Evaluation of atmospheric circulation of CMIP6 models for extreme temperature events using Latent Dirichlet Allocation

Nemo Malhomme, Bérengère Podvin, Davide Faranda, Lionel Mathelin

Evaluation of atmospheric circulation of CMIP6 models for extreme temperature events using Latent Dirichlet Allocation

Nemo Malhomme[1,2] , Bérengère Podvin[3] , Davide Faranda[1] , Lionel Mathelin[2]

[1] *ESTIMR, Université Paris-Saclay, CNRS, CEA, UVSQ, Laboratoire des sciences du climat et de l'environnement, 91191, Gif-sur-Yvette, France.*

[2] *LISN, CNRS, Université Paris-Saclay, 91405, Orsay, France.*

[3] *Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire EM2C, 91190, Gif-sur-Yvette, France.*

Corresponding author: Nemo Malhomme, nemo.malhomme@lsce.ipsl.fr

ABSTRACT: We study the ability of large-scale circulation models to reproduce extreme temperature events. To this end, we use a statistical clustering technique, Latent Dirichlet Allocation (LDA) to characterize sea-level pressure data over the north-Atlantic region. From the ERA5 reanalysis dataset, the method extracts a basis of interpretable objects at synoptic scale, that we call "motifs". Pressure data can be projected onto this basis, yielding motif weights that contain local information about the large-scale atmospheric circulation. We first examine how the weights statistics can be used to characterize extreme events in reanalysis data. We then compare the weights obtained from reanalysis data with those obtained from runs from four CMIP6 models. This allows us to quantify errors on each localized circulation pattern and identify model-agnostic and model-specific errors. On average, large-scale circulation is well predicted by all models, but model errors are increased for extreme events such as heatwaves and cold spells. A significant source of error was found to be associated with Mediterranean motifs for all models in all cases. Each model run can be characterized by a dynamic error associated with the global circulation pattern and a thermodynamic error associated with the predicted temperature. In the general case, this two-dimensional characterization is sufficient to discriminate between models. This remains possible in the cold spell case despite higher internal model variability, while all models perform similarly on heatwaves. The detailed characterization provided by LDA analysis is therefore well suited for model preselection for the study of extreme events.

1. Introduction

Heatwaves and cold spells both cause significant public health and safety risks (Weilnhammer et al. (2021)), as well as infrastructure damage (Añel et al. (2017)). They are generally defined as temperature events significantly higher or below average over a period of at least several days. Studies have shown that both the number and the duration of heatwaves in the European region have increased by up to 15% since pre-industrial times (Frich et al. (2002), Alexander et al. (2006)). Examples of severe heatwaves include the European heatwave of 2003 (Fink et al. (2004)), or that of 2018 (McCarthy et al. (2019)). Both events have caused tens of thousands of deaths. While cold spell frequency and intensity have decreased since pre-industrial times (Seneviratne et al. (2021)), they still represent a hazard (López-Bueno et al. (2021)). For instance, we can cite the cold spell of 2017 over the Balkans (Anagnostopoulou et al. (2017)), which had consequent socio-economic impacts. In addition, when occurring during spring, cold spells can have a devastating impact on the development of plants and cause major losses of agricultural yields (Papagiannaki et al. (2014)). One such example is the cold spell of April 2021 described in Vautard et al. (2023b).

Heatwaves and cold spells produce anomalies reaching up to ± 15°C for several consecutive days. This implies that these events cannot be due to local thermodynamic drivers alone. They are explained in large part by changes in atmospheric circulation patterns (Rousi et al. (2022)), namely the ensemble of cyclones and anticyclones affecting a certain region at a given time. Cyclones and anticyclones advect warm or cold air from polar to tropical latitude and vice-versa through the mechanism of baroclinic instability (Wallace and Hobbs (2006)). With the temperature difference between pole and equator reaching up to 60 degrees, cyclones and anticyclones can advect warm and cold air and trigger heatwaves or cold spells. These cyclones and anticyclones evolve most of the time from west to east, because they are embedded in the jet stream. Sporadically, the jet stream creates large meanders that trap cyclones and anticyclones in the same position for several days (Krishnamurti (1961)). This phenomenon, called blocking, can cause persistence of warm or cold conditions in the same areas and trigger heatwave and cold spells (Faranda et al. (2016); Lupo (2021)). Conditions of atmospheric circulation patterns that can cause extreme temperature events are often referred to as their dynamic drivers (Chan et al. (2022)). Simulating the large excursions from the mean temperature responsible for hot and cold prolonged periods in Europe is crucial to understand, anticipate and mitigate the impacts of heatwaves and cold-spells. Global

and regional climate models are extensively used for this purpose both in present, past and future climate conditions (Eyring et al. (2016)).

However, models still face severe limitations in performing this task. According to the Coupled Model Intercomparison Project (CMIP, Meehl et al. (2000)), the statistical properties of extreme events are reasonably well captured by the models, but challenges persist in reproducing their frequencies and intensities, as well as in capturing local specificities (Kharin et al. (2013), Li et al. (2021)). For example, Vautard et al. (2023a) show that models underestimate the trend of evolution of heatwaves, and Jeong et al. (2021) show that models still underestimate the frequency of cold spells. Models are still unable to accurately reproduce the behavior of the atmosphere and ocean. In particular, they tend to underestimate the warming induced by climate change (van Oldenborgh et al. (2009)), and still contain inaccuracies that affect local circulation patterns (Scaife et al. (2010)), including those linked with extreme heat (D'Andrea et al.), and extreme cold (Davini and D'Andrea (2020)). Despite these biases compared to reality, models have made significant progress over the years. There have been increases in grid resolution, reaching a resolution as high as 1 km in regional models (Lucas-Picher et al. (2021)). Tuning techniques have been developed to reduce biases, such as regarding arctic sea ice cover extent, or the amplitude of Atlantic Meridional Overturning Circulation (Mignot et al. (2021)). Such improvements of the models have resulted in an increased ability to represent observed circulation patterns (Rodrigues et al. (2018)). To assess the advances and the remaining challenges, it is necessary to develop evaluation methodologies that give a comprehensive and accurate measure of a model's ability to capture extremes and their drivers.

Regarding dynamic drivers, it is difficult to study directly atmospheric patterns, owning to their high dimensionality. Several methods attempt to produce a reduced-order representation of the atmospheric circulation. One option is to categorize circulation fields into a set of weather regimes, large-scale quasi-stable states of atmospheric circulation (as first introduced in Rex (1950)). Regimes are effective to describe persistent weather patterns (such as in Vautard (1990)). This is useful to the study of extreme events, since some weather patterns, such as the above-mentioned blockings, can induce extreme events such as cold spells or heatwaves. However, by construction, weather regimes are not localized in space (Michelangeli et al. (1995)). They combine various atmospheric structures that are local, such as, for example, cyclones or anticyclones, into

large-scale atmospheric states. This loses the locality and the ability to differentiate between the components. They also typically exist at a time scale too large to define individual extreme events. Another technique is to use climate indices. Climate indices are one-dimensional variables that characterize the state of large-scale patterns, typically oscillations in oceanic circulation patterns that have a large influence over the global and regional climate (Stenseth et al. (2003)). They condense information into a parameter that can be directly studied, and its correlation with all kinds of observables and events measured (de Freitas and Grigorieva (2017)). This is especially useful to study oceanic oscillations, for example (Hanley et al. (2003)). However, since climate indices aggregate a lot of data in a single variable, relevant information about the underlying circulation can be missing.

In this study, we show that a technique introduced in Fery et al. (2022) can provide new insight on the atmospheric circulation of extreme weather events and give both local and global quantitative measures of the performance of climate models. The technique relies on a statistical learning tool known as Latent Dirichlet Allocation (LDA) (Blei et al. (2003)). Originally developed for text analysis, it has shown promise in capturing latent structures within complex datasets outside of natural language processing, such as in fluid mechanics (Frihat et al. (2021)), or environmental sciences (Valle et al. (2018)). In Fery et al. (2022), application of the LDA method to NCEP/NCAR sea-level pressure (SLP) maps led to the identification of latent variables, or "motifs". Those motifs consist of synoptic objects, spatially localized pressure anomalies of the scale of 1000 km. Each map can be represented by a weighted combination of motifs. By monitoring the temporal evolution of the weights, they identified trends in impacts-defined extreme events.

In this paper, we show that LDA decomposition can be a useful tool to evaluate the performance of climate models, and in particular to quantify their ability to reproduce extreme temperature events. The paper is organized as follows. In section 2, we present the datasets to be analysed and our methods of analysis. In section 3, motifs extracted from the ERA5 SLP dataset are used to study the synoptic configuration of hot and cold temperature extremes occurring in France. A comparison between the reanalysis and climate models using this synoptic representation is reported in section 4. An evaluation of the climate models is carried out in section 5, based on the joint analysis of the synoptic representation error and the average temperature discrepancy. A conclusion is given in section 6.

2. Methods

a. Climate data

We choose the reanalysis dataset ERA5 (Hersbach et al. (2020)) as the ground truth to train LDA on and compare the models to. Our variable of study is the sea-level pressure (SLP), which contains the synoptic information relevant to a meteorological study, specifically the positions and extents of cyclones and anticyclones. An alternative for these properties would be 500 hPa geopotential height (z500). However, in ERA5 reanalysis data, z500 is computed from SLP rather than simulated directly. The data is converted into anomalies by removing the seasonal cycle, computed for each date as the average of all days corresponding to that date.

We chose to evaluate general circulation models, because they represent the physical detail of the atmospheric circulation. At time of writing, the CMIP6 project contains the state of the art in general circulation models. We select four CMIP6 models for which a high number of runs is available: IPSL-CM6A-LR (33 runs) (Boucher et al. (2020)), MIROC6 (50 runs) (Tatebe et al. (2019)), ACCESS-ESM1.5 (29 runs) (Ziehn et al. (2020)), and CanESM5 (25 runs) (Swart et al. (2019)).

b. Extreme event definition

Among extreme weather events, we study specifically cold spells and heatwaves. It is generally agreed upon that these terms refer to periods of temperatures significantly higher or below average for at least several days. However, any definition more precise is somewhat arbitrary, and there is no general consensus on a specific definition. A definition can be based on socio-economic impacts, on physical indicators, or the events can be automatically categorized through machine learning methods trained on data categorized by hand (such as in Liu et al. (2016)).

Since we are interested in evaluating model dynamics, while Fery et al. (2022) uses a definition based on impacts, we prefer to use a physics-based definition. In particular, we define a cold spell (resp. heatwave) as at least 3 consecutive days with average daily temperature below the 0.03 quantile (resp. beyond the 0.97 quantile) of average temperatures over the studied period, from 1950 to 2021. Extreme events are defined for specific regions by considering the average temperatures over that region. To illustrate the method, we will consider cold spells and heatwaves

occurring in France. The results for five other countries, Italy, Spain, Poland, Germany and the UK, are available in supplemental material.

c. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is an unsupervised statistical learning method originally devised in the field of natural language processing (Blei et al. (2003)). Its purpose is to extract, from a corpus of written documents, a set of latent variables called "topics" that describe their content. Documents are treated under the "bag of words" hypothesis, which assumes that the ordering of words in documents is irrelevant. A document is defined only by its counts of words belonging to a given vocabulary.

The number of topics K is a hyperparameter of the method, equivalent to a number of clusters. The topics are characterized by their distributions over the vocabulary. These are multinomial distributions parameterized by β, with components $\beta_k, k \in [1, K]$ each defining the distribution associated to the topic of index k.

LDA is a soft clustering technique: each of the D documents in the corpus is associated with a distribution over the topics. This distribution, denoted c, is a multinomial distribution drawn from a Dirichlet distribution of parameter α.

Given the parameters α and β, LDA assumes that each document d of the corpus has been generated as follows:

- A total number of word positions N in the document d is drawn from a Poisson distribution.
- A topic composition $c(d)$ is drawn for the document d (see Fig. 1) from a Dirichlet distribution of parameter α.
- For each word position n in the document:
  - a topic index $z_{d,n}$ is drawn from $c(d)$.
  - a word $w_{d,n}$ is drawn from the word-topic distribution $\beta_{z_{d,n}}$.

The generative process is summarized in Fig. 2.

The joint distribution of all observable and hidden variables, knowing the parameters α and β, is:

$$p\left(\left(w_{d,n}\right)_{d\in[1,D],n\in[1,N]}|\alpha,\beta\right)=\prod_{d=1}^{D}\int_{c}p(c(d)|\alpha)\prod_{n=1}^{N}\sum_{z_{d,n}=1}^{K}p(z_{d,n}|c(d))p(w_{d,n}|z_{d,n},\beta)dc(d) \quad (1)$$

with $c(d)$ drawn from the Dirichlet distribution of parameter α:

$$p(c(d)|\alpha)=\frac{1}{B(\alpha)}\prod_{k=1}^{K}c_k(d)^{\alpha_k-1},\qquad B(\alpha)=\frac{\prod_{k=1}^{K}\Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^{K}\alpha_k)} \quad (2)$$

$z_{d,n}$ drawn from the multinomial distribution $c(d)$:

$$p(z_{d,n}=k|c(d))=c_k(d) \quad (3)$$

$w_{d,n}$ drawn from the multinomial distribution β_z:

$$p(w_{d,n}=i|z_{d,n},\beta)=\beta_{z_{d,n},i} \quad (4)$$

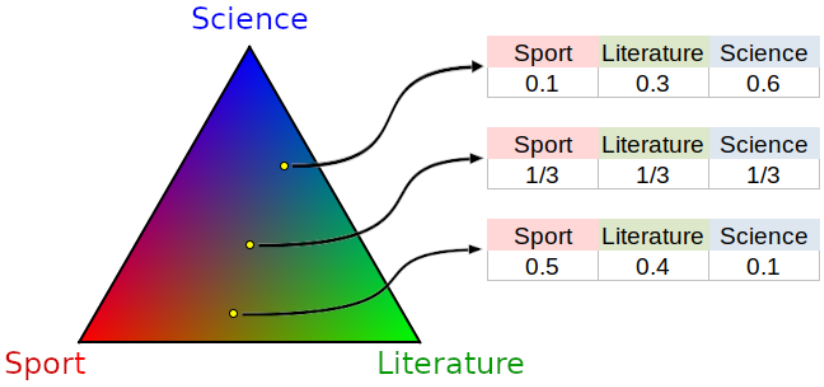

FIG. 1. Example representation of the space of possible topic compositions, on which we define the Dirichlet probability distribution parameterized by α.

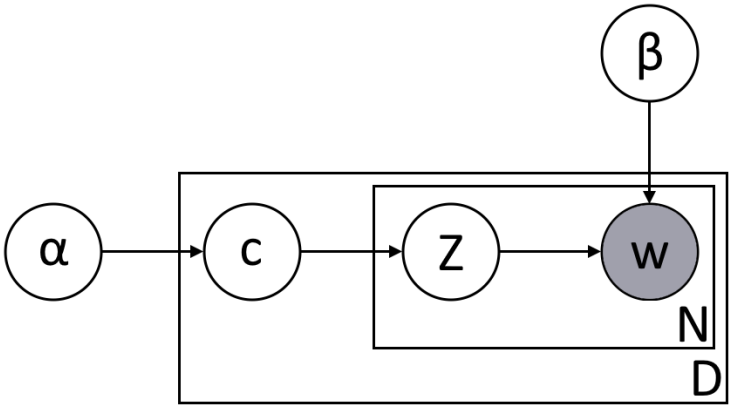

FIG. 2. Graphical model representation of the LDA generative process. Circles represent variables, greyed-out circles being observed variables. Arrows represent the process of drawing a random variable from a distribution. Rectangles represent reiteration of the process, with D being the number of documents, and N the number of words in a document.

This method is applied to datasets of bidimensional climate variables maps where each spatial map is reinterpreted as a document. Grid points, or cells, are reinterpreted as the words, with the list of cells taking the role of the vocabulary. Field values at each cell are reinterpreted as word counts. In this case, the cell-topics distributions β_k are defined over space and are called "motifs". Since the climate variable values are interpreted by LDA as word counts, they have to be digitized and made non-negative. The real variable maps are therefore separated into two channels, one for positive and one for negative values. This is equivalent to doubling the grid size over which the maps are defined. The reader is referred for more details to Fery et al. (2022).

Analysis of a corpus of documents with LDA consists in examining the posterior distribution of the topics β, topic proportions c, and topic assignments z. These are determined via a variational Bayes approach aiming to maximize the evidence lower bound, which is related to the likelihood of the observed data. An additional assumption of this approach is that the β_k are assumed to be drawn from a Dirichlet prior of parameter η. The Dirichlet parameters α and η respectively ensure the sparsity of the document-topic and the topic-word distributions: there are generally few topics in each document, and each topic is characterized by high occurrences of a few vocabulary words.

This sparsity property makes LDA particularly suited to provide models and decompositions that can be interpreted easily. For more information, see Hoffman et al. (2010).

For a given set of D maps, LDA returns motif distributions over grid cells $(\beta_k)_{k\in[1,K]}$, as well as the map compositions $(c_k(d))_{d\in[1,D],k\in[1,K]}$, where $c_k(d)$ denotes the weight of motif k in map d, such that $\forall d \in [1, D]$, $\sum_{k=1}^{K} c_k(d) = 1$. The motif weights $c(d)$ are always positive, unlike other decompositions such as Principal Component Analysis. The set of distributions $(\beta_k)_{k\in[1,K]}$ can be considered as a basis of motifs. Any map P defined on the grid (but not necessarily part of the original set) can be represented in this basis by its $K-$dimensional motif composition $c(P)$. Different sets of maps can thus be compared efficiently through examination of their motif compositions. In practice, numerical implementation of LDA is carried out with the python module Gensim (Řehůřek and Sojka (2010)).

d. Application of LDA

We apply LDA to ERA5 SLP data from the north-Atlantic region between 22.5° and 70° latitude and 80° and 50° longitude. Although higher resolutions are available, we used a spatial resolution of 1° as it was found to be sufficient to contain all relevant information about circulation patterns on the synoptic scale while maintaining manageable computation times. Our resolution is 48 points in latitude, 130 points in longitude, and we have two channels for positive and negative values. Therefore, the total number of values per map, noted N, is 12480. The temporal correlation time of synoptic circulation patterns is approximately 5 days. The full dataset (which will be referred to as general data) consists of daily averaged SLP anomaly fields from 1950 to 2021. The number of motifs was set to $K = 28$, as previous work (Fery et al. (2022)) showed, using a methodology from the field of dynamic systems (Faranda et al. (2017)), that this was the average local dimension of the SLP anomaly data.

These 28 motifs are shown in Fig. 3 and sorted by their average weights in decreasing order. To make discussion easier, names based on their signs and geographical locations were assigned to the motifs. Several motifs in the basis are approximate opposites of one another, such as Labrador high (1) and Labrador low (17), or Genoa low (25) and Mediterranean anticyclone (18). The resulting basis is similar to the one obtained in Fery et al. (2022), which was obtained for different reanalysis datasets at a lower resolution (NCEP/NCAR). Most of the motifs have recognizable equivalents

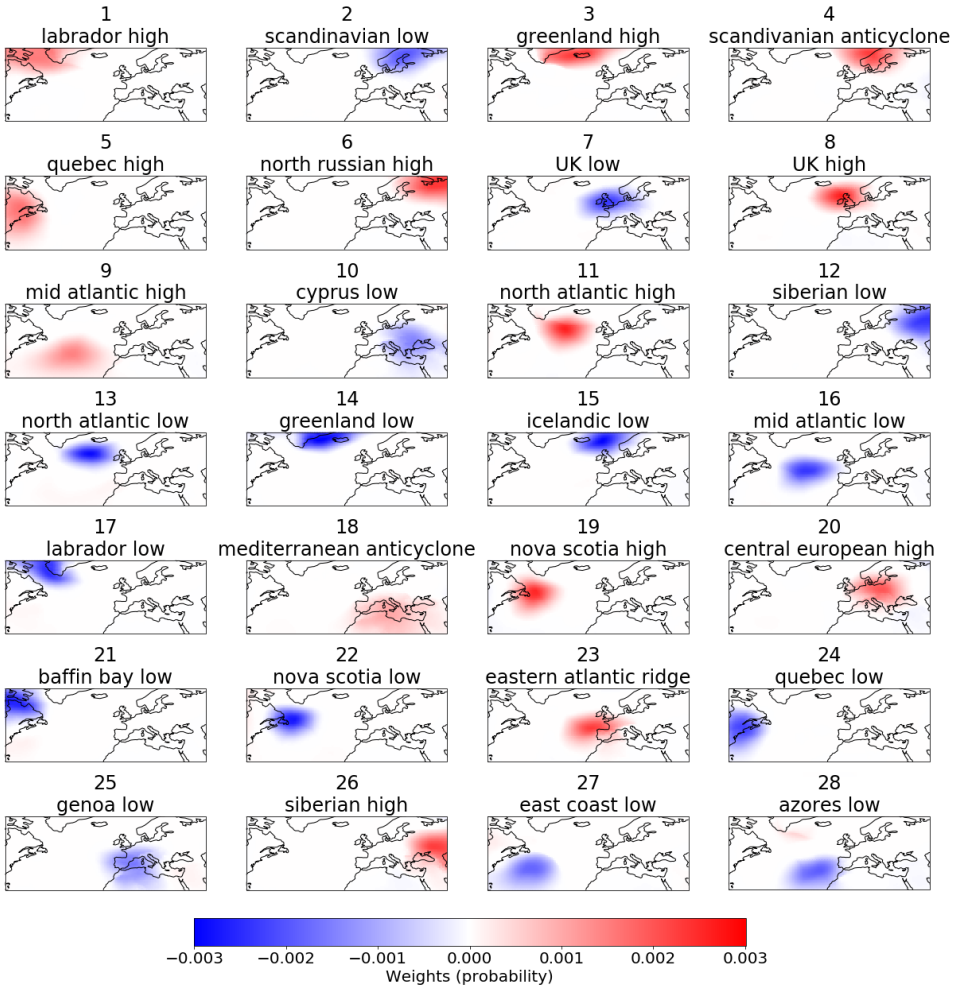

FIG. 3. The basis of 28 motifs learned by LDA from ERA5 SLP anomaly fields. Each motif is defined as a probability distribution over space, with positive and negative channels. The names were given based on sign and geographical location.

from one basis to the other, although some geographical locations may occasionally differ by a few hundred kilometers. Motifs can be seen to be analogous to localized synoptic objects of a given sign, such as cyclones and anticyclones. Therefore, motif weights in a SLP anomaly map directly measure the contribution of the relevant synoptic objects.

LDA offers the possibility of reconstructing maps from a motif composition. The reconstruction of map P, noted P^*, is obtained based on equation (5).

$$P^* = ||P||_1 \sum_{k=1}^{K} c_k(P)\beta_k \tag{5}$$

where:

- β_k is the spatial distribution associated with motif k.
- $c_k(P)$ is the weight of the $k-th$ motif in the weight vector associated with the pressure map P.
- $||P||_1 = \sum_{i=1}^{N} |P_i|$ is the ℓ_1 norm of map P over all N grid cells. This term is a renormalization factor, allowing for direct comparison with physical fields.

In this article, we reconstruct the average compositions of cold spells and heatwaves maps in a given model. In this case, $c_k(P)$ is replaced with $\langle\langle c_k(P)\rangle\rangle$, where $\langle\langle . \rangle\rangle$ designates a conditional average over maps corresponding to the extreme event, and $||P||_1$ is replaced with $||\langle\langle P\rangle\rangle||_1$.

3. Synoptic configuration of extreme events

We first use the decomposition into synoptic objects given by LDA to identify the atmospheric circulation patterns associated with cold spells and heatwaves. The patterns associated with extreme temperatures events in one country are expected to differ from those that would cause such events in another. As mentioned above, we focus our study on extreme temperature events occurring in France. The average synoptic configuration of reanalysis fields corresponding to cold spells (resp. heatwaves) is represented and compared to the average configuration of all reanalysis data in Fig. 4. Uncertainties are estimated by a resampling method called bootstrapping: many alternative sets of cold spell (resp. heatwave) days are generated by randomly sampling with replacements from the original cold spell (resp. heat wave) data. The average motif weights in these datasets are computed, and the 0.05 and 0.95 quantiles weights for each motif are used as lower and upper errors. We found that statistical convergence was reached with 500 datasets, with quantiles chosen to have a 90% confidence interval.

The synoptic configuration of extreme events is different from the average configuration of the general data. Cold spell circulation is dominated by northern anticyclones such as Greenland high, Scandinavian anticyclone and UK high, with more than 6% weights each. Correspondingly, the

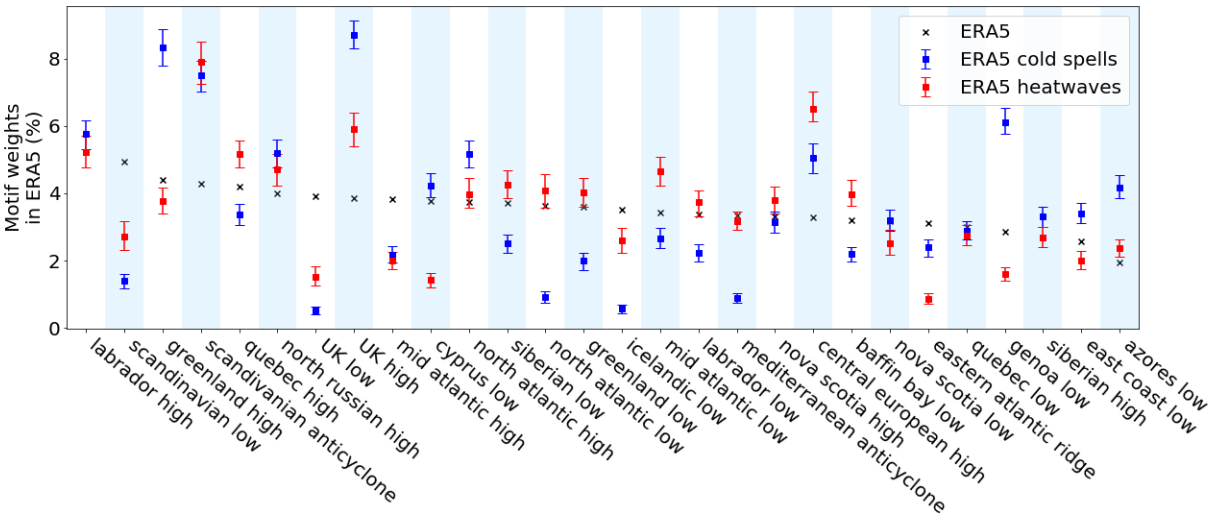

FIG. 4. Average motif weight in the configuration of ERA5 SLP anomaly fields, in the general case (black), in the case of cold spells in France (blue), and in the case of heatwaves in France (red). 90% confidence uncertainties are determined by bootstrapping.

low pressure objects over those regions have less than half the weights they have in the general data. Genoa low is also a key motif in French cold spells, being the fourth most represented motif. Its opposite, the Mediterranean anticyclone, also has during cold spells half the weights it has in general. Heatwave circulation is dominated by a smaller set of high-weights motifs, mainly consisting of Scandinavian anticyclone, and central European high. The UK high is also more prevalent during heatwaves than in general. Both types of extremes are associated with an above-average weights of Scandinavian anticyclone and of UK high.

4. Evaluation of model representation

a. Robustness of the basis

We first establish that a unique basis can be used to compare models with reanalysis data. Fig. 5 shows the correlation matrix between the reanalysis data basis and that obtained from a run from a IPSL-CM6A-LR model, which are respectively associated with cell-motif distributions β and β'.

The correlation matrix is obtained as follows: All fields are set to the same 1° resolution by linear interpolation. For each matrix entry, the Pearson correlation coefficient ρ_{kl} between between motif

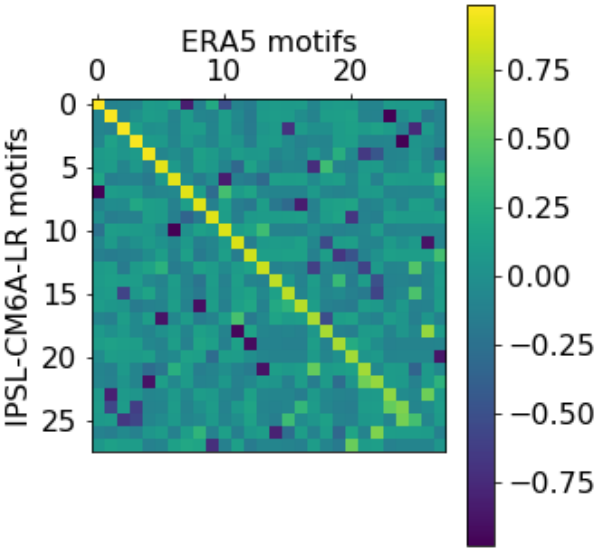

FIG. 5. Spatial correlation between the motifs of the bases obtained by applying LDA on ERA5 (vertical) and on IPSL-CM6A-LR run 1 (horizontal). The order of the motifs has been adjusted to put the highest correlations on the diagonal.

k of basis β and motif l of basis β' is computed as shown in equation (6).

$$\rho_{kl} = \frac{\overline{\left(\beta_k - \overline{\beta_k}\right)\left(\beta'_l - \overline{\beta'_l}\right)}}{\sqrt{\overline{\left(\beta_k - \overline{\beta_k}\right)^2}}\sqrt{\overline{\left(\beta'_l - \overline{\beta'_l}\right)^2}}} \tag{6}$$

where $\bar{\cdot}$ designates the spatial average.

Motifs were reordered in order to give the same rank in the bases to the motifs with the highest correlation. For the case considered, 22 out of 28 motifs have a clear equivalent in the other basis with correlation of at least 0.7 (other choices of models gave similar results). Based on these results, we consider that the motif basis learned from ERA5 is relevant to represent all model data.

b. General data case

We project each run of the four models onto the motif basis learned from ERA5, then average the resulting synoptic configuration of the fields over each run. We first consider all fields in the

datasets. For each run, the relative difference between the K motif weights in the model and that in the reanalysis is computed following equation (7).

$$\forall k \in [1,K], E_k = \frac{\langle c_k(P^{m,r})\rangle - \langle c_k(P)\rangle}{\langle c_k(P)\rangle} \tag{7}$$

where P corresponds to reanalysis maps, $P^{m,r}$ corresponds to maps from run r of model m, and $\langle\cdot\rangle$ designates the average over all maps in the dataset (model run or reanalysis). For each model, the statistics of the error computed for each model run are shown in Figure 6, using box plots. The mean weight of the motifs in the reanalysis data is also indicated for comparison.

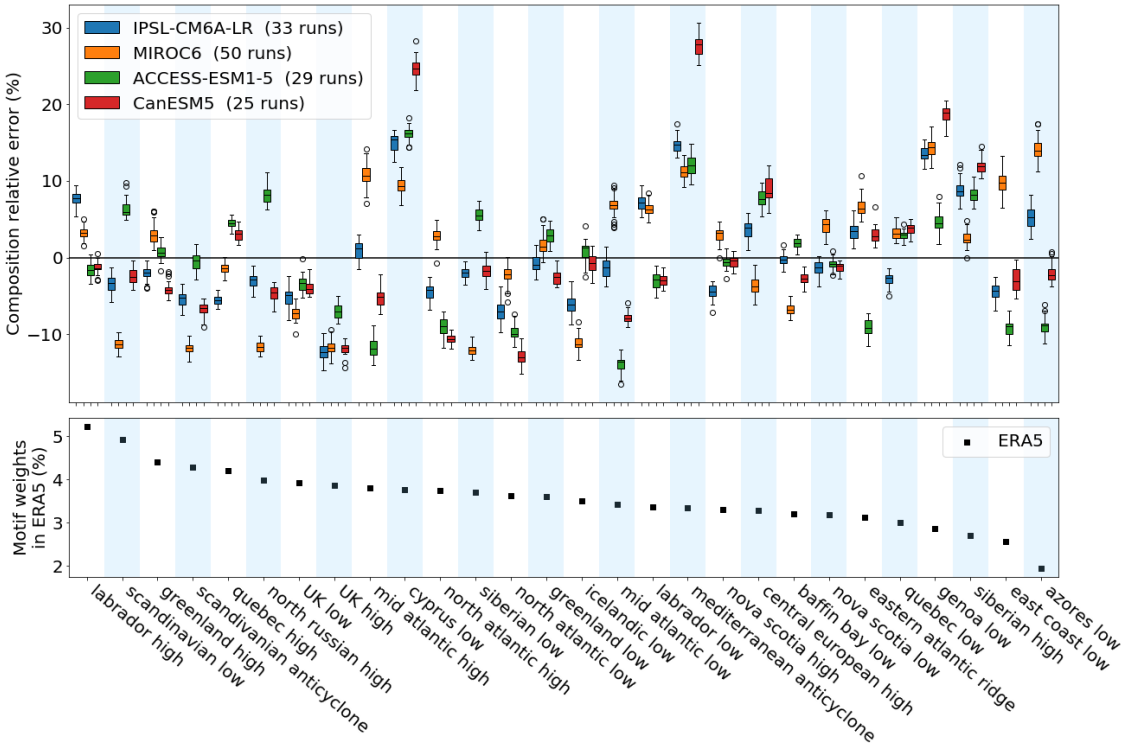

FIG. 6. Top: Relative error on average motif weight between models and ERA5 reanalysis. The box edges correspond to 1st and 3rd quartiles. The black line is the median. The whiskers extend to the furthest datapoint, up to 1.5 times the difference between the 1st and 3rd quartiles. Datapoints beyond the whiskers are represented as colorless circles. Bottom: average motif weight in the synoptic configuration of ERA5 fields.

The median relative errors, materialized by the black lines within the boxes, are relatively small. In particular, the error is less than 15% for the eight most prevalent motifs in the reanalysis. Overall, models represent well the reanalysis synoptic configuration. Relative errors made by

IPSL-CM6A-LR, MIROC6 and ACCESS-ESM1.5, which have resolutions of respectively 38×53, 34×92, and 39×69 are all below 20%. We note that the largest error (25%) is observed for CanESM5, which has a resolution of 17×46. It is possible that these larger errors could be due to its coarser resolution. Moreover, the inner variability of the models (corresponding to the width of the boxes) is typically much smaller than the error (in 96 cases out of the 112 (87.5%), the model's internal variability is lower than its bias). This shows that all runs make similar predictions and also indicates the presence of a bias inherent to each model.

In addition, the motifs associated with the largest relative errors tend to be the same from one model to another. A multimodel ensemble mean would therefore not eliminate these biases. The largest errors are made on motifs located on the Mediterranean region. The Cyprus low and Mediterranean anticyclone motifs are over-represented in all runs of all four models. Every model run also over-represents Genoa low and under-represents UK high and low. Finally, the Scandinavian anticyclone is the fourth most prevalent motif in the reanalysis, with an average weight of more than 4% yet all models but ACCESS-ESM1.5 systematically under-represent it. These similarities in the model errors suggest that the origin of the errors could be common to all models.

c. Model representation of cold spells

We study how models capture the circulation patterns of extreme events. For this part, we focus on cold spells occurring in France. The datasets are filtered following the definition proposed in section 2.

The fields corresponding to the real and the reconstructed averages are represented in Fig. 7. The real average is obtained by taking a conditional average over all daily fields associated with a cold spell. The reconstructed average is obtained from the average motif compositions of the daily fields included in the conditional average, using equation (5). To identify the most significant motifs associated with each model, the two most prevalent cyclonic and the two most prevalent anticyclonic motifs in each case are annotated on the figure.

The overall synoptic structure associated with French cold spells consists of an anticyclonic structure in the north and a cyclonic structure in the south, with a corridor between the two slanted northeast-southwest, passing through the middle of France. For all models, the real average

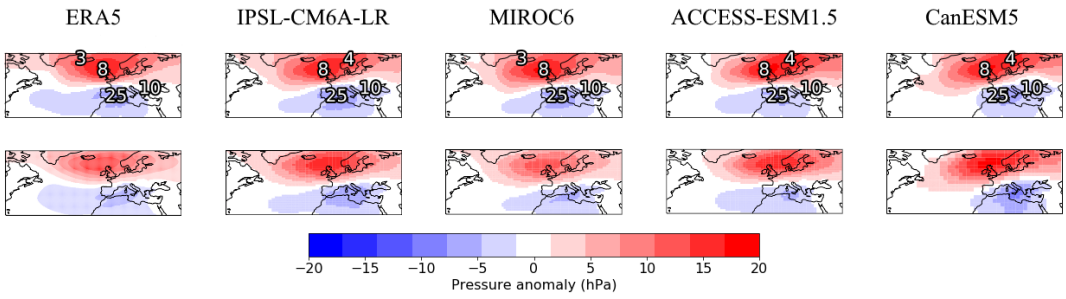

FIG. 7. Top line: Reconstruction of the average motif composition of cold spells in France according to different models (columns). The two cyclones and the two anticyclones with highest average weights in each case are annotated. Bottom line: Average SLP field for cold spells in France according to different models (columns).

is generally similar to its reconstructed average, which shows that LDA captures the synoptic information contained in the real fields.

The model average fields are also in in good agreement with those of ERA5. They have the same two most prevalent cyclones as ERA5, Cyprus low and Genoa low, and reproduce motif 8, UK high, as a dominant motif. However some discrepancies are present: all models underestimate the westward extent of the anticyclonic structure over the Atlantic. Only MIROC6 captures the fact that Greenland high (motif 3) is more prevalent than Scandinavian anticyclone (motif 4), though as seen in section 3, Greenland high and Scandinavian anticyclone are both relevant for French cold spells (near 8% weights). In addition, on CanESM5, Genoa low is too intense, and the cyclonic structure sees no extension to the west of the Mediterranean sea.

For a more detailed analysis, we show for each motif the relative errors in weights between the reanalyses and the models in the case of cold spells occurring in France, in Fig. 8. The biases are significantly higher for the cold extremes than for the general case. The variability among the runs of each models is also higher than for the general case. The five most prevalent reanalysis motifs during French cold spells are UK high, Greenland high, Scandinavian anticyclone, Genoa low, Central European high. Most of these motifs are correctly represented by the models. The significantly higher weights of UK high and Genoa low during cold spells are well captured by all models with an error within the internal variability of all four models. Central European high is also well represented by all models except by ACCESS-ESM1-5 which overestimates it by 25%. The weight of Scandinavian anticyclone high is well captured by IPSL-CM6A-LR and MIROC6,

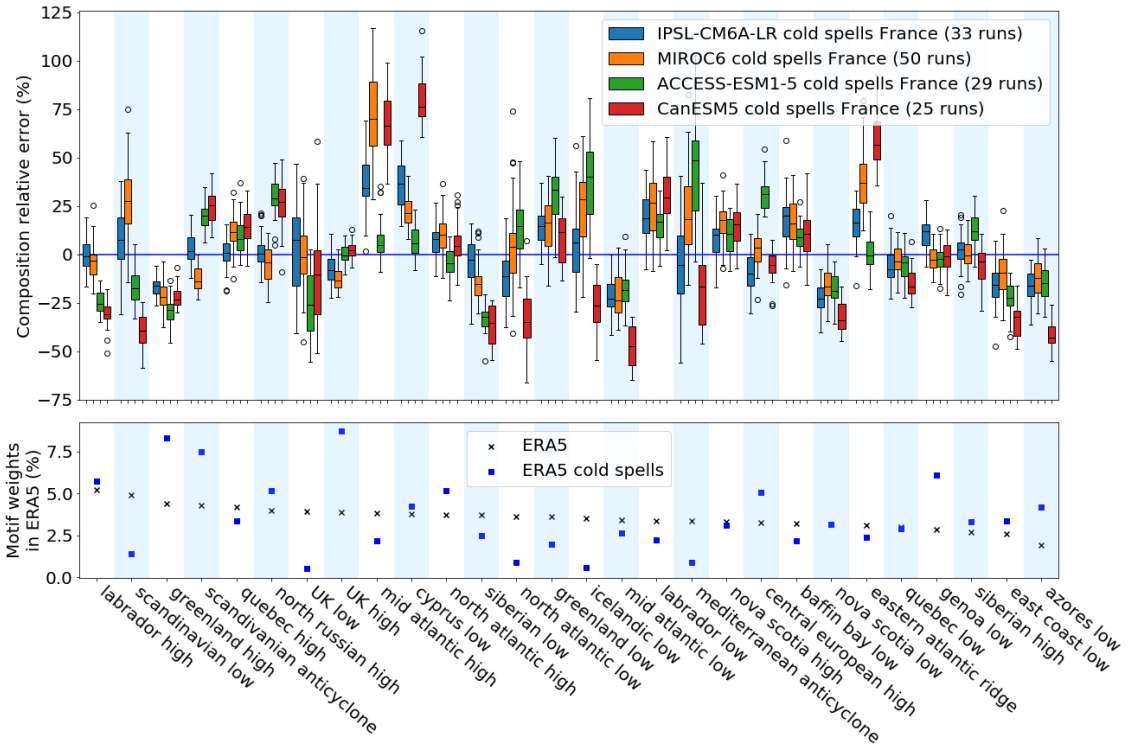

Fig. 8. Top: Relative error on average motif weight between models and ERA5 reanalysis in the case of cold spells occurring in France. Bottom: average motif weight in the synoptic configuration of ERA5 fields, for cold spells and in the general case.

while it is overestimated by 25% by the two other models. All models make about 25% error on Greenland high. Higher errors are made on less relevant motifs where the reanalysis values are lower. The most over-represented motifs are Cyprus low and Mid-Atlantic high for all models except ACCESS-ESM1.5. We note that larger errors are generally observed for the lower resolution model CAN-ESM5.

d. Model representation of heatwaves

We now focus on heatwaves occurring in France. We represent the real and reconstructed average heatwave fields in Fig. 9, using the same methodology as in the previous section.

The SLP anomaly values are weaker than in the case of cold spells. This is because heatwaves are more varied in configuration, leading to average error values closer to zero. There are differences between the real and reconstructed fields. In ERA5 and all models, the anticyclonic structure over

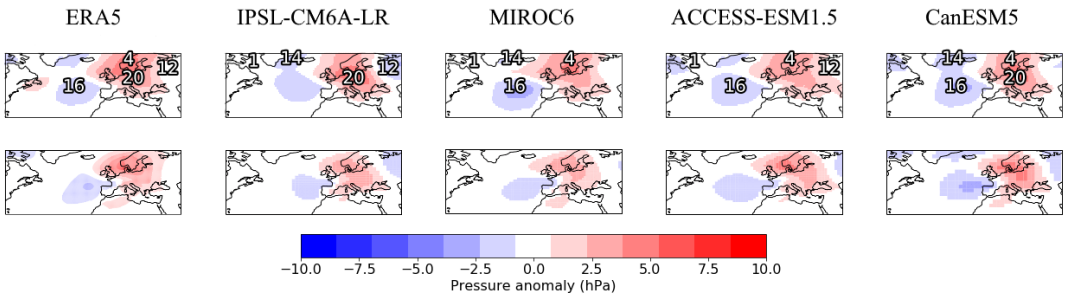

FIG. 9. Top line: Reconstruction of the average motif composition of heatwaves in France according to different models (columns). The two cyclones and the two anticyclones with highest average weights in each case are annotated. Bottom line: Average SLP field for heatwaves in France according to different models (columns).

Europe has a more crescent-like shape around the Atlantic cyclone, that changes into a arrow-like shape in the LDA reconstruction. Still, the overall structure consisting of anticyclones over northern and central Europe with a depression over the Atlantic is preserved by LDA reconstruction.

Models reproduce the overall structure of ERA5 circulation, with anticyclonic conditions on northern and central Europe and cyclones over the Atlantic. Models disagree, with ERA5 and each other, on the shape of those cyclones and the extent of the anticyclonic structure over northern Atlantic. The most prevalent anticyclones in the reanalysis are the Scandinavian anticyclone (motif 4) and the Central European high (motif 20). Only CanESM5 reproduces this property. For the other models, this leads to an anticyclonic structure that is weaker in the north for IPSL-CM6A-LR, in the south for MIROC6, and less intense overall for ACCESS-ESM1.5. The most prevalent cyclones are Siberian low (motif 12) and Mid-Atlantic low (motif 16). Only ACCESS-ESM1.5 reproduces this property.

For a more detailed analysis, we computed relative errors in motif weights between the reanalyses and the models for heatwaves occurring in France. They are shown in Fig. 10.

In the case of heatwaves too, model biases and internal variabilities are higher than in the general case. Which motifs are or are not relevant is generally well captured by the models. However, the most relevant motifs tend to be underpredicted by the models. All models except ACCESS-ESM1.5 under-represent by 20% on average the contribution of the most prevalent motif, which is the Scandinavian anticyclone. The second most prevalent motif, the central European high, is well represented by IPSL-CM6A-LR and CanESM5 but under-represented by about 20% by

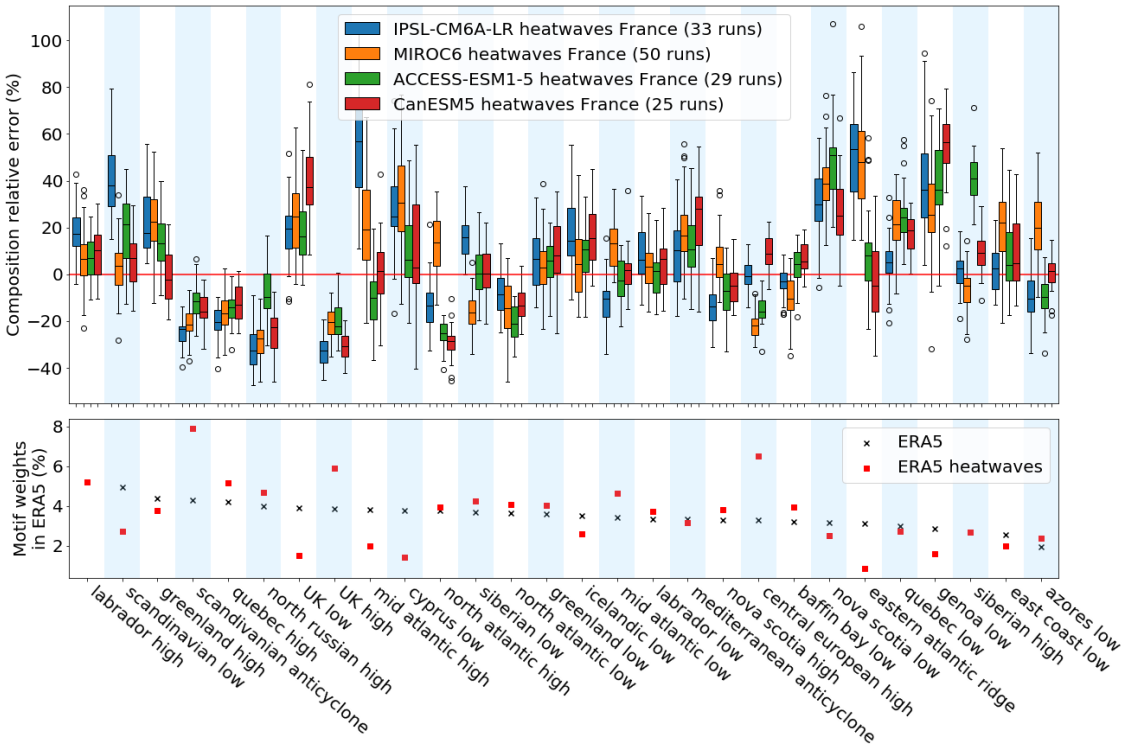

FIG. 10. Top: Relative error on average motif weight between models and ERA5 reanalysis in the case of heatwaves occurring in France. Bottom: average motif weight in the synoptic configuration of ERA5 fields, for heatwaves and in the general case.

MIROC6 and ACCESS-ESM1.5. UK high, the third most prevalent motif, is under-represented by 20% or more by almost all runs of all models. In general, motifs that have higher weights than in the general case tend to be under-represented (as for instance Quebec high and north Russian high), while motifs that have lower weights (UK low, Nova Scotia low, and Genoa low) are over-represented. This shows that models underestimate the changes in atmospheric circulation associated with heatwaves.

5. Global dynamic and thermodynamic error

a. General data case

LDA provides a decomposition of circulation patterns into motifs. Differences in motif weights provide a quantitative measure of model predictive ability in terms of dynamics. The dynamic error of run r of model m, $E_P^{m,r}$, is computed according to equation (8).

$$E_P^{m,r} = \sum_{k=1}^{K} |\langle c_k(P^{m,r})\rangle - \langle c_k(P)\rangle| \tag{8}$$

The dynamic error can be used to evaluate models comparatively, and produce rankings. An important question is to determine whether evaluating models based on thermodynamic error, i.e. the temperature difference between models and reanalysis data, would yield similar results. For run r of model m, the thermodynamic error is computed as shown in equation (9), with T denoting reanalysis temperature fields, and $T^{m,r}$ those from run r of model m.

$$E_T^{m,r} = \left\langle \overline{T^{m,r}} \right\rangle - \left\langle \overline{T} \right\rangle \tag{9}$$

Each model run is represented as a point in the error plane $(E_P^{m,r}, E_T^{m,r})$ shown in Fig. 11. In addition, we annotate for each run the index of the motif which contributes the most to the dynamic error. For each model, we show on the right of the figure the two motifs that appear most frequently as the largest contributor to the error of a run (the proportion of runs each motif corresponds to is indicated between parentheses) - except in the case of ACCESS-ESM1.5, where the largest contributor is always Cyprus low.

Although some overlap between the models would be observed if only one kind of error was considered, each model can be associated with a well-identified cluster in the 2-D error plane. MIROC6 is the model with the highest dynamic and thermodynamic error, but with the lowest thermodynamic variability. Unlike other models, it overpredicts the temperature. In contrast, the IPSL-CM6A-LR model has the highest thermodynamic variability for a relatively low error (similar to that of CanESM5), and it also corresponds to the lowest dynamic error. ACCESS-ESM1.5 has the lowest thermodynamic error for a relatively low dynamic error.

As mentioned earlier, each run is annotated with the index of the motif contributing the most to the dynamic error, which makes it possible to attribute the error to specific motifs and regions in space. Cyprus low (motif 10) is the least well represented motif for all or almost all runs of ACCESS-ESM1.5 and IPSL-CM6A-LR, as well as most runs of CanESM5. Another motif that is occasionally the least well represented in runs of CanESM5 and IPSL-CM6A-LR is Mediterranean anticyclone (motif 18), the opposite of Cyprus low. Both are eastern Mediterranean motifs.

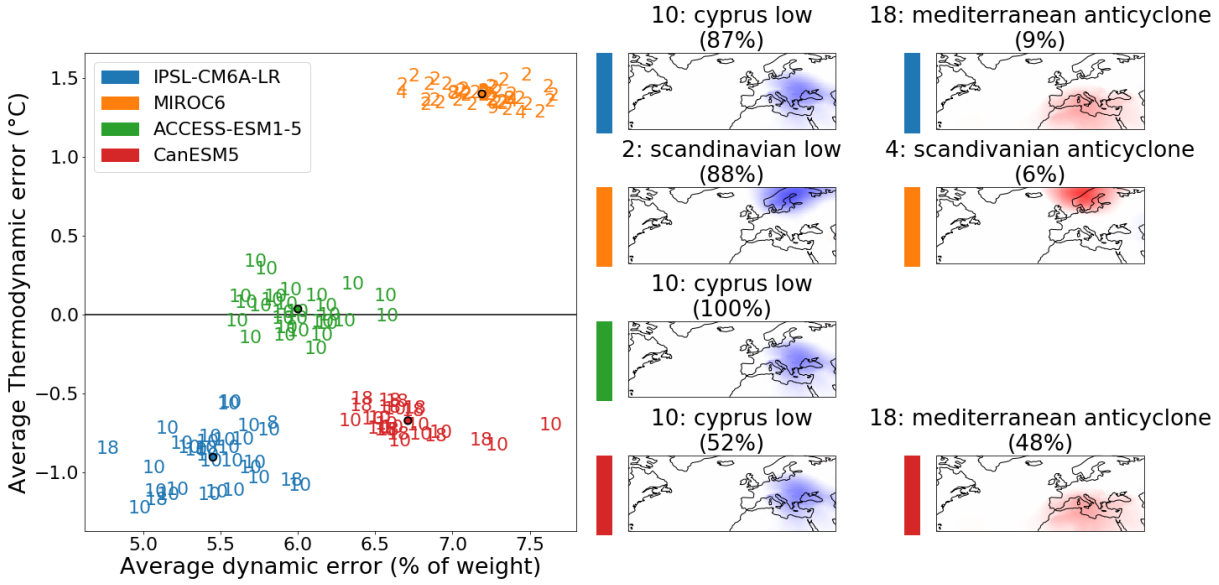

FIG. 11. Run-average thermodynamic model error (average temperature difference with reanalysis), versus run-average dynamic model error (average motif weights difference with reanalysis). The colored dots indicate the average of all runs of a model. Each number corresponds to the motif contributing the most to the error in a given run. The two most frequent such motifs for each model are displayed on the right.

We note that these motifs, which contribute the most to the error, are however not the most prevalent motifs. The associated relative error is therefore necessarily large. This confirms that the representation of the atmospheric circulation over the eastern Mediterranean region is a significant issue for all models, particularly for models IPSL-CM6A-LR, ACCESS-ESM1.5, and CanESM5. MIROC6 appears to differ from other models, as its error on the mean temperature is significantly higher, and its dynamic error is attributed to different motifs than other models, the Scandinavian low and Scandinavian anticyclone (motifs 2 and 4). This points to there being different sources of error between MIROC6 and the other models.

b. Model representation of extreme events

We now consider extreme temperature events and compute the dynamic and thermodynamic errors associated with heatwaves as well as cold spells. In that case, we eliminate the average bias, so as to only look at the component specific to extreme events. We therefore define the anomalous dynamic error $E^{m,r}_{P,ex}$ similarly for heatwaves and cold spells following equation (10).

$$E_{P,ex}^{m,r} = \sum_{k=1}^{K} |\langle\langle c_k(P^{m,r})\rangle\rangle - \langle\langle c_k(P)\rangle\rangle| - E_P^{m,r} \tag{10}$$

The anomalous thermodynamic error $E_{T,ex}^{m,r}$ is defined for heatwaves and cold spells, for run r of model m following equation (11).

$$E_{T,ex}^{m,r} = \left\langle\left\langle \overline{T^{m,r}} \right\rangle\right\rangle - \left\langle\left\langle \overline{T} \right\rangle\right\rangle - E_T^{m,r} \tag{11}$$

In subsequent figures, the dynamic and thermodynamic errors represented are only the anomalous errors defined above. The average errors studied in Fig. 11 are eliminated. However, we note that the general conclusions reported below did not change when these errors were taken into account.

Fig 12 shows model anomalous thermodynamic error against model anomalous dynamic error in the case of cold spells occurring in France.

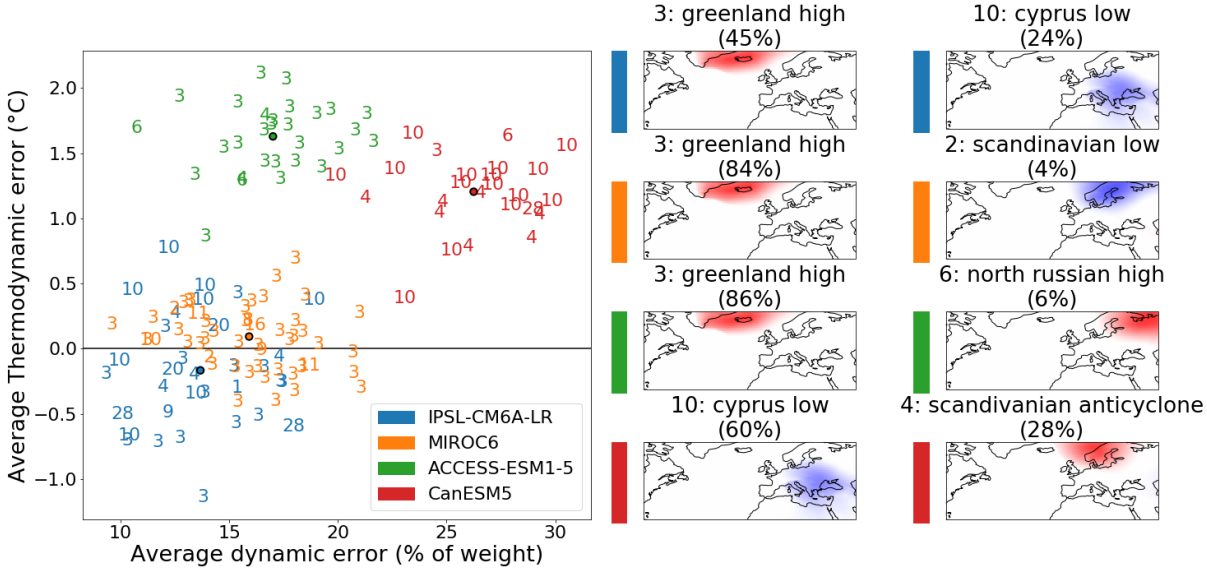

Fig. 12. Run-average thermodynamic model error (average temperature difference with reanalysis) on cold spells in France, versus run-average dynamic model error (average motif weights difference with reanalysis) on same extremes. We eliminate the errors computed in the general case, so as to look only at errors specific to extreme events. The colored dots indicate the average of all runs of a model. Each number corresponds to the motif contributing the most to the error in a given run. The two most frequent such motifs for each model are displayed on the right.

Inner model variability is higher in the cold extreme case than in the full dataset case, both for dynamic and thermodynamic error. There are three distinct clusters in error space, the differences between them being bigger than internal model variabilities. The first cluster corresponds to model CanESM5. It is the model with the highest dynamic error, and has a high thermodynamic error. The second cluster corresponds to model ACCESS-ESM1.5. ACCESS-ESM1.5 has the highest thermodynamic error, underpredicting the lowering of temperature due to cold spells by more than 1.5°C on average. Its dynamic error is comparable to that of MIROC6, and both are made in majority on the same motif (Greenland high). The third cluster consists of two models, IPSL-CM6A-LR and MIROC6. With the general bias removed, the temperature value from reanalysis is within the internal variability of both these models. They are also associated with the lowest dynamic error. This cluster appears to be closest to reanalysis. On average, IPSL-CM6A-LR has a slightly lower dynamic error than MIROC6, but the difference is lower than internal variability.

Greenland high (motif 3) is the least well represented motif on more than 80% of MIROC6 and ACCESS-ESM1.5 runs, as well as 45% of ISPL-CM6A-LR runs. However this does not signify a major model error in the local atmospheric circulation, as the relative error is small, and the significant contribution simply reflects the predominance of the motif in the composition of cold spells. In contrast, for a majority of CanESM5 runs, as well as 24% of IPSL-CM6A-LR runs, the largest contribution to the dynamic error is due to Cyprus low (motif 10). It is not a particularly dominant motif, but one on which the model makes a significant relative error (75% on median, see Fig. 8). Again this suggests a major flaw in the model representation of local circulation over the Mediterranean.

In Fig. 13 we plot model anomalous thermodynamic error against model anomalous dynamic error in the case of heatwaves occurring in France. The inner variability of the models for heatwaves is similar to the cold spell case. However, both thermodynamic and dynamic biases associated with the models are closer, so that in the 2-D error space, regions occupied by each model are overlapping. All four models are associated with similar thermodynamic errors, between +1.0 and +2.5°C - as these biases are all positive, they cannot be removed by use of a multimodel mean.

Still, some differences can be made between the models. CanESM5 has the lowest of both types of error on average and IPSL-CM6A-LR the highest, but the differences are lower than model

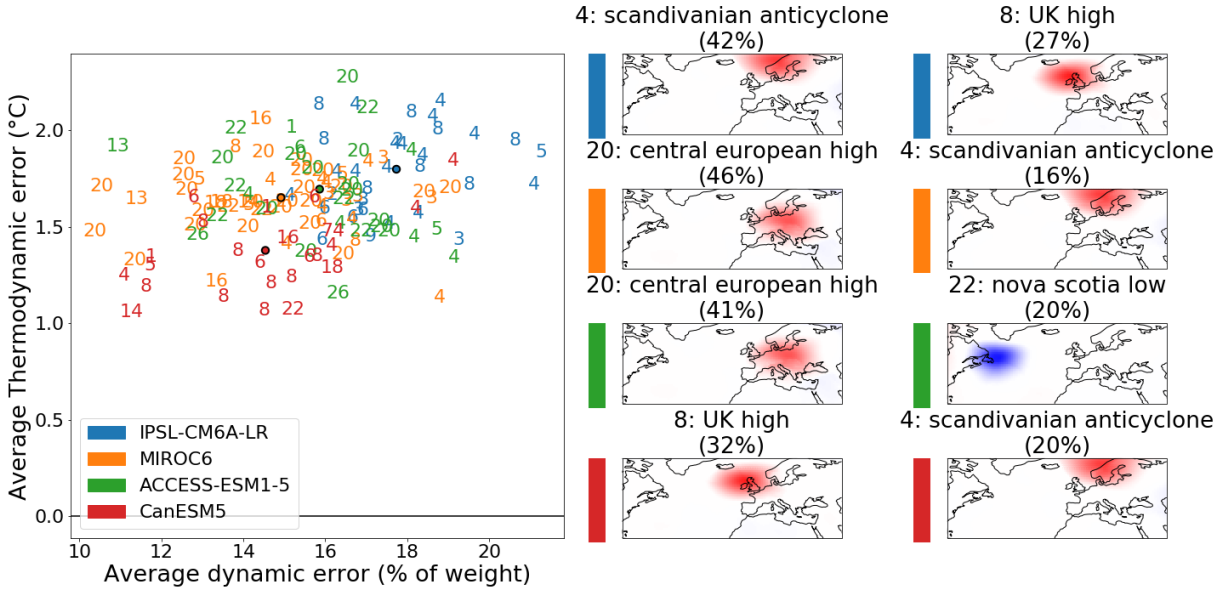

FIG. 13. Run-average thermodynamic model error (average temperature difference with reanalysis) on heatwaves in France, versus run-average dynamic model error (average motif weights difference with reanalysis) on same extremes. We eliminate the errors computed in the general case, so as to look only at errors specific to extreme events. The colored dots indicate the run-average value. Each number corresponds to the motif contributing the most to the error in a given run. The two most frequent such motifs for each model are displayed on the right.

internal variabilities. In addition, motifs that contribute the most to the error vary significantly more from run to run than for both general data and cold spells. In particular, no motif dominates the error in a majority of runs of any model, although some appear more often than others. Central European high (motif 20) appears most frequently as the most significant contributor to the error in runs of both MIROC6 and ACCESS-ESM1.5, while Scandinavian anticyclone (motif 4) makes the largest error contributions in multiple runs of IPSL-CM6A-LR, MIROC6 and CanESM5. However, we note that both Central European high and Scandinavian anticyclone are dominant motifs in heatwaves, so their presence does not reflect a significant relative motif error in the models. To sum up, all models appear to perform comparably for the representation of heat waves, and it seems difficult to identify specific error characteristics in the models.

6. Conclusion

In this paper, we use a statistical learning method called Latent Dirichlet Allocation (LDA) to study the circulation dynamics of ERA5 reanalysis data and CMIP6 general circulation models. Applied to sea-level pressure fields of the north-Atlantic region from ERA5 data, LDA yields a set of latent variables called "motifs" that are recognizable localized synoptic-scale meteorological objects, such as cyclones and anticyclones. By projecting daily sea-level pressure data onto this basis, we obtain the motif composition, which provides a sparse, low-dimensional representation of atmospheric circulation that can be physically interpreted as the associated synoptic configuration. We showed that synoptic configurations averaged over cold spells and heatwaves were both different from each other and from the average taken over the full data.

Using this reanalysis motif basis, we computed the synoptic configuration of runs from 4 different CMIP6 models. Evaluation of the models was based on comparing the statistics of model synoptic configurations with that of reanalysis ones. Differences between models and reanalysis could then be directly attributed to changes in the average weights of individual motifs. This local characterization of the circulation could help discriminate between model predictions, and also help identify the origin of model limitations. Generally speaking, a good agreement was found for general data, while discrepancies were larger for extreme events. In all cases, the largest source of model error was due to the circulation over the eastern Mediterranean region. Moreover, all models tended to underestimate the changes in atmospheric circulation associated with heat waves.

A global dynamic error, based on synoptic configuration differences with reanalysis, was compared with a thermodynamic error, based on the differences in average temperature. These two indicators were found to be sufficient to help discriminate between models when considering general data. Discriminating between models was still possible in the cold spell case, while models performed comparably on heatwaves. This method could therefore be used to determine whether specific models are best suited to the study of a given type of event. Characterization of the error is also relevant to knowing how to aggregate model data, and identifying the biases that can be eliminated this way.

7. Acknowledgements

This work is supported by CNRS-MITI (80 PRIME project ACLIM). We thank Robin Noyelle, Camille Cadiou and Mireia Ginesta-Fernandez for their help in processing the data.

8. Availability statement

This work makes use of the Gensim python module, which is publicly available for download through the pip interface. ERA5 reanalysis data are made publicly available by the Copernicus program, see https://doi.org/10.24381/cds.143582cf. Model datasets used in this article are simulations from the CMIP6 project for the "historical" experiment, by the models IPSL-CM6A-LR, MIROC6, ACCESS-ESM1.5 and CMIP6. The data is publicly available thanks to the World Climate Research Programme, and can be found at: https://esgf-node.llnl.gov/search/cmip6.

9. References

References

L. V. Alexander, X. Zhang, T. C. Peterson, J. Caesar, B. Gleason, A. M. G. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. Rupa Kumar, J. Revadekar, G. Griffiths, L. Vincent, D. B. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Taylor, M. New, P. Zhai, M. Rusticucci, and J. L. Vazquez-Aguirre. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111(D5), 2006. ISSN 2156-2202. doi: 10.1029/2005JD006290. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006290. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006290.

C. Anagnostopoulou, K. Tolika, G. Lazoglou, and P. Maheras. The Exceptionally Cold January of 2017 over the Balkan Peninsula: A Climatological and Synoptic Analysis. *Atmosphere*, 8(12): 252, Dec. 2017. ISSN 2073-4433. doi: 10.3390/atmos8120252. URL https://www.mdpi.com/2073-4433/8/12/252. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

J. A. Añel, M. Fernández-González, X. Labandeira, X. López-Otero, and L. De la Torre. Impact of Cold Waves and Heat Waves on the Energy Production Sector. *Atmosphere*, 8(11):209, Nov. 2017. ISSN 2073-4433. doi: 10.3390/atmos8110209. URL https://www.mdpi.com/2073-4433/8/11/209. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3(null):993–1022, Mar. 2003. ISSN 1532-4435.

O. Boucher, J. Servonnat, A. L. Albright, O. Aumont, Y. Balkanski, V. Bastrikov, S. Bekki, R. Bonnet, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, A. Caubel, F. Cheruy, F. Codron, A. Cozic, D. Cugnet, F. D'Andrea, P. Davini, C. de Lavergne, S. Denvil, J. Deshayes, M. Devilliers, A. Ducharne, J.-L. Dufresne, E. Dupont, C. Éthé, L. Fairhead, L. Falletti, S. Flavoni, M.-A. Foujols, S. Gardoll, G. Gastineau, J. Ghattas, J.-Y. Grandpeix, B. Guenet, E. Guez, Lionel, E. Guilyardi, M. Guimberteau, D. Hauglustaine, F. Hourdin, A. Idelkadi, S. Joussaume, M. Kageyama, M. Khodri, G. Krinner, N. Lebas, G. Levavasseur, C. Lévy, L. Li, F. Lott,

T. Lurton, S. Luyssaert, G. Madec, J.-B. Madeleine, F. Maignan, M. Marchand, O. Marti, L. Mellul, Y. Meurdesoif, J. Mignot, I. Musat, C. Ottlé, P. Peylin, Y. Planton, J. Polcher, C. Rio, N. Rochetin, C. Rousset, P. Sepulchre, A. Sima, D. Swingedouw, R. Thiéblemont, A. K. Traore, M. Vancoppenolle, J. Vial, J. Vialard, N. Viovy, and N. Vuichard. Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, 12(7):e2019MS002010, 2020. ISSN 1942-2466. doi: 10.1029/2019MS002010. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS002010.

P. W. Chan, J. L. Catto, and M. Collins. Heatwave–blocking relation change likely dominates over decrease in blocking frequency under global warming. *npj Climate and Atmospheric Science*, 5 (1):68, 2022.

P. Davini and F. D'Andrea. From CMIP3 to CMIP6: Northern Hemisphere Atmospheric Blocking Simulation in Present and Future Climate. *Journal of Climate*, 33(23):10021–10038, Dec. 2020. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-19-0862.1. URL https://journals.ametsoc.org/view/journals/clim/33/23/jcliD190862.xml. Publisher: American Meteorological Society Section: Journal of Climate.

C. R. de Freitas and E. A. Grigorieva. A comparison and appraisal of a comprehensive range of human thermal climate indices. *International Journal of Biometeorology*, 61(3):487–512, Mar. 2017. ISSN 1432-1254. doi: 10.1007/s00484-016-1228-6. URL https://doi.org/10.1007/s00484-016-1228-6.

F. D'Andrea, J.-P. Duvel, G. Rivière, R. Vautard, J. Cattiaux, D. Coumou, D. Faranda, T. Happé, A. Ribes, and P. Yiou. Summer Deep Depressions Increase Over the North Atlantic.

V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, May 2016. ISSN 1991-959X. doi: 10.5194/gmd-9-1937-2016. URL https://gmd.copernicus.org/articles/9/1937/2016/gmd-9-1937-2016.html. Publisher: Copernicus GmbH.

D. Faranda, G. Masato, N. Moloney, Y. Sato, F. Daviaud, B. Dubrulle, and P. Yiou. The switching between zonal and blocked mid-latitude atmospheric circulation: a dynamical system perspective. *Climate Dynamics*, 47:1587–1599, 2016.

D. Faranda, G. Messori, and P. Yiou. Dynamical proxies of North Atlantic predictability and extremes. *Scientific Reports*, 7(1):41278, Jan. 2017. ISSN 2045-2322. doi: 10.1038/srep41278. URL https://www.nature.com/articles/srep41278. Number: 1 Publisher: Nature Publishing Group.

L. Fery, B. Dubrulle, B. Podvin, F. Pons, and D. Faranda. Learning a Weather Dictionary of Atmospheric Patterns Using Latent Dirichlet Allocation. *Geophysical Research Letters*, 49(9):e2021GL096184, 2022. ISSN 1944-8007. doi: 10.1029/2021GL096184. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2021GL096184. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL096184.

A. H. Fink, T. Brücher, A. Krüger, G. C. Leckebusch, J. G. Pinto, and U. Ulbrich. The 2003 European summer heatwaves and drought - synoptic diagnosis and impacts. *Weather*, 59(8):209–216, 2004. ISSN 0043-1656. doi: 10.1256/wea.73.04. URL https://oceanrep.geomar.de/id/eprint/30311/. Number: 8 Publisher: Wiley-Blackwell.

P. Frich, L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. K. Tank, and T. Peterson. Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, 19(3):193–212, Jan. 2002. ISSN 0936-577X, 1616-1572. doi: 10.3354/cr019193. URL https://www.int-res.com/abstracts/cr/v19/n3/p193-212/.

M. Frihat, B. Podvin, L. Mathelin, Y. Fraigneau, and F. Yvon. Coherent structure identification in turbulent channel flow using Latent Dirichlet Allocation. *Journal of Fluid Mechanics*, 920: A27, Aug. 2021. ISSN 0022-1120, 1469-7645. doi: 10.1017/jfm.2021.444. URL http://arxiv.org/abs/2005.10010. arXiv: 2005.10010.

D. E. Hanley, M. A. Bourassa, J. J. O'Brien, S. R. Smith, and E. R. Spade. A Quantitative Evaluation of ENSO Indices. *Journal of Climate*, 16(8):1249–1258, Apr. 2003. ISSN 0894-8755, 1520-0442. doi: 10.1175/1520-0442(2003)16⟨1249:AQEOEI⟩2.0.CO;2.

URL https://journals.ametsoc.org/view/journals/clim/16/8/1520-0442_2003_16_1249_aqeoei_2.0.co_2.xml. Publisher: American Meteorological Society Section: Journal of Climate.

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. ISSN 1477-870X. doi: 10.1002/qj.3803. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803.

M. Hoffman, F. Bach, and D. Blei. Online Learning for Latent Dirichlet Allocation. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/hash/71f6278d140af599e06ad9bf1ba03cb0-Abstract.html.

D. I. Jeong, B. Yu, and A. J. Cannon. Links between atmospheric blocking and North American winter cold spells in two generations of Canadian Earth System Model large ensembles. *Climate Dynamics*, 57(7):2217–2231, Oct. 2021. ISSN 1432-0894. doi: 10.1007/s00382-021-05801-0. URL https://doi.org/10.1007/s00382-021-05801-0.

V. V. Kharin, F. W. Zwiers, X. Zhang, and M. Wehner. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, 119(2):345–357, July 2013. ISSN 1573-1480. doi: 10.1007/s10584-013-0705-8. URL https://doi.org/10.1007/s10584-013-0705-8.

T. N. Krishnamurti. The subtropical jet stream of winter. *Journal of the Atmospheric Sciences*, 18 (2):172–191, 1961.

C. Li, F. Zwiers, X. Zhang, G. Li, Y. Sun, and M. Wehner. Changes in Annual Extremes of Daily Temperature and Precipitation in CMIP6 Models. *Journal of Climate*, 34(9):3441–3460, May 2021. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-19-1013.1. URL https:

//journals.ametsoc.org/view/journals/clim/34/9/JCLI-D-19-1013.1.xml. Publisher: American Meteorological Society Section: Journal of Climate.

Y. Liu, E. Racah, Prabhat, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, and W. Collins. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. *arXiv:1605.01156 [cs]*, May 2016. URL http://arxiv.org/abs/1605.01156. arXiv: 1605.01156.

P. Lucas-Picher, D. Argüeso, E. Brisson, Y. Tramblay, P. Berg, A. Lemonsu, S. Kotlarski, and C. Caillaud. Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, 12(6):e731, 2021. ISSN 1757-7799. doi: 10.1002/wcc.731. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.731. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.731.

A. R. Lupo. Atmospheric blocking events: A review. *Annals of the New York Academy of Sciences*, 1504(1):5–24, 2021.

J. A. López-Bueno, M. Navas-Martín, J. Díaz, I. J. Mirón, M. Y. Luna, G. Sánchez-Martínez, D. Culqui, and C. Linares. The effect of cold waves on mortality in urban and rural areas of Madrid. *Environmental Sciences Europe*, 33(1):72, June 2021. ISSN 2190-4715. doi: 10.1186/s12302-021-00512-z. URL https://doi.org/10.1186/s12302-021-00512-z.

M. McCarthy, N. Christidis, N. Dunstone, D. Fereday, G. Kay, A. Klein-Tank, J. Lowe, J. Petch, A. Scaife, and P. Stott. Drivers of the UK summer heatwave of 2018. *Weather*, 74(11):390–396, 2019. ISSN 1477-8696. doi: 10.1002/wea.3628. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wea.3628. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wea.3628.

G. A. Meehl, G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer. The Coupled Model Intercomparison Project (CMIP). *Bulletin of the American Meteorological Society*, 81(2):313–318, Feb. 2000. ISSN 0003-0007, 1520-0477. doi: 10.1175/1520-0477(2000)081⟨0313:TCMIPC⟩2.3.CO;2. URL http://journals.ametsoc.org/doi/10.1175/1520-0477(2000)081<0313:TCMIPC>2.3.CO;2.

P.-A. Michelangeli, R. Vautard, and B. Legras. Weather Regimes: Recurrence and Quasi Stationarity. *Journal of the Atmospheric Sciences*, 52(8):1237–1256, Apr. 1995. ISSN 0022-4928, 1520-0469. doi: 10.1175/1520-0469(1995)052⟨1237:WRRAQS⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/52/8/1520-0469_1995_052_1237_wrraqs_2_0_co_2.xml. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences.

J. Mignot, F. Hourdin, J. Deshayes, O. Boucher, G. Gastineau, I. Musat, M. Vancoppenolle, J. Servonnat, A. Caubel, F. Chéruy, S. Denvil, J.-L. Dufresne, C. Ethé, L. Fairhead, M.-A. Foujols, J.-Y. Grandpeix, G. Levavasseur, O. Marti, M. Menary, C. Rio, C. Rousset, and Y. Silvy. The Tuning Strategy of IPSL-CM6A-LR. *Journal of Advances in Modeling Earth Systems*, 13(5):e2020MS002340, 2021. ISSN 1942-2466. doi: 10.1029/2020MS002340. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002340. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002340.

K. Papagiannaki, K. Lagouvardos, V. Kotroni, and G. Papagiannakis. Agricultural losses related to frost events: use of the 850 hPa level temperature as an explanatory variable of the damage cost. *Natural Hazards and Earth System Sciences*, 14(9):2375–2386, Sept. 2014. ISSN 1561-8633. doi: 10.5194/nhess-14-2375-2014. URL https://nhess.copernicus.org/articles/14/2375/2014/. Publisher: Copernicus GmbH.

R. Řehůřek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. http://is.muni.cz/publication/884893/en.

D. F. Rex. Blocking Action in the Middle Troposphere and its Effect upon Regional Climate. *Tellus*, 2(4):275–301, Jan. 1950. ISSN 0040-2826. doi: 10.3402/tellusa.v2i4.8603. URL https://doi.org/10.3402/tellusa.v2i4.8603. Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/tellusa.v2i4.8603.

D. Rodrigues, M. C. Alvarez-Castro, G. Messori, P. Yiou, Y. Robin, and D. Faranda. Dynamical Properties of the North Atlantic Atmospheric Circulation in the Past 150 Years in CMIP5 Models and the 20CRv2c Reanalysis. *Journal of Climate*, 31(15):6097–6111, Aug. 2018. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-17-0176.1. URL https://journals.ametsoc.

org/view/journals/clim/31/15/jcli-d-17-0176.1.xml. Publisher: American Meteorological Society Section: Journal of Climate.

E. Rousi, K. Kornhuber, G. Beobide-Arsuaga, F. Luo, and D. Coumou. Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia. *Nature Communications*, 13(1):3851, July 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31432-y. URL https://www.nature.com/articles/s41467-022-31432-y. Number: 1 Publisher: Nature Publishing Group.

A. A. Scaife, T. Woollings, J. Knight, G. Martin, and T. Hinton. Atmospheric Blocking and Mean Biases in Climate Models. *Journal of Climate*, 23(23):6143–6152, Dec. 2010. ISSN 0894-8755, 1520-0442. doi: 10.1175/2010JCLI3728.1. URL https://journals.ametsoc.org/view/journals/clim/23/23/2010jcli3728.1.xml. Publisher: American Meteorological Society Section: Journal of Climate.

S. I. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S. M. Vicente-Serrano, M. Wehner, and B. Zhou. Weather and climate extreme events in a changing climate. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, Yelekçi, R. Yu, and B. Zhou, editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1513–1766. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021. doi: 10.1017/9781009157896.001.

N. C. Stenseth, G. Ottersen, J. W. Hurrell, A. Mysterud, M. Lima, K. Chan, N. G. Yoccoz, and B. Ådlandsvik. Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1529):2087–2096, Oct. 2003. doi: 10.1098/rspb.2003.2415. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2003.2415. Publisher: Royal Society.

N. C. Swart, J. N. S. Cole, V. V. Kharin, M. Lazare, J. F. Scinocca, N. P. Gillett, J. Anstey, V. Arora, J. R. Christian, S. Hanna, Y. Jiao, W. G. Lee, F. Majaess, O. A. Saenko, C. Seiler,

C. Seinen, A. Shao, M. Sigmond, L. Solheim, K. von Salzen, D. Yang, and B. Winter. The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12 (11):4823–4873, Nov. 2019. ISSN 1991-959X. doi: 10.5194/gmd-12-4823-2019. URL https://gmd.copernicus.org/articles/12/4823/2019/. Publisher: Copernicus GmbH.

H. Tatebe, T. Ogura, T. Nitta, Y. Komuro, K. Ogochi, T. Takemura, K. Sudo, M. Sekiguchi, M. Abe, F. Saito, M. Chikira, S. Watanabe, M. Mori, N. Hirota, Y. Kawatani, T. Mochizuki, K. Yoshimura, K. Takata, R. O'ishi, D. Yamazaki, T. Suzuki, M. Kurogi, T. Kataoka, M. Watanabe, and M. Kimoto. Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7):2727–2765, July 2019. ISSN 1991-959X. doi: 10.5194/gmd-12-2727-2019. URL https://gmd.copernicus.org/articles/12/2727/2019/. Publisher: Copernicus GmbH.

D. Valle, P. Albuquerque, Q. Zhao, A. Barberan, and R. J. Fletcher Jr. Extending the Latent Dirichlet Allocation model to presence/absence data: A case study on North American breeding birds and biogeographical shifts expected from climate change. *Global Change Biology*, 24(11):5560–5572, 2018. ISSN 1365-2486. doi: 10.1111/gcb.14412. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14412. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14412.

G. J. van Oldenborgh, S. Drijfhout, A. van Ulden, R. Haarsma, A. Sterl, C. Severijns, W. Hazeleger, and H. Dijkstra. Western Europe is warming much faster than expected. *Climate of the Past*, 5(1):1–12, Jan. 2009. ISSN 1814-9324. doi: 10.5194/cp-5-1-2009. URL https://cp.copernicus.org/articles/5/1/2009/. Publisher: Copernicus GmbH.

R. Vautard. Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors. *Monthly Weather Review*, 118(10):2056–2081, Oct. 1990. ISSN 1520-0493, 0027-0644. doi: 10.1175/1520-0493(1990)118⟨2056:MWROTN⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/mwre/118/10/1520-0493_1990_118_2056_mwrotn_2_0_co_2.xml. Publisher: American Meteorological Society Section: Monthly Weather Review.

R. Vautard, J. Cattiaux, T. Happé, J. Singh, R. Bonnet, C. Cassou, D. Coumou, F. D'Andrea, D. Faranda, E. Fischer, A. Ribes, P. Yiou, and S. Sippel. Heat extremes in Western Europe are

increasing faster than simulated due to missed atmospheric circulation trends, Jan. 2023a. URL https://hal.science/hal-03937057.

R. Vautard, G. J. van Oldenborgh, R. Bonnet, S. Li, Y. Robin, S. Kew, S. Philip, J.-M. Soubeyroux, B. Dubuisson, N. Viovy, M. Reichstein, F. Otto, and I. Garcia de Cortazar-Atauri. Human influence on growing-period frosts like in early April 2021 in central France. *Natural Hazards and Earth System Sciences*, 23(3):1045–1058, Mar. 2023b. ISSN 1561-8633. doi: 10.5194/nhess-23-1045-2023. URL https://nhess.copernicus.org/articles/23/1045/2023/. Publisher: Copernicus GmbH.

J. M. Wallace and P. V. Hobbs. *Atmospheric science: an introductory survey*, volume 92. Elsevier, 2006.

V. Weilnhammer, J. Schmid, I. Mittermeier, F. Schreiber, L. Jiang, V. Pastuhovic, C. Herr, and S. Heinze. Extreme weather events in europe and their health consequences – A systematic review. *International Journal of Hygiene and Environmental Health*, 233:113688, Apr. 2021. ISSN 1438-4639. doi: 10.1016/j.ijheh.2021.113688. URL https://www.sciencedirect.com/science/article/pii/S1438463921000018.

T. Ziehn, M. A. Chamberlain, R. M. Law, A. Lenton, R. W. Bodman, M. Dix, L. Stevens, Y.-P. Wang, J. Srbinovsky, T. Ziehn, M. A. Chamberlain, R. M. Law, A. Lenton, R. W. Bodman, M. Dix, L. Stevens, Y.-P. Wang, and J. Srbinovsky. The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, 70(1):193–214, Aug. 2020. ISSN 2206-5865, 2206-5865. doi: 10.1071/ES19035. URL https://www.publish.csiro.au/es/ES19035. Publisher: CSIRO PUBLISHING.