 simulation databases.

Model	Simulation	$\overline{u_c}$ [m/s]	std(u_c) [m/s]	$\overline{\phi_c}$ [°]	std(ϕ_c) [°]
CESM2-WACCM	abrupt4xCO2	32.84	6.37	59.29	7.96
CanESM5	abrupt4xCO2	46.05	6.46	61.41	3.54
IPSL-CM6A-LR	abrupt4xCO2	43.91	8.81	58.62	4.45

Table 7
Frequency of type 1 events for JRA-55 and climate models CESM2-WACCM, CanESM5 and IPSL-CM6A-LR for abrupt4xCO2 simulation. Averaged value of zonal wind speed u is also provided.

Model	Simulation	f [%]	$\overline{u_c}$ [m/s]
CESM2-WACCM	abrupt4xCO2	1.11	20.61
IPSL-CM6A-LR	abrupt4xCO2	0.07	18.21

other methodologies, but we are able to capture all these aspects at once.

4.6. Identification and classification of SSWs

So far, our algorithm has been applied to analyze the PNJ on a monthly time scale. However, a large part of the variance of the polar vortex is concentrated on a daily time scale. The most important instances of this variability are the SSWs as mentioned in Section 1. In these cases, the PNJ is largely perturbed and weak. In addition, the polar vortex is either shifted out of the pole (displacement SSWs) or split into two pieces (split SSWs). Consequently, the PNJ would also move southward similarly to what happens in type 1 events detected in previous sections or split in several pieces (mainly two) in line with the type 2 events anticipated in Section 4.3. We have thus applied our classifier to daily data of JRA-55 to investigate if our algorithm is able to both identify split and displacement SSWs. We have only done it for reanalysis data so that we can compare our results with those detected in literature by using other methodology.

When applying the classifier trained with only monthly data to daily JRA-55 data in DJF, we are able to detect a high number of the displacement events identified by other methodologies in the literature (Gerber and Martineau, 2022). Indeed, we retrieve more than 70% (6 out of 9) of the unanimously classified as displacement by different schemes based on PV field, geometric moment diagnostics of geopotential height or absolute vorticity as shown by Gerber and Martineau (2022) and 11 out of 16 events classified as displacement SSWs by at least two of these methods. The onset dates of these events also match the first dates of the reversal of the wind with a difference of less than 6 days.

As for split events, these are associated with type 2 events mentioned in Section 4.3, which, at that time, were pulled out of the main study for further investigation. These events are not common in monthly data in reanalysis and far infrequent in model simulations. As a result, the pool of data with a PNJ structure of type 2 is not enough to train the classifier. In order to correctly identify type 2 events, we have added additional training data to our classifier. These data are composed of 84 additional images from the JRA-55 daily simulation, containing extreme events. A new classification feature has also been implemented, the geopotential height. With these modifications, we have been able to detect type 2 events in daily JRA-55 data. We have also compared the initial dates of our daily results with Charlton and Polvani (2007) and Lawrence and Manney (2018), in Table 8, which shows that our algorithm obtains 14 out of 17 split events (82%) of those other two studies combined. The splitting of the PNJ sequence is shown for 1978–1979 winter in Fig. 6, in which two regions of interest are identified at the end days of February, associated with minimum geopotential height points. Lastly, the detected type 2 event regions on monthly JRA-55 data are shown on Fig. 5(d), over the minimum geopotential height.

Table 8
Comparison of our classifier's dates for type 2 event detection with Charlton and Polvani (2007)'s algorithm (period 1958–1999), and also with results in Lawrence and Manney (2018) (period 1982–2013). Regarding our classifier's results, only onset dates have been included in this table, selected from the full result.

Our classifier	Charlton and Polvani (2007)	Lawrence and Manney (2018)
30 Jan 1958	31 Jan 1958	
17 Feb 1963	28 Jan 1963	
27 Feb 1966	23 Feb 1966	
08 Jan 1968	7 Jan 1968	
22 Jan 1971	17 Jan 1971	
06 Feb 1973	31 Jan 1973	
	9 Jan 1977	
21 Feb 1979	22 Feb 1979	
30 Dec 1984	1 Jan 1985	30 Dec 1984
20 Jan 1985		20 Jan 1985
	7 Dec 1987	16 Dec 1987
19 Feb 1989	21 Feb 1989	19 Feb 1989
20 Dec 1998	15 Dec 1998	
	25 Feb 1999	
18 Feb 2003		20 Jan 2003
26 Jan 2009		28 Jan 2009
		11 Feb 2010
08 Jan 2013		8 Jan 2013

5. Conclusions

In this study, we introduce a method for the characterization of the polar night jet (PNJ) that can be applied in both observations and climate models and under different climate conditions. This characterization is commonly performed based on the intensity of the zonal mean zonal wind field at 10 hPa and a fixed latitude (60° N), but this standard method presents some issues. It does not allow for a complete description of the PNJ state regarding its exact location and can even produce misleading results when the core of the PNJ is shifted away from 60° N parallel. On the other hand, the existing 2D methods for the analysis of the polar vortex and its edge usually require several variables that in many cases involve fields such as potential vorticity that are not direct output from model simulations. In addition, these 2D techniques imply several calculations and so, long computational time especially when working with long datasets. The method presented in this study seeks for achieving these challenges and is based on artificial intelligence. More specifically, we have trained a classifier with expert-labeled stratospheric monthly averaged u. The kind of ground truth we have built was not previously available, due to the high human effort that is required and the relatively recent introduction of machine learning to the field of climate science. The ground truth dataset binary-classified images did not previously exist in any public database, and were treated by the expert specifically for this study. In order to reduce the amount of human effort, we have helped the expert classify the images by preprocessing them with a region growing image segmentation method.

The validation of our classification algorithm reveals that it computes a climatological latitude of the PNJ in reanalysis very close to 60° N (the reference latitude of the PNJ standard characterization method). This is encouraging, but it can also mean that if we are only interested in assessing the intensity of PNJ, our method does not suppose any advantage with respect to the standard one. Moreover, the standard method is faster both computationally and in terms of human effort than applying a classification algorithm. A classifier requires high human effort for correct training, and more computational resources, which increase with dataset size. Thus, at a first sight the standard method would be more efficient than our method. However, the use of our classification algorithm does have several advantage:

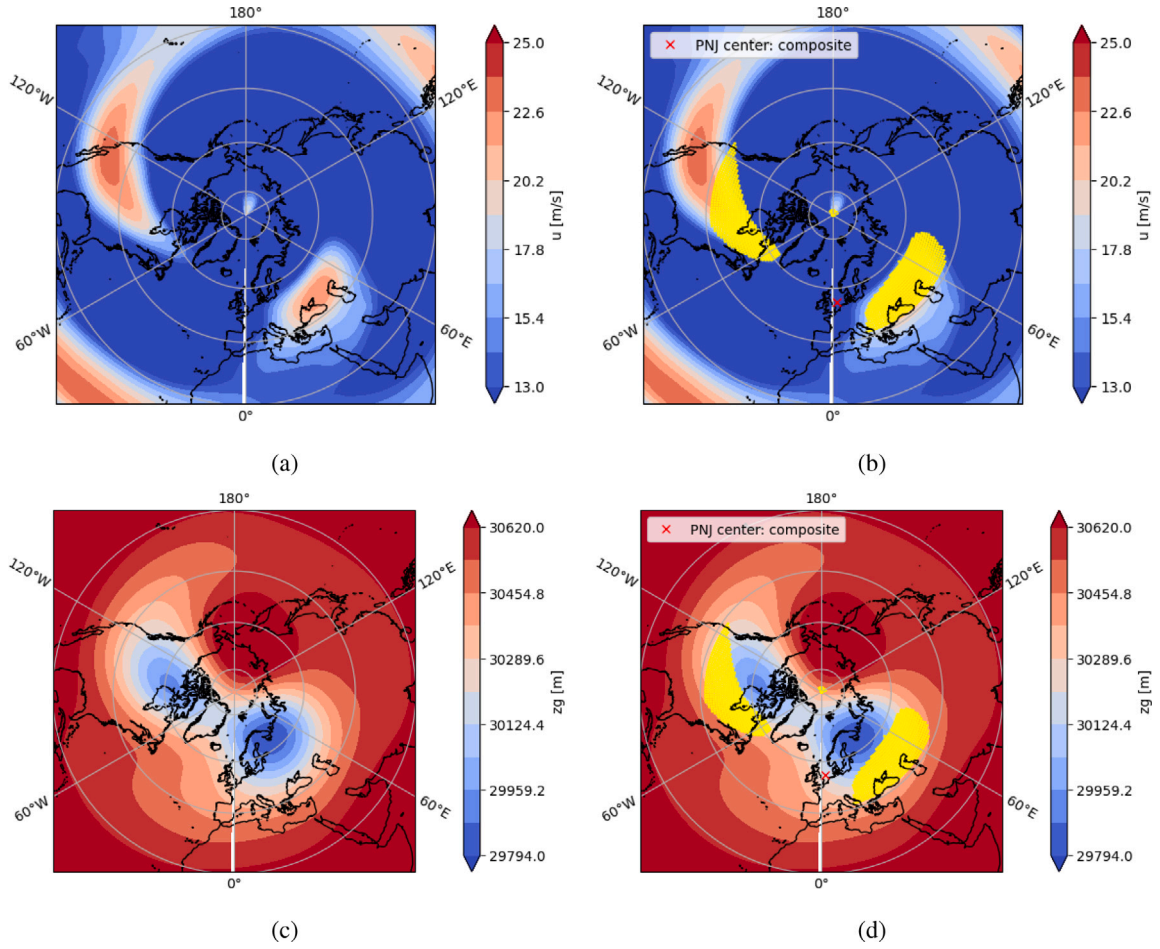

Fig. 5. **(a)** Composite map of zonal wind speed u in JRA-55 dataset for type 2 events. Contour interval 1.2 m/s. North-polar stereographic projection. **(b)** Same as (a) but showing, in yellow, the PNJ region as detected by a decision trees classifier, and in red, the point representing the weighted averaged latitude and longitude of the yellow region. **(c)** Composite map of geopotential height z_g in JRA-55 dataset for type 1 events. Contour interval 82.6 m. North-polar stereographic projection. **(d)** Same as (b) but drawn over the composite map of z_g in JRA-55 dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

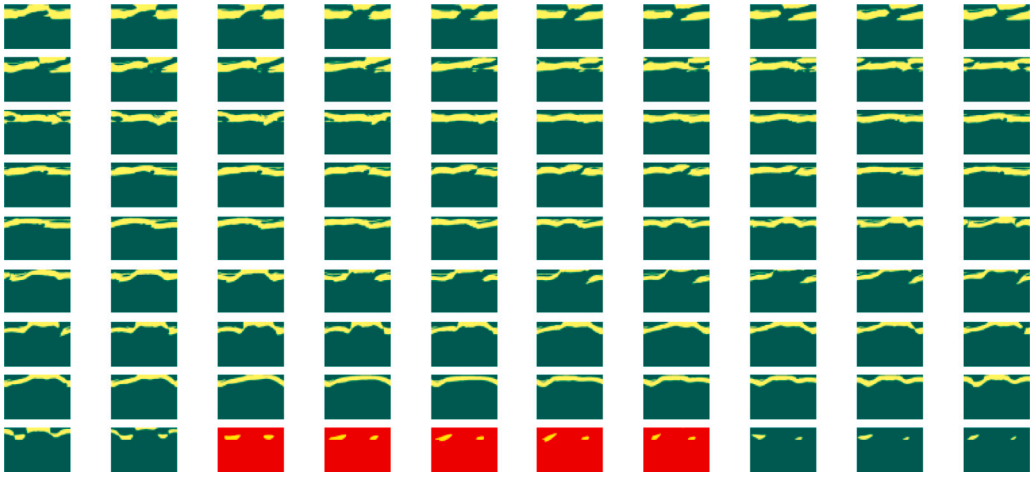

Fig. 6. Polar night jet and type 2 regions, marked in yellow, for 90 days 1978-1979 winter in the JRA-55 daily simulation results, detected over a latitude–longitude grid as a function of zonal wind speed u and geopotential height z_g features. Type 2 events detected by our algorithm have a red background, the remaining days have a green background. The dates of the type 2 events detected by our algorithm are 1979-02-21, 1979-02-22, 1979-02-23, 1979-02-24, and 1979-02-25. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Once trained, our classifier, as mentioned in Section 1, is able to obtain, in one step, different parameters, from one image as an input: extension of the polar night jet, averaged magnitudes over the PNJ area, and type of PNJ event. Machine learning classifiers offer the possibility to obtain, at least, the combined accuracy of existing scientific methods and human labeling. In our case, we

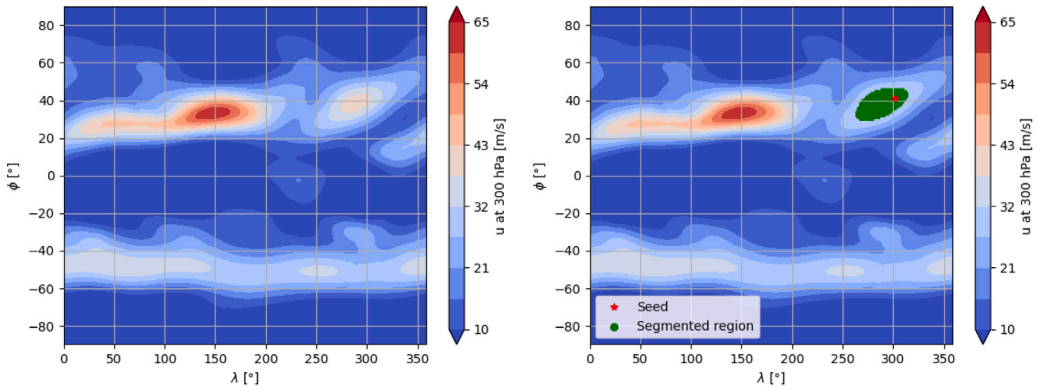

Fig. 7. **(a)** Zonal wind u at 300 hPa for the composite winter monthly data of the CESM2-WACCM piControl simulation. Contour interval 5.5 m/s. **(b)** Same as (a) but including the point of maximum zonal wind speed in the North Atlantic region, in red. The figure also includes, in green, the North Atlantic jet contour selected by the region growing algorithm as described in Section 3.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

have reached an accuracy of 95.5%. Having been trained with a robust ground truth, machine learning classifiers can be trusted to function unsupervised in new datasets (Muller et al., 2021).

- Our algorithm is able to characterize the PNJ in more detail with respect to area and location, preventing from undervaluing the intensity of the PNJ when its core is southward shifted as the standard method might do. This also implies that the classifier is not only able to detect the PNJ region and intensity, but it also identifies extreme episodes of the PNJ variability such as the aforementioned type 1 events, in which the PNJ is southward shifted. These kinds of events have been found by setting the analysis as a function of variables such as latitude and intensity, variables which would not be possible to be set as conditions if using the standard method. More variables could be added in the future, in order to refine the algorithm, as well as introducing optimization techniques to make a parametric analysis. This kind of postprocessing could also be useful for inferring which variables drive PNJ changes, and which ones do not, paving the way for causation or other studies (Runge et al., 2019).

Knowing where the PNJ (roughly the edge of the polar vortex) is located and its intensity, information that is provided by our algorithm, is important for the reasons described in Section 1 and has many potential applications. In this study we have briefly explored some of them such as the analysis of the effects of climate on the boreal polar stratosphere or the study of the variability of the PNJ, in particular, the SSWs. In the first case, the application of our classifier results in a more accurate analysis of the impact of climate change on the PNJ and so, the polar vortex. We have observed the variability of climate change response of the PNJ, which highly depends on the climate model, both in terms of intensity and frequency of extreme events. We have confirmed the lack of consensus in this response among models already documented in the literature, which applied the 60° N characterization method for calculations. The fact that our algorithm can additionally provide the latitude of the PNJ, would, in a more extended analysis, allow to draw more conclusions about changes in intensity and extreme event frequency in relation to polar vortex location, both with and without a CO2 forcing. So far, our findings show that under the same climate conditions, a higher PNJ intensity corresponds to a relatively higher latitude. Interestingly, we also find that the PNJ latitude decreases with carbon emissions regardless of the sign of the change in the intensity. This last result should be investigated in more detail in future analysis. Thus, the application of our classifier to a multi-model analysis containing more climate models would allow us to draw more conclusions, which could potentially be used for making fast and simplified predictions about the future.

As for the variability of the PNJ, our algorithm is able to detect reasonably well the extreme weak PNJ events on a monthly and daily time scale, and even to characterize the type of these disturbances (vortex displacement or split). The classification of these events is relevant for the atmospheric community as there is still a large uncertainty regarding the influence of extreme polar vortex events on near-surface circulation, since not all extreme events are followed by a signal near surface (Baldwin et al., 2021). Some authors (e.g Mitchell et al., 2013; Hall et al., 2021) have suggested that this signal depends on the shape of the perturbed vortex in observations, but this was not further analyzed in detail in model simulations. In the case of our type 2 events (close to split vortex events), our classifier needs to be re-trained with daily fields and z_g images and the performance has been proven to be very successful in comparison with existing algorithms in the literature. The good agreement between our results and those coming from the analysis by Lawrence and Manney (2018) (CAVE-ART) is particularly important because CAVE-ART also applies computer vision techniques, but it involves more requirements that might imply more constrains for the detection and investigation of split events. For instances, while CAVE-ART requires scaled PV, temperature, geopotential height, and zonal and meridional wind as inputs, ours only uses zonal wind and occasionally geopotential height. Further, CAVE-ART usually needs sufficient time sampling as it uses spatial overlap between time steps to detect the split event. In contrast, our algorithm can work with one image on one time step, once it is trained. Lastly, Lawrence and Manney (2018) uses subjective criteria in the final interpretation of results, in order to decide whether a split like event is a SSW, by observing the geometric evolution on the days surrounding each event. Our classifier directly outputs a label for different types of polar night jet extreme events, on the basis of both geometric and physical criteria.

Apart from the applications of our classifier explored in this work, there are many others. For instance, establishing an accurate contour of the polar vortex, and its spatial patterns, would be crucial for explaining the linkage between distribution of trace gases (particularly ozone) and polar vortex. Actually, studies have shown that the morphological changes of the polar vortex can impact the horizontal distribution of ozone in the Southern Hemisphere (Zhang et al., 2017).

Future works are the use of more advanced machine learning algorithms, like CNNs, for the same aim of this study, something which would also require increasing the amount of expert-classified images, with procedures similar to the one explained in Section 3.2. Millions of binary-classified images would be the ideal dataset size for training deep neural networks for image segmentation from scratch (Sharma et al., 2020), and this type of classification has not been performed for stratospheric PNJ images as of now. Another extension of our algorithm might be its application to 3-D data, i.e. fields in longitude–latitude

plane and at different pressure levels. This would be interesting for the analysis of the stratospheric polar vortex given that the core of PNJ also changes its vertical structure during different types of extreme PNJ events (e.g., Serra et al., 2017; Matthewman et al., 2009). Finally, the methods in this paper might be adapted to be applied to other jets, and use with other climate variables for the identification of other atmospheric and stratospheric structures. Fig. 7, for example, shows a first attempt of their use to detect the North Atlantic jet (NAJ) based on a map of the monthly climatology of zonal wind at 300 hPa (a tropopheric level) and when using the region growing algorithm described in Section 3.1. Thus, the algorithm described in this study opens new opportunities to the analysis of atmospheric circulation structures, particularly in large datasets that would otherwise involve extensive computations and so, computational time.

CRediT authorship contribution statement

María Rodríguez-Montes: Data processing, Software, Drafting and revising the manuscript, Take intellectual responsibility for its content. **Blanca Ayarzagüena:** Climate advice, Climate problem design, Drafting and revising the manuscript, Take intellectual responsibility for its content. **María Guijarro:** Artificial Intelligent design, Drafting and revising the manuscript, Take intellectual responsibility for its content.

Acknowledgments

This work was partially supported by the Spanish Ministry of Science, Innovation and Universities under MCI/AEI/FEDER Project number RTI2018-094902-B-C21, and also through the JeDiS (RTI2018-096402-B-I00) project. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. CMIP6 data are allocated at the ESGF archive (https://esgf-node.llnl.gov/projects/cmip6/). The Japanese 55-year Reanalysis (JRA-55) project was carried out by the Japan Meteorological Agency (JMA). JRA-55 data were accessed through NCAR- UCAR Research Data Archive (https://rda.ucar.edu). All authors approved the version of the manuscript to be published.

References

Alfeilat, et al., 2019. Effects of distance measure choice on K-nearest neighbor classifier performance: A review. Big Data 7 (4), 221–248.

Andrews, et al., 1987. Middle Atmosphere Dynamics. In: International Geophysics Series, vol. 40, Academic Press, p. 489.

Awwalu, J., Ogwueleka, F., 2019. On holdout and cross validation: a comparison between neural network and support vector machine. 6, pp. 2394–9333.

Ayarzagüena, et al., 2020. Uncertainty in the response of sudden stratospheric warmings and stratosphere-troposphere coupling to quadrupled CO_2 concentrations in CMIP6 models. J. Geophys. Res.: Atmos. 125 (6), e2019JD032345. http://dx.doi.org/10.1029/2019JD032345.

Baldwin, et al., 2021. Sudden stratospheric warmings. Rev. Geophys. 59 (1), e2020RG000708. http://dx.doi.org/10.1029/2020RG000708.

Bell, et al., 2010. Changes in Northern Hemisphere stratospheric variability under increased CO_2 concentrations. Q. J. R. Meteorol. Soc. 136 (650), 1181–1190. http://dx.doi.org/10.1002/qj.633.

Bieniecki, W., Grabowski, S., 2004. Nearest neighbor classifiers for color image segmentation. http://dx.doi.org/10.1109/TCSET.2004.241511.

Boucher, et al., 2018. IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP. Earth System Grid Federation, http://dx.doi.org/10.22033/ESGF/CMIP6.1534.

Bunzel, F., Schmidt, H., 2013. The brewer–dobson circulation in a changing climate: Impact of the model configuration. J. Atmos. Sci. 70 (5), 1437–1455. http://dx.doi.org/10.1175/JAS-D-12-0215.1.

Charlton, A.J., Polvani, L.M., 2007. A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. J. Clim. 20 (3), 449–469. http://dx.doi.org/10.1175/JCLI3996.1.

Chattopadhyay, et al., 2020. Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. Sci. Rep. 10 (1), 1–13.

Chun, H.Y., Song, B.G., Song, I.S., 2019. The role of gravity waves in the evolution of the vortex-splitting stratospheric sudden warming in January 2009.

Danabasoglu, G., 2019. NCAR CESM2 Model Output Prepared for CMIP6 CMIP 1pctCO2. Earth System Grid Federation, http://dx.doi.org/10.22033/ESGF/CMIP6.7497.

Daniya, et al., 2020. Classification and regression trees with Gini index. Adv. Math. Sci. J. 9, 1857–8438. http://dx.doi.org/10.37418/amsj.9.10.53.

Domeisen, D.I., Garfinkel, C.I., Butler, A.H., 2019. The teleconnection of el niño Southern Oscillation to the stratosphere. Rev. Geophys. 57 (1), 5–47.

Efford, N., 2000. Digital Image Processing: A Practical Introduction using Java (with CD-ROM), first ed. Addison-Wesley Longman Publishing Co., Inc., USA.

Eyring, et al., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organizations. Geosci. Model Dev. 9, 1937–1958.

Fawagreh, et al., 2014. Random forests: from early developments to recent advancements. Syst. Sci. Control Eng. 2 (1), 602–609. http://dx.doi.org/10.1080/21642583.2014.956265.

Fels, et al., 1980. Stratospheric sensitivity to perturbations in ozone and carbon dioxide: Radiative and dynamical response. J. Atmos. Sci. 37 (10), 2265–2297. http://dx.doi.org/10.1175/1520-0469(1980)037<2265:SSTPIO>2.0.CO;2.

Gerber, E.P., Martineau, P., 2022. Extratropical Stratosphere–Troposphere Coupling. SPARC Reanalysis Intercomparison Project (S-RIP) Final Report, SPARC, pp. 221–264. http://dx.doi.org/10.17874/800dee57d13, (Chapter 6).

Gettelman, et al., 2019. The whole atmosphere community climate model version 6 (WACCM6). J. Geophys. Res.: Atmos. 124 (23), 12380–12403. http://dx.doi.org/10.1029/2019JD030943.

Hall, et al., 2021. Tracking the stratosphere-to-surface impact of sudden stratospheric warmings. J. Geophys. Res.: Atmos. 126 (3), e2020JD033881. http://dx.doi.org/10.1029/2020JD033881.

Hardiman, et al., 2019. The impact of strong el niño and la niña events on the north atlantic. Geophys. Res. Lett. 46 (5), 2874–2883. http://dx.doi.org/10.1029/2018GL081776.

He, et al., 2019. Scan-flood fill(SCAFF): an efficient automatic precise region filling algorithm for complicated regions.

Ise, T., Oba, Y., 2019. Forecasting climatic trends using neural networks: An experimental study using global historical data. Front. Robot. AI 6, 32. http://dx.doi.org/10.3389/frobt.2019.00032.

Kalnay, et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–472. http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Karpechko, et al., 2018. Predicting sudden stratospheric warming 2018 and its climate impacts with a multimodel ensemble. Geophys. Res. Lett. 45, http://dx.doi.org/10.1029/2018GL081091.

Keeble, et al., 2021. Using machine learning to make computationally inexpensive projections of 21st century stratospheric column ozone changes in the tropics. Front. Earth Sci. 8, 591. http://dx.doi.org/10.3389/feart.2020.592667.

Kobayashi, et al., 2015. The JRA-55 reanalysis: General specifications and basic characteristics. J. Meteorol. Soc. Japan 93, 5–48.

Kotsiantis, et al., 2006. Machine learning: a review of classification and combining techniques. Artif. Intell. Rev. 26 (3), 159–190. http://dx.doi.org/10.1007/s10462-007-9052-3.

Krützmann, et al., 2008. Identification of mixing barriers in chemistry-climate model simulations using rényi entropy. Geophys. Res. Lett. 35, L06806.

Labitzke, K., 1977. Interannual variability of the winter stratosphere in the northern hemisphere. Mon. Weather Rev. 105 (6), 762–770. http://dx.doi.org/10.1175/1520-0493(1977)105<0762:IVOTWS>2.0.CO;2.

Labitzke, K.B., 1981. Stratospheric-mesospheric midwinter disturbances - A summary of observed characteristics. J. Geophys. Res. 86, 9665–9678.

Lawrence, Z.D., Manney, G.L., 2018. Characterizing stratospheric polar vortex variability with computer vision techniques. J. Geophys. Res.: Atmos. 123 (3), 1510–1535. http://dx.doi.org/10.1002/2017JD027556.

Liu, et al., 2019. Parallel comparison of major sudden stratospheric warming events in CESM1-WACCM and CESM2-WACCM. Atmosphere 10 (11), http://dx.doi.org/10.1088/1748-9326/ac12f4.

Lu, et al., 2021. The sudden stratospheric warming in January 2021. 16 (8), 084029. http://dx.doi.org/10.1088/1748-9326/ac12f4.

Martineau, P., Chen, G., Son, S.W., Kim, J., 2018. Lower-stratospheric control of the frequency of sudden stratospheric warming events. J. Geophys. Res.: Atmos. 123 (6), 3051–3070.

Matthewman, et al., 2009. A new look at stratospheric sudden warmings. Part III: Polar vortex evolution and vertical structure. J. Clim. 22 (6), 1566–1585. http://dx.doi.org/10.1175/2008JCLI2365.1.

McDonald, A.J., Smith, M., 2013. A technique to identify vortex air using carbon monoxide observations. J. Geophys. Res.: Atmos. 118 (22), 12,719–12,733. http://dx.doi.org/10.1002/2012JD019257.

McInturff, R., 1978. Stratospheric Warmings: Synoptic, Dynamic and General-Circulation Aspects, NASA Reference Publ. Technical Report. NASA-RP-101, NASA, Natl. Meteorol. Cent., Washington, DC.

McLandress, C., Shepherd, T.G., 2009. Simulated anthropogenic changes in the brewer–dobson circulation, including its extension to high latitudes. J. Clim. 22 (6), 1516–1540. http://dx.doi.org/10.1175/2008JCLI2679.1.

Meng, et al., 2020. Interpreting deep learning-based networking systems. In: Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication. pp. 154–171.

Mitchell, et al., 2011. Characterizing the variability and extremes of the stratospheric polar vortices using 2D moment analysis. J. Atmos. Sci. 68 (6), 1194–1213. http://dx.doi.org/10.1175/2010JAS3555.1.

Mitchell, et al., 2012. The effect of climate change on the variability of the Northern hemisphere stratospheric polar vortex. J. Atmos. Sci. 69 (8), 2608–2618.

Mitchell, et al., 2013. The influence of stratospheric vortex displacements and splits on surface climate. J. Clim. 26 (8), 2668–2682. http://dx.doi.org/10.1175/JCLI-D-12-00030.1.

Muller, et al., 2021. Machine learning FOR microstructure classification: how to ASSIGN the ground truth in the most objective way. Adv. Mater. Process. 179 (1), 16–22.

Nash, et al., 1996. An objective determination of the polar vortex using Ertel's potential vorticity. J. Geophys. Res.: Atmos. 101 (D5), 9471–9478. http://dx.doi.org/10.1029/96JD00066.

National Weather Service, 2021. Climate prediction center- Global temperature time series. URL: https://www.cpc.ncep.noaa.gov/products/stratosphere/temperature/.

Nowack, et al., 2018. Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. Environ. Res. Lett. 13 (10), 104016. http://dx.doi.org/10.1088/1748-9326/aae2be.

Oberländer, et al., 2013. Unraveling impact factors for future changes in the Brewer–Dobson circulation. J. Geophys. Res.: Atmos. 118 (18), 10,296–10,312. http://dx.doi.org/10.1002/jgrd.50775.

Pedregosa, et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Ping, et al., 2001. The origin of the subtropical anticyclones. J. Atmos. Sci. 58 (13), 1827–1835. http://dx.doi.org/10.1175/1520-0469(2001)058<1827:TOOTSA>2.0.CO;2.

Pomerantz, M.A., 1963. WMO/IQSY international years of the quiet sun, 1964–65. Science 142 (3596), 1136–1143.

Rajini, et al., 2011. Classification of MRI brain images using k-nearest neighbor and artificial neural network. In: 2011 International Conference on Recent Trends in Information Technology. ICRTIT, pp. 563–568. http://dx.doi.org/10.1109/ICRTIT.2011.5972341.

Rao, J., Garfinkel, C.I., 2021. CMIP5/6 models project little change in the statistical characteristics of sudden stratospheric warmings in the 21st century. Environ. Res. Lett. 16 (3), 034024. http://dx.doi.org/10.1088/1748-9326/abd4fe.

Rao, et al., 2021. Northern hemisphere sudden stratospheric warming and its downward impact in four chinese CMIP6 models. Adv. Atmos. Sci. 38, 187–202, URL: https://doi.org/10.1007/s00376-020-0250-0.

Runge, et al., 2019. Inferring causation from time series in Earth system sciences. Nature Commun. 10.

Salmi, N., Rustam, Z., 2019. Naïve bayes classifier models for predicting the colon cancer. IOP Conf. Ser.: Mater. Sci. Eng. 546, 052068. http://dx.doi.org/10.1088/1757-899x/546/5/052068.

Sebestyén, et al., 2021. The applicability of big data in climate change research: The importance of system of systems thinking. Front. Environ. Sci. 9, 70. http://dx.doi.org/10.3389/fenvs.2021.619092.

Sedona, et al., 2020. Exploration of machine learning methods for the classification of infrared limb spectra of polar stratospheric clouds. Atmos. Meas. Tech. 13 (7), 3661–3682. http://dx.doi.org/10.5194/amt-13-3661-2020.

Serra, et al., 2017. Uncovering the edge of the polar vortex. J. Atmos. Sci. 74 (11), 3871–3885. http://dx.doi.org/10.1175/JAS-D-17-0052.1.

Seviour, et al., 2013. A practical method to identify displaced and split stratospheric polar vortex events. Geophys. Res. Lett. 40 (19), 5268–5273. http://dx.doi.org/10.1002/grl.50927.

Sharma, et al., 2020. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. 7 (4), 566–574. http://dx.doi.org/10.1016/j.inpa.2019.11.001.

Sheeren, et al., 2009. Discriminating small wooded elements in rural landscape from aerial photography: A hybrid pixel/object-based analysis approach. Int. J. Remote Sens. 30, 4979–4990. http://dx.doi.org/10.1080/01431160903022928.

Shi, et al., 2020. Comparison of major sudden stratospheric warming impacts on the mid-latitude mesosphere based on local microwave radiometer CO observations in 2018 and 2019. Remote Sens. 12 (23), http://dx.doi.org/10.3390/rs12233950, URL: https://www.mdpi.com/2072-4292/12/23/3950.

Simpson, et al., 2018. The downward influence of uncertainty in the northern hemisphere stratospheric polar vortex response to climate change. J. Clim. 31 (16), 6371–6391. http://dx.doi.org/10.1175/JCLI-D-18-0041.1.

Smith, M.L., McDonald, A.J., 2014. A quantitative measure of polar vortex strength using the function M. J. Geophys. Res.: Atmos. 119 (10), 5966–5985. http://dx.doi.org/10.1002/2013JD020572.

Swart, et al., 2019a. CCCma CanESM5 Model Output Prepared for CMIP6 CMIP 1pctCO2. Earth System Grid Federation, http://dx.doi.org/10.22033/ESGF/CMIP6.3151.

Swart, et al., 2019b. The canadian earth system model version 5 (CanESM5.0.3). Geosci. Model Dev. 12 (11), 4823–4873. http://dx.doi.org/10.5194/gmd-12-4823-2019.

Taylor, et al., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 (4), 485–498. http://dx.doi.org/10.1175/BAMS-D-11-00094.1.

Trisal, S., Kaul, A., 2019. K-RCC: A novel approach to reduce the computational complexity of KNN algorithm for detecting human behavior on social networks. J. Intell. Fuzzy Systems 36, 1–23. http://dx.doi.org/10.3233/JIFS-181336.

Vargin, et al., 2020. Study of the variability of spring breakup dates and Arctic stratospheric polar vortex parameters from simulation and reanalysis data. Izv. Atmos. Ocean. Phys. 56 (5), 458–469.

Waugh, D.N.W., 1997. Elliptical diagnostics of stratospheric polar vortices. Q. J. R. Meteorol. Soc. 123 (542), 1725–1748. http://dx.doi.org/10.1002/qj.49712354213.

Waugh, D.W., Polvani, L.M., 2010. Stratospheric polar vortices. In: The Stratosphere: Dynamics, Transport, and Chemistry. American Geophysical Union (AGU), pp. 43–57. http://dx.doi.org/10.1002/9781118666630.ch3.

Waugh, D.W., Randel, W.J., 1999. Climatology of arctic and antarctic polar vortices using elliptical diagnostics. J. Atmos. Sci. 56 (11), 1594–1613. http://dx.doi.org/10.1175/1520-0469(1999)056<1594:COAAAP>2.0.CO;2.

Wu, et al., 2008. Texture feature based Automated Seeded Region growing in abdominal MRI segmentation. J. Biomed. Sci. Eng. 02, 263–267. http://dx.doi.org/10.1109/BMEI.2008.352.

Wu, et al., 2019. Intermodel spread in the northern hemisphere stratospheric polar vortex response to climate change in the CMIP5 models. Geophys. Res. Lett. 46 (22), 13290–13298. http://dx.doi.org/10.1029/2019GL085545.

Zhang, et al., 2016. Persistent shift of the Arctic polar vortex towards the Eurasian continent in recent decades. Nature Clim. Change 6, 1094.

Zhang, et al., 2017. The relationship between polar vortex and ozone depletion in the antarctic stratosphere during the period 1979–2016. Adv. Meteorol. 2017, 1–12. http://dx.doi.org/10.1155/2017/3078079.

Zhang, et al., 2018. Stratospheric ozone loss over the Eurasian continent induced by the polar vortex shift. Nature Commun. 9, 206.

Zhou, et al., 2018. Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning. PLOS Comput. Biol. 14 (7), e1006337. http://dx.doi.org/10.1371/journal.pcbi.1006337.