



HAL
open science

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

Nemo Malhomme, Bérengère Podvin, Davide Faranda, Lionel Mathelin

► **To cite this version:**

Nemo Malhomme, Bérengère Podvin, Davide Faranda, Lionel Mathelin. Evaluation of atmospheric circulation of CMIP6 models for extreme temperature events using Latent Dirichlet Allocation. 2024. hal-04484617v2

HAL Id: hal-04484617

<https://cnrs.hal.science/hal-04484617v2>

Preprint submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3 Nemo Malhomme^{1,2}, Bérengère Podvin³, Davide Faranda¹, Lionel Mathelin²

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

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

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

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

10 **ABSTRACT:** For climate models to continue improving, we need to uncover as many discrep-
11 ancies they have with reality as possible. In particular, evaluating the representation of extreme
12 events is important but challenging owing to their rarity. Here, we study how general circulation
13 models reproduce large-scale atmospheric circulation associated with extreme temperature events.
14 To this end, we apply Latent Dirichlet Allocation (LDA), a dimensionality reduction method, to a
15 set of sea-level pressure ERA5 maps over the north-Atlantic region. LDA provides a basis of sparse
16 latent modes called “motifs” that consist of localized objects at synoptic scale. Any pressure map
17 can be approximated by a generally sparse combination of motifs, whose coefficients are called
18 the weights, containing local information about large-scale circulation. Weights statistics can be
19 used to locally characterize circulation patterns, in general and during extreme events, allowing
20 for detailed comparison of datasets. For four CMIP6 models and reanalysis, we quantify local
21 circulation errors and identify model-agnostic and model-specific biases. On average, large-scale
22 circulation is well predicted by all models, but model errors are increased for heatwaves and cold
23 spells. Significant errors were found to be associated with Mediterranean motifs for all models in
24 all cases. In addition, the combination of motif and temperature error can discriminate between
25 models in the general and cold spell cases, while models perform similarly on heatwaves. The
26 sparse characterization provided by LDA analysis is therefore well suited for model preselection
27 for the study of extreme events.

28 **1. Introduction**

29 Heatwaves and cold spells both cause significant public health and safety risks (Weilhammer
30 et al. (2021)), as well as infrastructure damage (Añel et al. (2017)). They are generally defined
31 as temperature events significantly higher or below average over a period of at least several days.
32 Studies have shown that both the number and the duration of heatwaves in the European region
33 have increased by up to 15% since pre-industrial times (Frich et al. (2002), Alexander et al. (2006)).
34 Examples of severe heatwaves include the European heatwave of 2003 (Fink et al. (2004)), or that
35 of 2018 (McCarthy et al. (2019)). Both events have caused tens of thousands of deaths. While cold
36 spell frequency and intensity have decreased since pre-industrial times (Seneviratne et al. (2021)),
37 they still represent a hazard (López-Bueno et al. (2021)). For instance, we can cite the cold spell
38 of 2017 over the Balkans (Anagnostopoulou et al. (2017)), which had consequent socio-economic
39 impacts. In addition, when occurring during spring, cold spells can have a devastating impact
40 on the development of plants and cause major losses of agricultural yields (Papagiannaki et al.
41 (2014)). One such example is the cold spell of April 2021 described in Vautard et al. (2023b).

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

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

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

79 Regarding dynamic drivers, it is difficult to study directly atmospheric patterns, owing to their
80 high dimensionality. Several methods attempt to produce a reduced-order representation of the
81 atmospheric circulation. One option is to categorize circulation fields into a set of weather
82 regimes, large-scale quasi-stable states of atmospheric circulation (as first introduced in Rex
83 (1950)). Regimes are effective to describe persistent weather patterns (such as in Vautard (1990)).
84 This is useful to the study of extreme events, since some weather patterns, such as the above-
85 mentioned blockings, can induce extreme events such as cold spells or heatwaves. However, by
86 construction, weather regimes are not localized in space (Michelangeli et al. (1995)). They combine
87 various atmospheric structures that are local, such as, for example, cyclones or anticyclones, into

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

98 A solution to these difficulties might be offered by machine learning, which has already shown
99 promise in a variety of applications in climate sciences (Reichstein et al. (2019), Chen et al. (2023)).
100 Significant contributions to modelling and prediction problems have been obtained through the
101 application of machine learning, such as in the case of hydrology (Ardabili et al. (2020)), or
102 predictions on subseasonal to seasonal timescales (Cohen et al. (2019)). Deep learning models
103 have been used to model climate change and predict its consequences (Ren et al. (2020)). Machine
104 learning has also been applied to the study of extreme weather events (see Salcedo-Sanz et al.
105 (2024) for a review). Extreme events applications that make use of these methods range from
106 detection and classification (Liu et al. (2016), Gardoll and Boucher (2022)), to prediction (Fang
107 et al. (2021)). However, machine learning, and especially deep learning, have the drawback of
108 being black boxes. Their inner workings can be unknown, which causes the output to be difficult
109 to interpret, and available physical knowledge on the matter to be unused. There are several ways
110 to tackle this issue. One such way is to use “explainable AI”, i.e. machine learning methods that
111 offer ways to understand the processes behind the model predictions. Such methods have seen
112 applications to the subject of extreme weather events (Bommer et al. (2024)). However, as noted
113 by O’Loughlin et al. (2024), to increase our trust in the models, it is important not to only focus on
114 the mapping between model inputs and outputs, but also to provide component-level understanding
115 that makes it possible to attribute the performance of a model to a specific part of its architecture.
116 As a step towards reaching this goal, we present an alternative solution, a dimensionality reduction
117 technique facilitating physical insight through the use of easily interpretable latent factors inferred

118 from the data. An advantage of the learned representation is that it is both sparse in the physical
119 and the parameter space.

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

131 As seen in Fery et al. (2022), the LDA methodology offers advantages in the context of atmo-
132 spheric circulation analysis. This methods reduces the dimension of the data to a limited number
133 of modes while preserving most of the underlying information. Because LDA modes are localized
134 and physically interpretable, information on the localization of relevant synoptic components fol-
135 lows directly from LDA application. Additionally, compared to other statistical learning methods
136 of dimensionality reduction, LDA is doubly sparse. The learned motifs are sparse, with significant
137 values attributed only to a limited number of grid points, and the data representation is also sparse,
138 with only a few significant motifs weights in the LDA composition of any given map.

139 In this article, we show that because of these properties, LDA can be used to locally characterize
140 large-scale atmospheric circulation. In turn, this characterization can be used to comparatively
141 evaluate climate models on their representation of circulation during extreme events. The paper
142 is organized as follows. In section 2, we present the datasets to be analysed and our methods of
143 analysis. In section 3, motifs extracted from the ERA5 SLP dataset are used to study the synoptic
144 configuration of hot and cold temperature extremes occurring in France. A comparison between
145 the reanalysis and climate models using this synoptic representation is reported in section 4. An
146 evaluation of the climate models is carried out in section 5, based on the joint analysis of the

147 synoptic representation error and the average temperature discrepancy. A conclusion is given in
148 section 6.

149 **2. Methods**

150 *a. Climate data*

151 We choose the reanalysis dataset ERA5 (Hersbach et al. (2020)) as the ground truth to train
152 LDA on and compare the models to. Our variable of study is the sea-level pressure (SLP), which
153 contains the synoptic information relevant to a meteorological study, specifically the positions
154 and extents of cyclones and anticyclones. An alternative for these properties would be 500 hPa
155 geopotential height (z500). However, in ERA5 reanalysis data, z500 is computed from SLP rather
156 than simulated directly. An average map is computed for each day of the year. In order to eliminate
157 the seasonal cycle from the data, the corresponding average is subtracted from each map for each
158 day of the year. The result is called anomalies.

159 We chose to evaluate general circulation models, because they represent the physical detail of
160 the atmospheric circulation. At time of writing, the CMIP6 project contains the state of the art
161 in general circulation models. We select four CMIP6 models for which a high number of runs is
162 available: IPSL-CM6A-LR (33 runs) (Boucher et al. (2020)), MIROC6 (50 runs) (Tatebe et al.
163 (2019)), ACCESS-ESM1.5 (29 runs) (Ziehn et al. (2020)), and CanESM5 (25 runs) (Swart et al.
164 (2019)).

165 *b. Extreme event definition*

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

172 Since we are interested in evaluating model dynamics, while Fery et al. (2022) uses a definition
173 based on impacts, we prefer to use a physics-based definition. In particular, we define a cold
174 spell (resp. heatwave) as at least 3 consecutive days with average daily temperature below the

175 0.03 quantile (resp. beyond the 0.97 quantile) of average temperatures over the studied period,
176 from 1950 to 2021. Extreme events are defined for specific regions by considering the average
177 temperatures over that region. To illustrate the method, we will consider cold spells and heatwaves
178 occurring in France. The results for five other countries, Italy, Spain, Poland, Germany and the
179 UK, are available in supplemental material.

180 *c. Latent Dirichlet Allocation*

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

187 The number of topics K is a hyperparameter of the method, equivalent to a number of clusters.
188 Topics are characterized by a distribution over the vocabulary. For each $k \in \{1, 2, \dots, K\}$, the topic
189 of index k is associated with the multinomial distribution parameterized by the vector β_k . We
190 define β as the matrix containing all the β_k parameters. Assuming the vocabulary of possible
191 words to be of size V , the β_k are of length V , and β of shape $K \times V$. An additional assumption of
192 this approach is that the β_k are drawn from a Dirichlet prior of parameter η .

193 LDA is a soft clustering technique: it can associate to any document, included or not in the
194 corpus, a distribution over topics indices. The distribution associated to a document indexed by
195 d is a multinomial distribution over topic indices parameterized by the vector $\mathbf{c}(d)$, of length K ,
196 assumed to be drawn from a Dirichlet distribution of parameter α .

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

- 199 • A total number of word positions N in the document d is drawn from a Poisson distribution.
- 200 • A topic composition $\mathbf{c}(d)$ is drawn for the document d (see Fig. 1) from a Dirichlet distribution
201 of parameter α .
- 202 • For each word position n in the document:

- 203 – a topic index $z_{d,n}$ is drawn from the document-topic distribution parameterized by $\mathbf{c}(d)$.
- 204 – a word $w_{d,n}$ is drawn from the topic-word distribution parameterized by $\beta_{z_{d,n}}$.

205 The generative process is summarized in Fig. 2.

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

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

208 where

- 209 • $\mathbf{c}(d)$ is drawn from the Dirichlet distribution of parameter α :

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

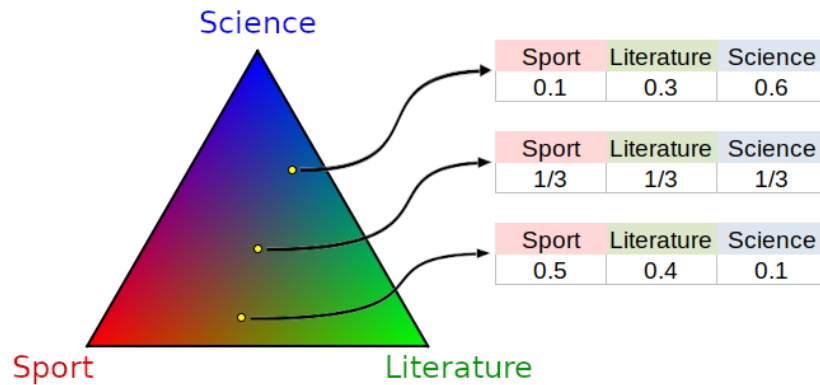
- 210 • $z_{d,n}$ is drawn from the multinomial distribution parameterized by $\mathbf{c}(d)$:

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

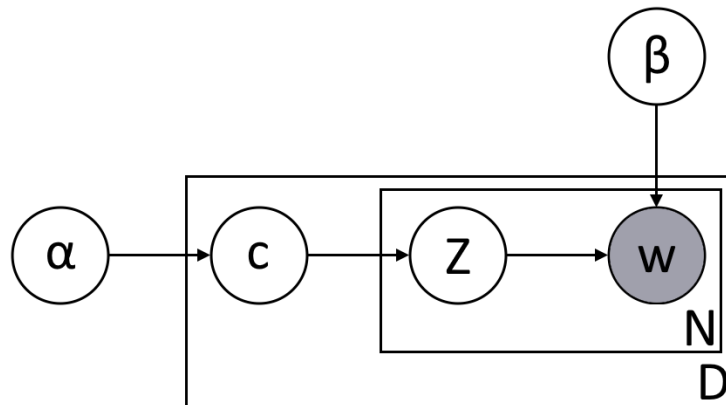
- 211 • $w_{d,n}$ is drawn from the multinomial distribution parameterized by $\beta_{z_{d,n}}$:

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

218 This method is applied to datasets of bidimensional climate variables maps where each spatial
 219 map is reinterpreted as a document. Grid points, or cells, are reinterpreted as the words, with the
 220 list of cells taking the role of the vocabulary. Field values at each cell are reinterpreted as word
 221 counts. In this case, the topic-cell distributions, parameterized by β_k , are defined over space and
 222 are called “motifs”. Since the climate variable values are interpreted by LDA as word counts, they
 223 have to be digitized and made non-negative. To ensure non-negativity, the real variable maps are
 224 separated into two channels, one for positive and one for negative values. This is equivalent to
 225 doubling the grid size over which the maps are defined (see also Fery et al. (2022)). Moreover, a
 226 rescaling factor A is applied to the data before digitization in order to manage computation times,



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



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

227 which depend on the total sum of field values in the maps. To sum up, with \mathbf{p}_i designating the
 228 values of an original map \mathbf{p} at cell i , of a total number of I cells, the LDA formatted version of the

229 map, \mathbf{P} , is defined as follows:

$$\forall i \in \{1, \dots, I\}, \begin{cases} \mathbf{P}_i = \max(\text{int}(A \times \mathbf{p}_i), 0) \\ \mathbf{P}_{i+I} = \max(-\text{int}(A \times \mathbf{p}_i), 0) \end{cases} \quad (5)$$

230 where $\text{int}()$ designates the truncate to integer function. \mathbf{P} is therefore a non-negative, integer-valued
 231 map defined on $V = 2I$ cells.

232 Analysis of a corpus of documents with LDA consists in examining the posterior distribution of
 233 the topics β , topic proportions c , and topic assignments z . These are determined via a variational
 234 Bayes approach aiming to maximize the evidence lower bound, which is related to the likelihood
 235 of the observed data. Small values for the Dirichlet parameters α and η respectively ensure the
 236 sparsity of the document-topic and the topic-word distributions: there are generally few topics in
 237 each document, and each topic is characterized by high occurrences of a few vocabulary words.
 238 This sparsity property makes LDA particularly suited to provide models and decompositions that
 239 can be interpreted easily. For more information, see Hoffman et al. (2010).

240 For a given set of D maps, LDA returns motif distributions over grid cells $(\beta_k)_{k \in [1, K]}$, as well
 241 as the map compositions $(\mathbf{c}_k(d))_{d \in [1, D], k \in [1, K]}$. Throughout the paper, the $\mathbf{c}_k(d)$ will be called
 242 the *weights* of motif k in map d . We note that $\forall d \in [1, D], \sum_{k=1}^K \mathbf{c}_k(d) = 1$. The motif weights
 243 $\mathbf{c}(d)$ are always positive, unlike other decompositions such as Principal Component Analysis. The
 244 set of distributions $(\beta_k)_{k \in [1, K]}$ can be considered as a basis of motifs. Any map \mathbf{P} defined on
 245 the grid (but not necessarily part of the learning set) can be approximated in this basis by its
 246 K -dimensional motif composition $\mathbf{c}(\mathbf{P})$. Different sets of maps can thus be compared efficiently
 247 through examination of their motif compositions. In practice, numerical implementation of LDA
 248 is carried out with the python module Gensim (Řehůřek and Sojka (2010)).

249 *d. Application of LDA*

250 We apply LDA to ERA5 SLP data from the north-Atlantic region between 22.5° and 70° latitude
 251 and 80° and 50° longitude. Although higher resolutions are available, we used a spatial resolution
 252 of 1° as it was found to be sufficient to contain all relevant information about circulation patterns on
 253 the synoptic scale while maintaining manageable computation times. Our resolution is 48 points
 254 in latitude, 130 points in longitude, and we have two channels for positive and negative values.

255 Therefore, the total number of values per map, noted N , is 12480. The temporal correlation time
 256 of synoptic circulation patterns is approximately 5 days. The full dataset (which will be referred to
 257 as general data) consists of daily averaged SLP anomaly fields from 1950 to 2021. The number of
 258 motifs was set to $K = 28$, as previous work (Fery et al. (2022)) showed, using a methodology from
 259 the field of dynamic systems (Faranda et al. (2017)), that this was the average local dimension of
 260 the SLP anomaly data. The rescaling factor is set to $A = 0.5$ to alleviate computation time. Some
 261 arbitrariness exists in the choice of the factor. However, increasing A to 1 did not significantly
 262 change the basis.

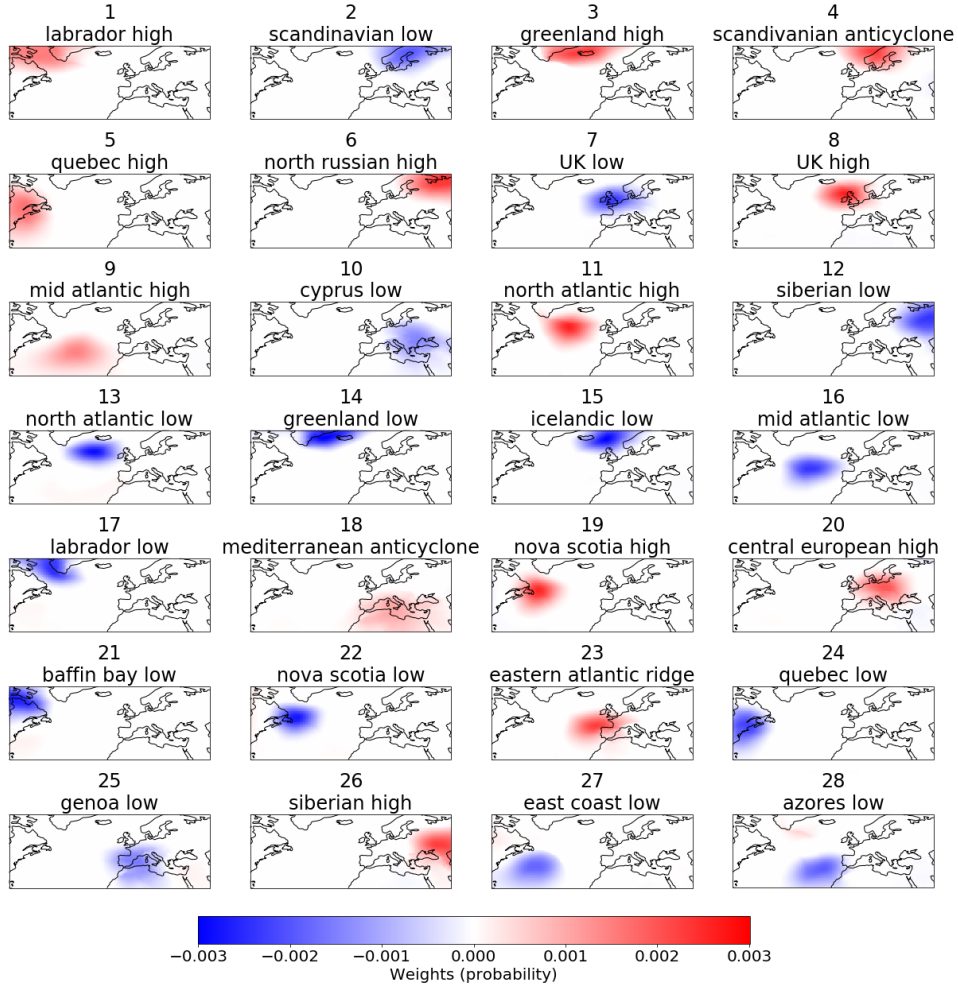
266 These 28 motifs are shown in Fig. 3 and sorted by their average weights in decreasing order. To
 267 make discussion easier, names based on their signs and geographical locations were assigned to the
 268 motifs. Several motifs in the basis are approximate opposites of one another, such as Labrador high
 269 (1) and Labrador low (17), or Genoa low (25) and Mediterranean anticyclone (18). The resulting
 270 basis is similar to the one obtained in Fery et al. (2022), which was obtained for different reanalysis
 271 datasets at a lower resolution (NCEP/NCAR). Most of the motifs have recognizable equivalents
 272 from one basis to the other, although some geographical locations may occasionally differ by a few
 273 hundred kilometers. Motifs can be seen to be analogous to localized synoptic objects of a given
 274 sign, such as cyclones and anticyclones. Therefore, motif weights in a SLP anomaly map directly
 275 measure the contribution of the relevant synoptic objects.

276 LDA offers the possibility of reconstructing maps from a motif composition. The reconstruction
 277 of map \mathbf{P} , noted \mathbf{P}^* , is obtained based on equation (6).

$$\mathbf{P}^* = \|\mathbf{P}\|_1 \sum_{k=1}^K \mathbf{c}_k(\mathbf{P}) \beta_k \quad (6)$$

278 where:

- 279 • β_k is the spatial distribution associated with motif k .
- 280 • $\mathbf{c}_k(\mathbf{P})$ is the weight of the k -th motif in the weight vector associated with the pressure map
 281 \mathbf{P} .
- 282 • $\|\mathbf{P}\|_1 = \sum_{i=1}^V |\mathbf{P}_i|$, i iterating over the V gridpoints, is the ℓ_1 norm of map \mathbf{P} . This term is a
 283 renormalization factor, allowing for direct comparison with physical fields.



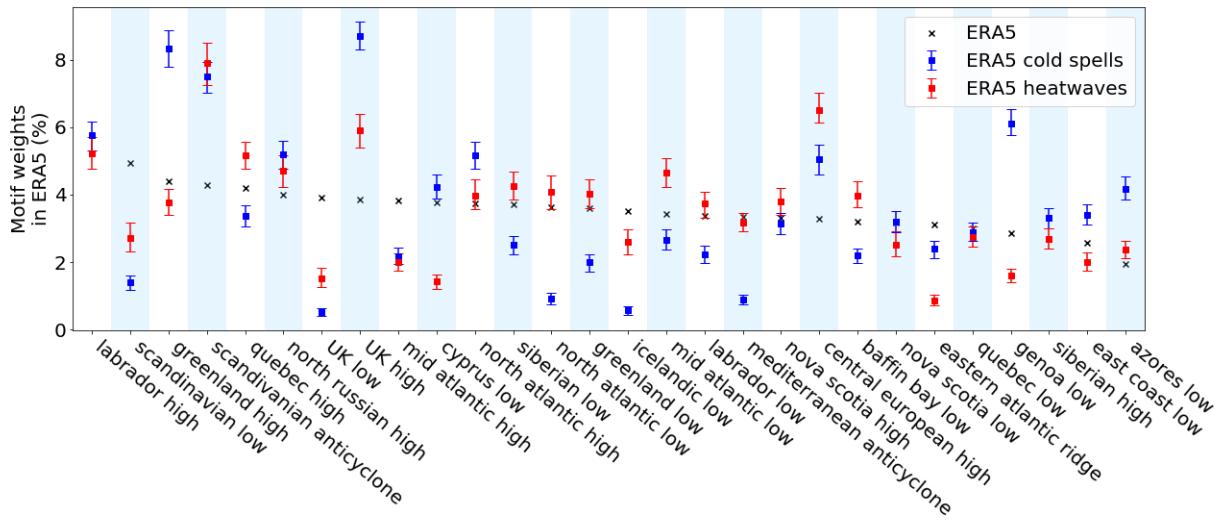
263 FIG. 3. The basis of 28 motifs learned by LDA from ERA5 SLP anomaly fields. Each motif is defined as a
 264 probability distribution over space, with positive and negative channels. The names were given based on sign
 265 and geographical location.

284 In this article, we reconstruct the average compositions of cold spells and heatwaves maps in a
 285 given model. In this case, $\mathbf{c}_k(\mathbf{P})$ is replaced with $\langle\langle\mathbf{c}_k(\mathbf{P})\rangle\rangle$, where $\langle\langle\cdot\rangle\rangle$ designates the average over
 286 all timesteps associated with a given type of extreme event, and $\|\mathbf{P}\|_1$ is replaced with $\|\langle\langle\mathbf{P}\rangle\rangle\|_1$.

287 3. Synoptic configuration of extreme events

288 We first use the decomposition into synoptic objects given by LDA to identify the atmospheric
 289 circulation patterns associated with cold spells and heatwaves. The patterns associated with extreme

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



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

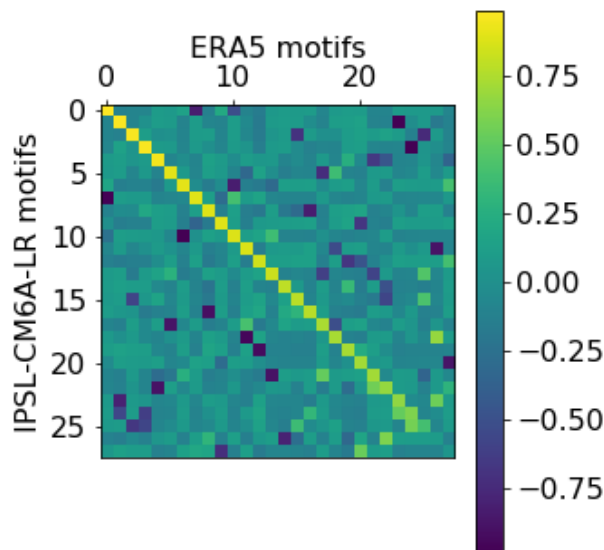
303 The synoptic configuration of extreme events is different from the average configuration of the
 304 general data. Cold spell circulation is dominated by northern anticyclones such as Greenland high,
 305 Scandinavian anticyclone and UK high, with more than 6% weights each. Correspondingly, the
 306 low pressure objects over those regions have less than half the weights they have in the general
 307 data. Genoa low is also a key motif in French cold spells, being the fourth most represented

308 motif. Its opposite, the Mediterranean anticyclone, also has during cold spells half the weights
 309 it has in general. Heatwave circulation is dominated by a smaller set of high-weights motifs,
 310 mainly consisting of Scandinavian anticyclone, and central European high. The UK high is also
 311 more prevalent during heatwaves than in general. Both types of extremes are associated with an
 312 above-average weights of Scandinavian anticyclone and of UK high.

313 4. Evaluation of model representation

314 a. Robustness of the basis

315 We first establish that a unique basis can be used to compare models with reanalysis data. Fig. 5
 316 shows the correlation matrix between the reanalysis data basis and that obtained from a run from
 317 a IPSL-CM6A-LR model, which are respectively associated with cell-motif distributions β and
 318 β' . Since motifs are defined by a spatial distribution, we use spatial correlation to measure their
 319 similarity.



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

323 The correlation matrix is obtained as follows: All fields are set to the same 1° resolution by linear
 324 interpolation. For each matrix entry, the Pearson correlation coefficient ρ_{kl} between between motif
 325 k of basis β and motif l of basis β' is computed as shown in equation (7).

$$\rho_{kl} = \frac{\overline{(\beta_k - \bar{\beta}_k)(\beta'_l - \bar{\beta}'_l)}}{\sqrt{\overline{(\beta_k - \bar{\beta}_k)^2}} \sqrt{\overline{(\beta'_l - \bar{\beta}'_l)^2}}} \quad (7)$$

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

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

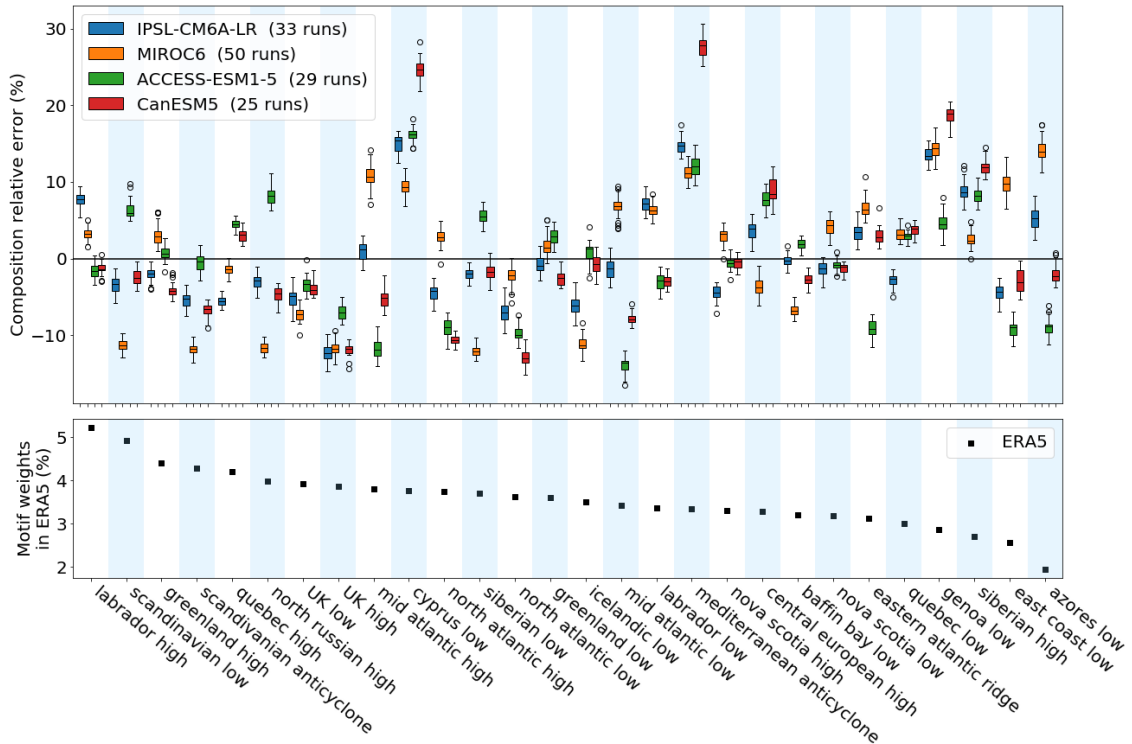
331 *b. General data case*

332 We project each run of the four models onto the motif basis learned from ERA5, then average
 333 the resulting synoptic configuration of the fields over each run. We first consider all fields in the
 334 datasets. For each run, the relative difference between the K motif weights in the model and that
 335 in the reanalysis is computed following equation (8).

$$\forall k \in [1, K], E_k = \frac{\langle \mathbf{c}_k(\mathbf{P}^{m,r}) \rangle - \langle \mathbf{c}_k(\mathbf{P}) \rangle}{\langle \mathbf{c}_k(\mathbf{P}) \rangle} \quad (8)$$

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

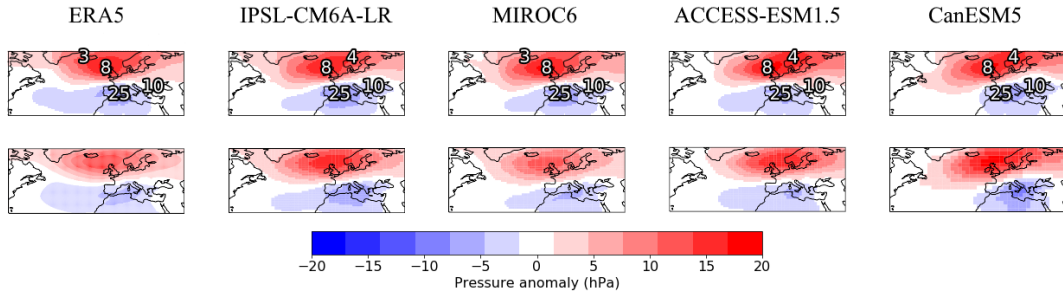
344 The median relative errors, materialized by the black lines within the boxes, are relatively small.
 345 In particular, the error is less than 15% for the eight most prevalent motifs in the reanalysis.
 346 Overall, models represent well the reanalysis synoptic configuration. Relative errors made by
 347 IPSL-CM6A-LR, MIROC6 and ACCESS-ESM1.5, which have resolutions of respectively 38×53 ,
 348 34×92 , and 39×69 are all below 20%. We note that the largest error (25%) is observed for



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

349 CanESM5, which has a resolution of 17×46 . It is possible that these larger errors could be due to
 350 its coarser resolution. Moreover, the inner variability of the models (corresponding to the width of
 351 the boxes) is typically much smaller than the error (in 96 cases out of the 112 (87.5%), the model's
 352 internal variability is lower than its bias). This shows that all runs make similar predictions and
 353 also indicates the presence of a bias inherent to each model.

354 In addition, the motifs associated with the largest relative errors tend to be the same from one
 355 model to another. A multimodel ensemble mean would therefore not eliminate these biases.
 356 The largest errors are made on motifs located on the Mediterranean region. The Cyprus low
 357 and Mediterranean anticyclone motifs are over-represented in all runs of all four models. Every
 358 model run also over-represents Genoa low and under-represents UK high and low. Finally, the
 359 Scandinavian anticyclone is the fourth most prevalent motif in the reanalysis, with an average



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

360 weight of more than 4% yet all models but ACCESS-ESM1.5 systematically under-represent it.
 361 These similarities in the model errors suggest that the origin of the errors could be common to all
 362 models.

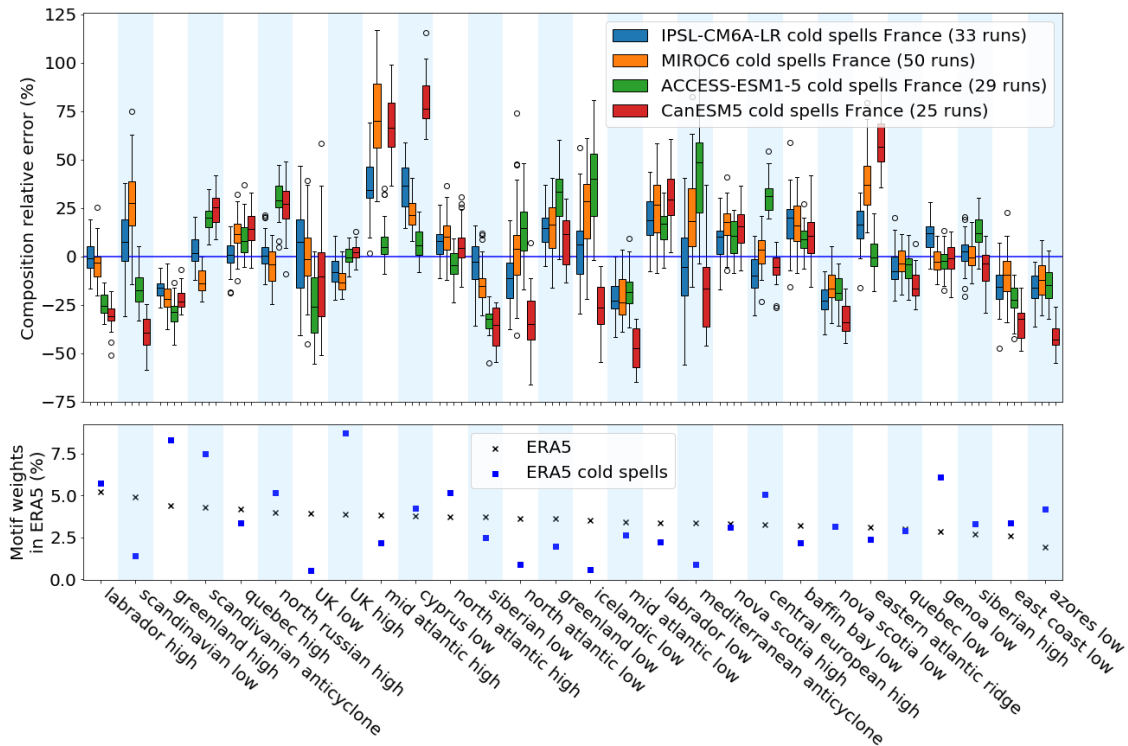
363 *c. Model representation of cold spells*

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

367 The fields corresponding to the real and the reconstructed averages are represented in Fig. 7.
 368 The real average is obtained by taking a conditional average over all daily fields associated with
 369 a cold spell. The reconstructed average is obtained from the average motif compositions of the
 370 daily fields included in the conditional average, using equation (6). To identify the most significant
 371 motifs associated with each model, the two most prevalent cyclonic and the two most prevalent
 372 anticyclonic motifs in each case are annotated on the figure.

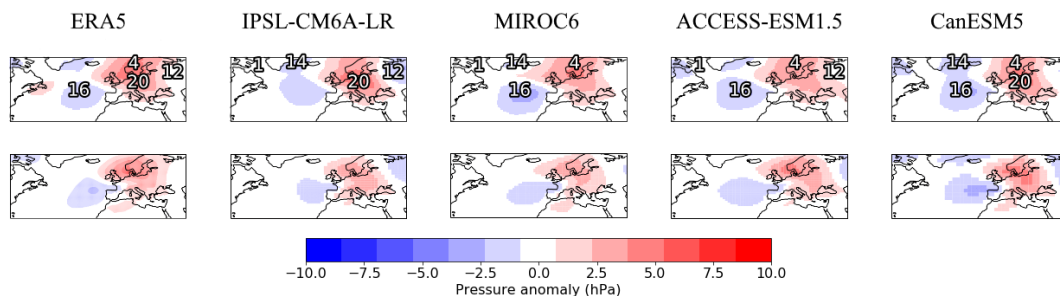
376 The overall synoptic structure associated with French cold spells consists of an anticyclonic
 377 structure in the north and a cyclonic structure in the south, with a corridor between the two
 378 slanted northeast-southwest, passing through the middle of France. For all models, the real average
 379 is generally similar to its reconstructed average, which shows that LDA captures the synoptic
 380 information contained in the real fields.

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



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

392 For a more detailed analysis, we show for each motif the relative errors in weights between the
 393 reanalyses and the models in the case of cold spells occurring in France, in Fig. 8. The biases
 394 are significantly higher for the cold extremes than for the general case. The variability among the
 395 runs of each models is also higher than for the general case. The five most prevalent reanalysis



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

396 motifs during French cold spells are UK high, Greenland high, Scandinavian anticyclone, Genoa
 397 low, Central European high. Most of these motifs are correctly represented by the models. The
 398 significantly higher weights of UK high and Genoa low during cold spells are well captured by all
 399 models with an error within the internal variability of all four models. Central European high is
 400 also well represented by all models except by ACCESS-ESM1-5 which overestimates it by 25%.
 401 The weight of Scandinavian anticyclone high is well captured by IPSL-CM6A-LR and MIROC6,
 402 while it is overestimated by 25% by the two other models. All models make about 25% error
 403 on Greenland high. Higher errors are made on less relevant motifs where the reanalysis values
 404 are lower. The most over-represented motifs are Cyprus low and Mid-Atlantic high for all models
 405 except ACCESS-ESM1.5. We note that larger errors are generally observed for the lower resolution
 406 model CAN-ESM5.

407 *d. Model representation of heatwaves*

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

413 The SLP anomaly values are weaker than in the case of cold spells. This is because heatwaves are
 414 more varied in configuration, leading to average error values closer to zero. There are differences
 415 between the real and reconstructed fields. In ERA5 and all models, the anticyclonic structure over
 416 Europe has a more crescent-like shape around the Atlantic cyclone, that changes into a arrow-like

417 shape in the LDA reconstruction. Still, the overall structure consisting of anticyclones over northern
418 and central Europe with a depression over the Atlantic is preserved by LDA reconstruction.

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

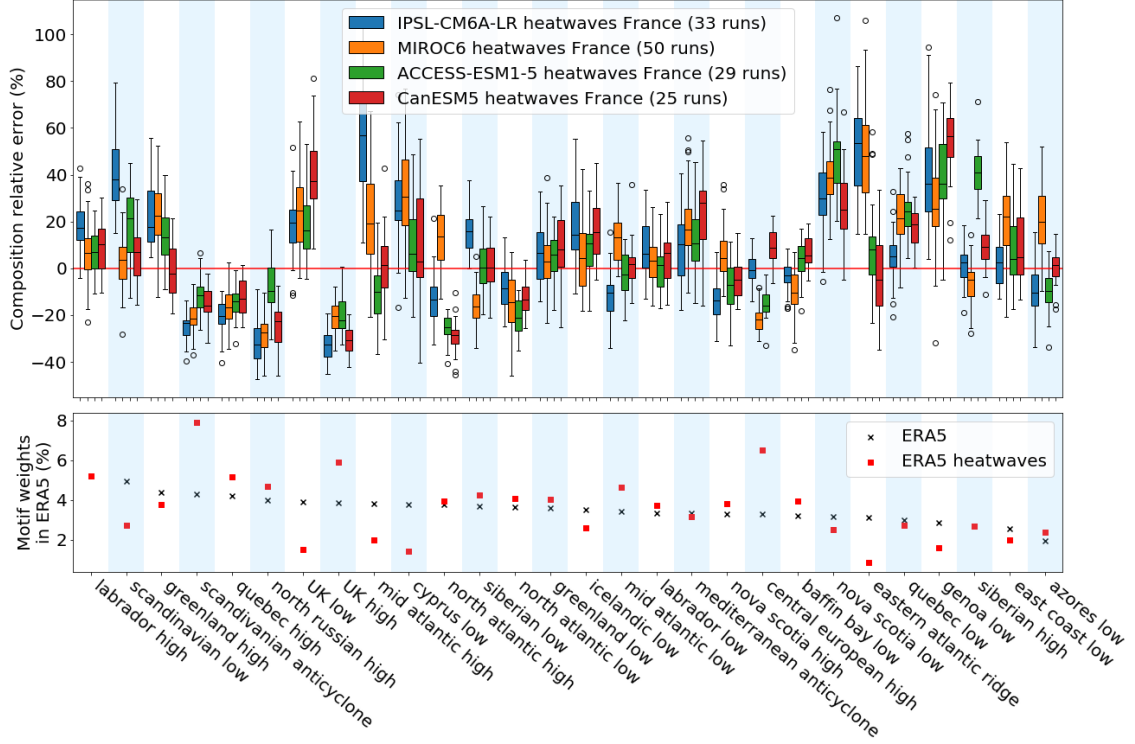
428 For a more detailed analysis, we computed relative errors in motif weights between the reanalyses
429 and the models for heatwaves occurring in France. They are shown in Fig. 10.

433 In the case of heatwaves too, model biases and internal variabilities are higher than in the general
434 case. Which motifs are or are not relevant is generally well captured by the models. However,
435 the most relevant motifs tend to be underpredicted by the models. All models except ACCESS-
436 ESM1.5 under-represent by 20% on average the contribution of the most prevalent motif, which
437 is the Scandinavian anticyclone. The second most prevalent motif, the central European high,
438 is well represented by IPSL-CM6A-LR and CanESM5 but under-represented by about 20% by
439 MIROC6 and ACCESS-ESM1.5. UK high, the third most prevalent motif, is under-represented
440 by 20% or more by almost all runs of all models. In general, motifs that have higher weights
441 than in the general case tend to be under-represented (as for instance Quebec high and north
442 Russian high), while motifs that have lower weights (UK low, Nova Scotia low, and Genoa low) are
443 over-represented. This shows that models underestimate the changes in atmospheric circulation
444 associated with heatwaves.

445 **5. Global dynamic and temperature error**

446 *a. General data case*

447 LDA provides a decomposition of circulation patterns into motifs. As seen in section 4, differ-
448 ences in motif weights provide a quantitative measure of model ability to reproduce dynamics



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

449 observed in reanalysis data. The dynamic error of run r of model m , $E_{\mathbf{P}}^{m,r}$ is defined as the sum
 450 of individual motif errors, where the error associated with each motif is defined as the absolute
 451 difference between the average weight of this motif in the model and that in the reanalysis data. It
 452 is therefore expressed as:

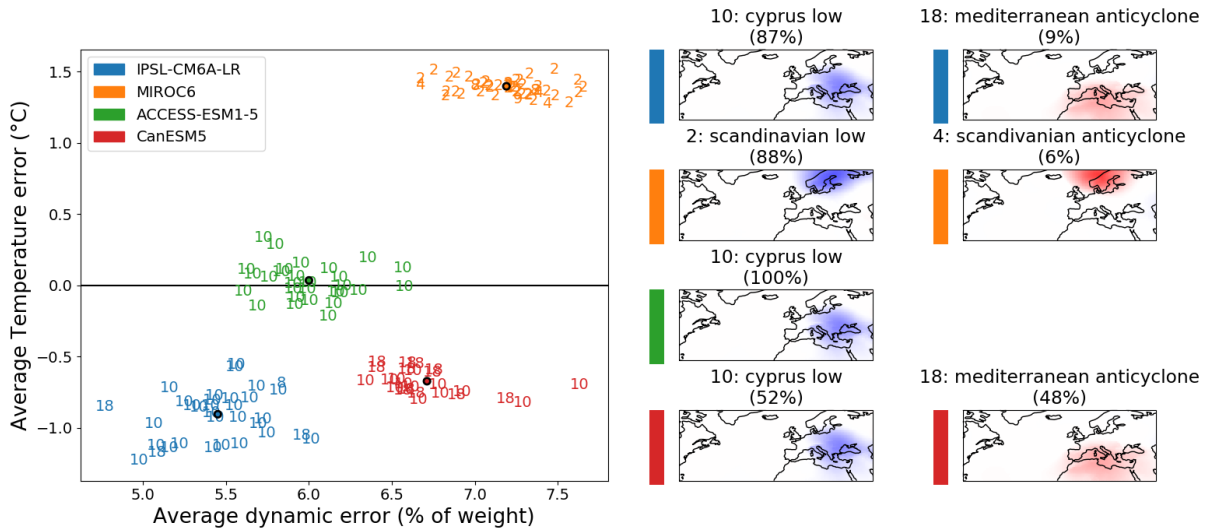
$$E_{\mathbf{P}}^{m,r} = \sum_{k=1}^K |\langle \mathbf{c}_k(\mathbf{P}^{m,r}) \rangle - \langle \mathbf{c}_k(\mathbf{P}) \rangle| \quad (9)$$

453 The dynamic error can be used to evaluate models comparatively, and produce rankings. Models
 454 can also be evaluated based on temperature error, i.e. the temperature difference between models
 455 and reanalysis data. Two sources can contribute to the temperature error: errors in the underlying
 456 thermodynamic processes (regulating the exchanges of heat), and errors on the dynamic processes
 457 (regulating the circulation of air masses in the atmosphere (Wehrli et al. (2018))). The dynamic

458 error, as previously defined, may propagate to the temperature error, but it is unclear to what
 459 extent the latter is determined by the former. Therefore, an important question is to determine
 460 whether the measure of temperature error brings additional information to the comparison of model
 461 performance. For run r of model m , the temperature error is computed as shown in equation (10),
 462 with T denoting reanalysis temperature fields, and $T^{m,r}$ those from run r of model m .

$$E_T^{m,r} = \langle \overline{T^{m,r}} \rangle - \langle \overline{T} \rangle \quad (10)$$

463 Each model run is represented as a point in the error plane $(E_{\mathbf{P}}^{m,r}, E_T^{m,r})$ shown in Fig. 11.
 464 In addition, we annotate for each run the index of the motif with the highest contribution to the
 465 dynamic error: $\max_k |\langle \mathbf{c}_k(\mathbf{P}^{m,r}) \rangle - \langle \mathbf{c}_k(\mathbf{P}) \rangle|$. For each model, we show on the right of the figure
 466 the two motifs that appear most frequently as the largest contributor to the dynamic error of a run
 467 (the proportion of runs each motif corresponds to is indicated between parentheses) - except in the
 468 case of ACCESS-ESM1.5, where the largest contributor is always Cyprus low.



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

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

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

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

494 *b. Model representation of extreme events*

495 We now consider extreme temperature events and compute the dynamic and temperature errors
 496 associated with heatwaves as well as cold spells. In that case, we eliminate the average bias, so
 497 as to only look at the component specific to extreme events. We therefore define the anomalous
 498 dynamic error $E_{\mathbf{P},ex}^{m,r}$ similarly for heatwaves and cold spells following equation (11).

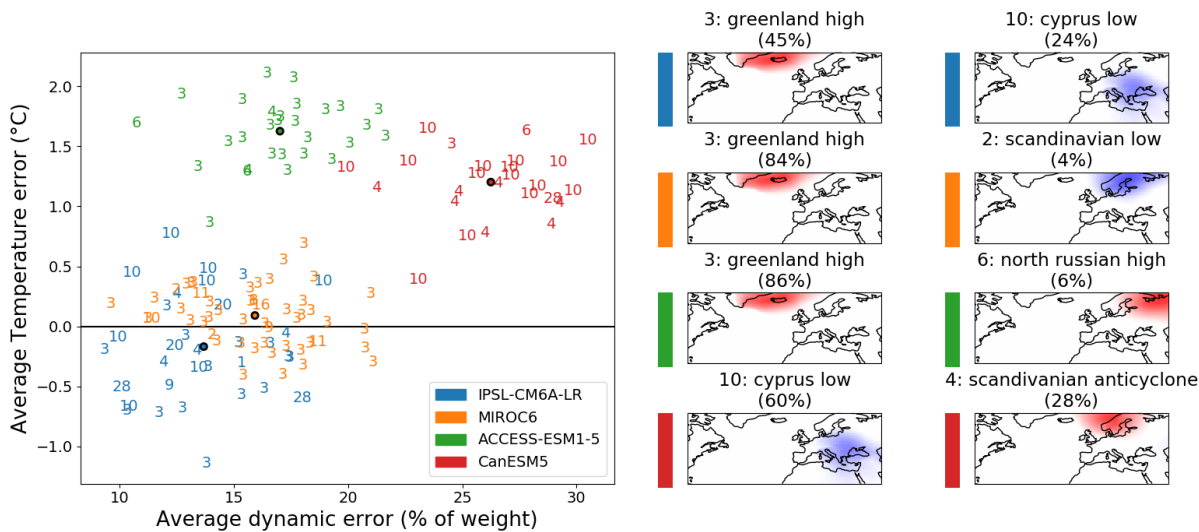
$$E_{\mathbf{P},ex}^{m,r} = \sum_{k=1}^K |\langle \langle \mathbf{c}_k(\mathbf{P}^{m,r}) \rangle \rangle - \langle \langle \mathbf{c}_k(\mathbf{P}) \rangle \rangle| - E_{\mathbf{P}}^{m,r} \quad (11)$$

499 The anomalous temperature error $E_{T,ex}^{m,r}$ is defined for heatwaves and cold spells, for run r of
 500 model m following equation (12).

$$E_{T,ex}^{m,r} = \left\langle \left\langle \overline{T^{m,r}} \right\rangle \right\rangle - \left\langle \left\langle \overline{T} \right\rangle \right\rangle - E_T^{m,r} \quad (12)$$

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

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



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

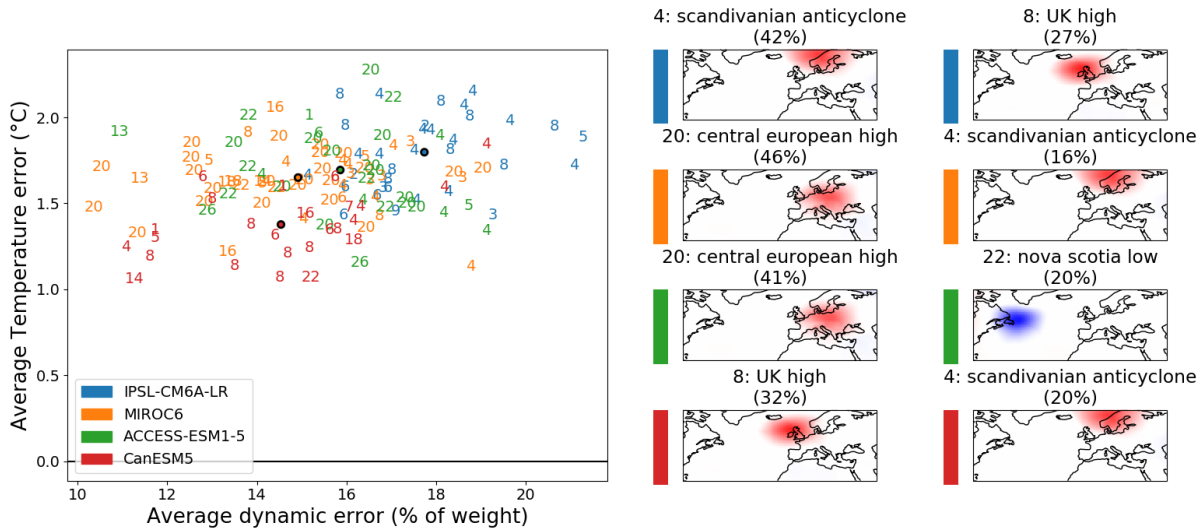
511
 512 Inner model variability is higher in the cold extreme case than in the full dataset case, for both
 513 dynamic and temperature errors. There are three distinct clusters in error space, the differences
 514 between them being bigger than internal model variabilities. The first cluster corresponds to model

515 CanESM5. It is the model with the highest dynamic error, and has a high temperature error.
516 The second cluster corresponds to model ACCESS-ESM1.5. ACCESS-ESM1.5 has the highest
517 temperature error, underpredicting the lowering of temperature due to cold spells by more than
518 1.5°C on average. Its dynamic error is comparable to that of MIROC6, and both are made in
519 majority on the same motif (Greenland high). The third cluster consists of two models, IPSL-
520 CM6A-LR and MIROC6. With the general bias removed, the temperature value from reanalysis
521 is within the internal variability of both these models. They are also associated with the lowest
522 dynamic error. This cluster appears to be closest to reanalysis. On average, IPSL-CM6A-LR has
523 a slightly lower dynamic error than MIROC6, but the difference is lower than internal variability.

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

533
534 In Fig. 13 we plot model anomalous temperature error against model anomalous dynamic error
535 in the case of heatwaves occurring in France. The inner variability of the models for heatwaves is
536 similar to the cold spell case. However, both temperature and dynamic biases associated with the
537 models are closer, so that in the 2-D error space, regions occupied by each model are overlapping.
538 All four models are associated with similar temperature errors, between +1.0 and +2.5°C - as these
539 biases are all positive, they cannot be removed by the use of a multimodel mean.

540
541 Still, some differences can be made between the models. CanESM5 has the lowest of both types
542 of error on average and IPSL-CM6A-LR the highest, but the differences are lower than model
543 internal variabilities. In addition, motifs that contribute the most to the error vary significantly
544 more from run to run than for both general data and cold spells. In particular, no motif dominates
545 the error in a majority of runs of any model, although some appear more often than others. Central
546 European high (motif 20) appears most frequently as the most significant contributor to the error
547
548
549
550



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

551 in runs of both MIROC6 and ACCESS-ESM1.5, while Scandinavian anticyclone (motif 4) makes
 552 the largest error contributions in multiple runs of IPSL-CM6A-LR, MIROC6 and CanESM5.
 553 However, we note that both Central European high and Scandinavian anticyclone are dominant
 554 motifs in heatwaves, so their presence does not reflect a significant relative motif error in the
 555 models. To sum up, all models appear to perform comparably for the representation of heat waves,
 556 and it seems difficult to identify specific error characteristics in the models.

557 6. Conclusion

558 In this paper, we use a statistical learning method called Latent Dirichlet Allocation (LDA) to
 559 study the circulation dynamics of ERA5 reanalysis data and CMIP6 general circulation models.
 560 Applied to sea-level pressure fields of the north-Atlantic region from ERA5 data, LDA yields a set of
 561 sparse latent variables called “motifs” that are recognizable localized synoptic-scale meteorological
 562 objects, such as cyclones and anticyclones. By projecting daily sea-level pressure data onto this

563 basis, we obtain the motif weights, which provides a sparse, low-dimensional representation of
564 atmospheric circulation that can be physically interpreted as the associated synoptic configuration.
565 We showed that synoptic configurations averaged over cold spells and heatwaves were both different
566 from each other and from the average taken over the full data.

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

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

584 **7. Acknowledgements**

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

587 **8. Availability statement**

588 This work makes use of the Gensim python module, which is publicly available for download
589 through the pip interface. ERA5 reanalysis data are made publicly available by the Copernicus
590 program, see <https://doi.org/10.24381/cds.143582cf>. Model datasets used in this article are sim-

591 ulations from the CMIP6 project for the “historical” experiment, by the models IPSL-CM6A-LR,
592 MIROC6, ACCESS-ESM1.5 and CanESM5. The data is publicly available thanks to the World
593 Climate Research Programme, and can be found at: <https://esgf-node.llnl.gov/search/cmip6>.

594 **9. References**

595 **References**

- 596 L. V. Alexander, X. Zhang, T. C. Peterson, J. Caesar, B. Gleason, A. M. G. Klein Tank,
597 M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. Rupa Kumar, J. Re-
598 vadekar, G. Griffiths, L. Vincent, D. B. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Tay-
599 lor, M. New, P. Zhai, M. Rusticucci, and J. L. Vazquez-Aguirre. Global observed changes
600 in daily climate extremes of temperature and precipitation. *Journal of Geophysical Re-*
601 *search: Atmospheres*, 111(D5), 2006. ISSN 2156-2202. doi: 10.1029/2005JD006290.
602 URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006290>. _eprint:
603 <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006290>.
- 604 C. Anagnostopoulou, K. Tolika, G. Lazoglou, and P. Maheras. The Exceptionally Cold January of
605 2017 over the Balkan Peninsula: A Climatological and Synoptic Analysis. *Atmosphere*, 8(12):
606 252, 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
607 Institute.
- 608
- 609 S. Ardabili, A. Mosavi, M. Dehghani, and A. R. Várkonyi-Kóczy. Deep Learning and Ma-
610 chine Learning in Hydrological Processes Climate Change and Earth Systems a System-
611 atic Review. In A. R. Várkonyi-Kóczy, editor, *Engineering for Sustainable Future*, pages
612 52–62, Cham, 2020. Springer International Publishing. ISBN 978-3-030-36841-8. doi:
613 10.1007/978-3-030-36841-8_5.
- 614 J. A. Añel, M. Fernández-González, X. Labandeira, X. López-Otero, and L. De la Torre. Impact
615 of Cold Waves and Heat Waves on the Energy Production Sector. *Atmosphere*, 8(11):209,
616 Nov. 2017. ISSN 2073-4433. doi: 10.3390/atmos8110209. URL [https://www.mdpi.com/](https://www.mdpi.com/2073-4433/8/11/209)
617 [2073-4433/8/11/209](https://www.mdpi.com/2073-4433/8/11/209). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
- 618 D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of Machine*
619 *Learning Research*, 3(null):993–1022, Mar. 2003. ISSN 1532-4435.
- 620 P. L. Bommer, M. Kretschmer, A. Hedström, D. Bareeva, and M. M.-C. Höhne. Finding the
621 right XAI Method — A Guide for the Evaluation and Ranking of Explainable AI Methods in

622 Climate Science. *Artificial Intelligence for the Earth Systems*, -1(aop), Mar. 2024. ISSN 2769-
623 7525. doi: 10.1175/AIES-D-23-0074.1. URL [https://journals.ametsoc.org/view/
624 journals/aies/aop/AIES-D-23-0074.1/AIES-D-23-0074.1.xml](https://journals.ametsoc.org/view/journals/aies/aop/AIES-D-23-0074.1/AIES-D-23-0074.1.xml). Publisher: American
625 Meteorological Society Section: Artificial Intelligence for the Earth Systems.

626 O. Boucher, J. Servonnat, A. L. Albright, O. Aumont, Y. Balkanski, V. Bastrikov, S. Bekki, R. Bon-
627 net, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, A. Caubel, F. Cheruy, F. Codron,
628 A. Cozic, D. Cugnet, F. D'Andrea, P. Davini, C. de Lavergne, S. Denvil, J. Deshayes, M. Dev-
629 illiers, A. Ducharne, J.-L. Dufresne, E. Dupont, C. Éthé, L. Fairhead, L. Falletti, S. Flavoni,
630 M.-A. Foujols, S. Gardoll, G. Gastineau, J. Ghattas, J.-Y. Grandpeix, B. Guenet, E. Guez, Li-
631 onel, E. Guilyardi, M. Guimberteau, D. Hauglustaine, F. Hourdin, A. Idelkadi, S. Joussaume,
632 M. Kageyama, M. Khodri, G. Krinner, N. Lebas, G. Levavasseur, C. Lévy, L. Li, F. Lott,
633 T. Lurton, S. Luysaert, G. Madec, J.-B. Madeleine, F. Maignan, M. Marchand, O. Marti,
634 L. Mellul, Y. Meurdesoif, J. Mignot, I. Musat, C. Ottlé, P. Peylin, Y. Planton, J. Polcher,
635 C. Rio, N. Rochetin, C. Rousset, P. Sepulchre, A. Sima, D. Swingedouw, R. Thiéblemont,
636 A. K. Traore, M. Vancoppenolle, J. Vial, J. Vialard, N. Viovy, and N. Vuichard. Presenta-
637 tion and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling
638 Earth Systems*, 12(7):e2019MS002010, 2020. ISSN 1942-2466. doi: 10.1029/2019MS002010.
639 URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002010>. _eprint:
640 <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS002010>.

641 P. W. Chan, J. L. Catto, and M. Collins. Heatwave–blocking relation change likely dominates over
642 decrease in blocking frequency under global warming. *npj Climate and Atmospheric Science*, 5
643 (1):68, 2022.

644 L. Chen, Z. Chen, Y. Zhang, Y. Liu, A. I. Osman, M. Farghali, J. Hua, A. Al-Fatesh, I. Ihara,
645 D. W. Rooney, and P.-S. Yap. Artificial intelligence-based solutions for climate change: a
646 review. *Environmental Chemistry Letters*, 21(5):2525–2557, Oct. 2023. ISSN 1610-3661. doi:
647 10.1007/s10311-023-01617-y. URL <https://doi.org/10.1007/s10311-023-01617-y>.

648 J. Cohen, D. Coumou, J. Hwang, L. Mackey, P. Orenstein, S. Totz, and E. Tziperman. S2S
649 reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal
650 forecasts. *WIREs Climate Change*, 10(2):e00567, 2019. ISSN 1757-7799. doi: 10.1002/wcc.

651 567. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.567>. eprint:
652 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.567>.

653 P. Davini and F. D'Andrea. From CMIP3 to CMIP6: Northern Hemisphere Atmospheric Block-
654 ing Simulation in Present and Future Climate. *Journal of Climate*, 33(23):10021–10038,
655 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:
656 American Meteorological Society Section: Journal of Climate.
657

658 C. R. de Freitas and E. A. Grigorieva. A comparison and appraisal of a comprehensive range
659 of human thermal climate indices. *International Journal of Biometeorology*, 61(3):487–512,
660 Mar. 2017. ISSN 1432-1254. doi: 10.1007/s00484-016-1228-6. URL <https://doi.org/10.1007/s00484-016-1228-6>.
661

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

664 V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor.
665 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design
666 and organization. *Geoscientific Model Development*, 9(5):1937–1958, May 2016. ISSN 1991-
667 959X. doi: 10.5194/gmd-9-1937-2016. URL [https://gmd.copernicus.org/articles/
668 9/1937/2016/gmd-9-1937-2016.html](https://gmd.copernicus.org/articles/9/1937/2016/gmd-9-1937-2016.html). Publisher: Copernicus GmbH.

669 W. Fang, Q. Xue, L. Shen, and V. S. Sheng. Survey on the Application of Deep Learning
670 in Extreme Weather Prediction. *Atmosphere*, 12(6):661, June 2021. ISSN 2073-4433. doi:
671 10.3390/atmos12060661. URL <https://www.mdpi.com/2073-4433/12/6/661>. Number:
672 6 Publisher: Multidisciplinary Digital Publishing Institute.

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

676 D. Faranda, G. Messori, and P. Yiou. Dynamical proxies of North Atlantic predictability and
677 extremes. *Scientific Reports*, 7(1):41278, Jan. 2017. ISSN 2045-2322. doi: 10.1038/srep41278.

678 URL <https://www.nature.com/articles/srep41278>. Number: 1 Publisher: Nature Pub-
679 lishing Group.

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

685 A. H. Fink, T. Brücher, A. Krüger, G. C. Leckebusch, J. G. Pinto, and U. Ulbrich. The 2003
686 European summer heatwaves and drought - synoptic diagnosis and impacts. *Weather*, 59(8):209–
687 216, 2004. ISSN 0043-1656. doi: 10.1256/wea.73.04. URL <https://oceanrep.geomar.de/id/eprint/30311/>. Number: 8 Publisher: Wiley-Blackwell.

688

689 P. Frich, L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. K. Tank, and
690 T. Peterson. Observed coherent changes in climatic extremes during the second half of the
691 twentieth century. *Climate Research*, 19(3):193–212, Jan. 2002. ISSN 0936-577X, 1616-
692 1572. doi: 10.3354/cr019193. URL [https://www.int-res.com/abstracts/cr/v19/n3/
693 p193-212/](https://www.int-res.com/abstracts/cr/v19/n3/p193-212/).

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

698 S. Gardoll and O. Boucher. Classification of tropical cyclone containing images using a
699 convolutional neural network: performance and sensitivity to the learning dataset. *Geo-*
700 *scientific Model Development*, 15(18):7051–7073, Sept. 2022. ISSN 1991-959X. doi:
701 10.5194/gmd-15-7051-2022. URL [https://gmd.copernicus.org/articles/15/7051/
702 2022/](https://gmd.copernicus.org/articles/15/7051/2022/). Publisher: Copernicus GmbH.

703 D. E. Hanley, M. A. Bourassa, J. J. O’Brien, S. R. Smith, and E. R. Spade. A Quan-
704 titative Evaluation of ENSO Indices. *Journal of Climate*, 16(8):1249–1258, Apr. 2003.
705 ISSN 0894-8755, 1520-0442. doi: 10.1175/1520-0442(2003)16<1249:AQEOEI>2.0.CO;2.

706 URL https://journals.ametsoc.org/view/journals/clim/16/8/1520-0442_2003_
707 [16_1249_aqoei_2.0.co_2.xml](https://journals.ametsoc.org/view/journals/clim/16/8/1520-0442_2003_16_1249_aqoei_2.0.co_2.xml). Publisher: American Meteorological Society Section:
708 Journal of Climate.

709 H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas,
710 C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo,
711 P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Dia-
712 mantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy,
713 R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti,
714 P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. The ERA5 global reanal-
715 ysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. ISSN
716 1477-870X. doi: 10.1002/qj.3803. URL [https://onlinelibrary.wiley.com/doi/abs/](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803)
717 [10.1002/qj.3803](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803>.

718 M. Hoffman, F. Bach, and D. Blei. Online Learning for Latent Dirichlet Allocation.
719 In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates,
720 Inc., 2010. URL [https://proceedings.neurips.cc/paper_files/paper/2010/hash/](https://proceedings.neurips.cc/paper_files/paper/2010/hash/71f6278d140af599e06ad9bf1ba03cb0-Abstract.html)
721 [71f6278d140af599e06ad9bf1ba03cb0-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2010/hash/71f6278d140af599e06ad9bf1ba03cb0-Abstract.html).

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

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

729 T. N. Krishnamurti. The subtropical jet stream of winter. *Journal of the Atmospheric Sciences*, 18
730 (2):172–191, 1961.

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

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

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

740 P. Lucas-Picher, D. Argüeso, E. Brisson, Y. Trambly, P. Berg, A. Lemonsu, S. Kotlarski, and
741 C. Caillaud. Convection-permitting modeling with regional climate models: Latest develop-
742 ments and next steps. *WIREs Climate Change*, 12(6):e731, 2021. ISSN 1757-7799. doi: 10.
743 1002/wcc.731. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.731>.
744 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.731>.

745 A. R. Lupo. Atmospheric blocking events: A review. *Annals of the New York Academy of Sciences*,
746 1504(1):5–24, 2021.

747 J. A. López-Bueno, M. Á. Navas-Martín, J. Díaz, I. J. Mirón, M. Y. Luna, G. Sánchez-Martínez,
748 D. Culqui, and C. Linares. The effect of cold waves on mortality in urban and rural areas
749 of Madrid. *Environmental Sciences Europe*, 33(1):72, June 2021. ISSN 2190-4715. doi:
750 10.1186/s12302-021-00512-z. URL <https://doi.org/10.1186/s12302-021-00512-z>.

751 M. McCarthy, N. Christidis, N. Dunstone, D. Fereday, G. Kay, A. Klein-Tank, J. Lowe,
752 J. Petch, A. Scaife, and P. Stott. Drivers of the UK summer heatwave of
753 2018. *Weather*, 74(11):390–396, 2019. ISSN 1477-8696. doi: 10.1002/wea.
754 3628. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wea.3628>. _eprint:
755 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wea.3628>.

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

- 761 P.-A. Michelangeli, R. Vautard, and B. Legras. Weather Regimes: Recurrence and
762 Quasi Stationarity. *Journal of the Atmospheric Sciences*, 52(8):1237–1256, Apr. 1995.
763 ISSN 0022-4928, 1520-0469. doi: 10.1175/1520-0469(1995)052<1237:WRRMQS>2.0.CO;2.
764 URL [https://journals.ametsoc.org/view/journals/atsc/52/8/1520-0469_1995_](https://journals.ametsoc.org/view/journals/atsc/52/8/1520-0469_1995_052_1237_wrraqs_2_0_co_2.xml)
765 [052_1237_wrraqs_2_0_co_2.xml](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:
766 Journal of the Atmospheric Sciences.
- 767 J. Mignot, F. Hourdin, J. Deshayes, O. Boucher, G. Gastineau, I. Musat, M. Vancoppenolle,
768 J. Servonnat, A. Caubel, F. Ch eruy, S. Denvil, J.-L. Dufresne, C. Eth e, L. Fairhead, M.-
769 A. Foujols, J.-Y. Grandpeix, G. Levavasseur, O. Marti, M. Menary, C. Rio, C. Rousset, and
770 Y. Silvy. The Tuning Strategy of IPSL-CM6A-LR. *Journal of Advances in Modeling Earth*
771 *Systems*, 13(5):e2020MS002340, 2021. ISSN 1942-2466. doi: 10.1029/2020MS002340.
772 URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002340>. _eprint:
773 <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002340>.
- 774 R. O’Loughlin, D. Li, and T. O’Brien. Moving beyond post-hoc XAI: Lessons learned
775 from dynamical climate modeling. *EGUsphere*, pages 1–24, Jan. 2024. doi: 10.5194/
776 egusphere-2023-2969. URL [https://egusphere.copernicus.org/preprints/2024/](https://egusphere.copernicus.org/preprints/2024/egusphere-2023-2969/)
777 [egusphere-2023-2969/](https://egusphere.copernicus.org/preprints/2024/egusphere-2023-2969/). Publisher: Copernicus GmbH.
- 778 K. Papagiannaki, K. Lagouvardos, V. Kotroni, and G. Papagiannakis. Agricultural losses related to
779 frost events: use of the 850 hPa level temperature as an explanatory variable of the damage cost.
780 *Natural Hazards and Earth System Sciences*, 14(9):2375–2386, Sept. 2014. ISSN 1561-8633.
781 doi: 10.5194/nhess-14-2375-2014. URL [https://nhess.copernicus.org/articles/14/](https://nhess.copernicus.org/articles/14/2375/2014/)
782 [2375/2014/](https://nhess.copernicus.org/articles/14/2375/2014/). Publisher: Copernicus GmbH.
- 783 R. Řeh r ek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In
784 *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages
785 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.
- 786 M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat.
787 Deep learning and process understanding for data-driven Earth system science. *Nature*, 566
788 (7743):195–204, Feb. 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-0912-1. URL <https://www.nature.com/articles/s41586-019-0912-1>. Publisher: Nature Publishing Group.
789

- 790 X. Ren, L. Li, Y. Yu, Z. Xiong, S. Yang, W. Du, and M. Ren. A Simplified Climate Change Model
791 and Extreme Weather Model Based on a Machine Learning Method. *Symmetry*, 12(1):139,
792 Jan. 2020. ISSN 2073-8994. doi: 10.3390/sym12010139. URL [https://www.mdpi.com/
793 2073-8994/12/1/139](https://www.mdpi.com/2073-8994/12/1/139). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.
- 794 D. F. Rex. Blocking Action in the Middle Troposphere and its Effect upon Regional Climate.
795 *Tellus*, 2(4):275–301, Jan. 1950. ISSN 0040-2826. doi: 10.3402/tellusa.v2i4.8603. URL
796 <https://doi.org/10.3402/tellusa.v2i4.8603>. Publisher: Taylor & Francis eprint:
797 <https://doi.org/10.3402/tellusa.v2i4.8603>.
- 798 D. Rodrigues, M. C. Alvarez-Castro, G. Messori, P. Yiou, Y. Robin, and D. Faranda. Dynamical
799 Properties of the North Atlantic Atmospheric Circulation in the Past 150 Years in CMIP5 Models
800 and the 20CRv2c Reanalysis. *Journal of Climate*, 31(15):6097–6111, Aug. 2018. ISSN 0894-
801 8755, 1520-0442. doi: 10.1175/JCLI-D-17-0176.1. URL [https://journals.ametsoc.
802 org/view/journals/clim/31/15/jcli-d-17-0176.1.xml](https://journals.ametsoc.org/view/journals/clim/31/15/jcli-d-17-0176.1.xml). Publisher: American Meteoro-
803 logical Society Section: Journal of Climate.
- 804 E. Rousi, K. Kornhuber, G. Beobide-Arsuaga, F. Luo, and D. Coumou. Accelerated western
805 European heatwave trends linked to more-persistent double jets over Eurasia. *Nature Com-
806 munications*, 13(1):3851, July 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31432-y.
807 URL <https://www.nature.com/articles/s41467-022-31432-y>. Number: 1 Publisher:
808 Nature Publishing Group.
- 809 S. Salcedo-Sanz, J. Pérez-Aracil, G. Ascenso, J. Del Ser, D. Casillas-Pérez, C. Kadow, D. Fister,
810 D. Barriopedro, R. García-Herrera, M. Giuliani, and A. Castelletti. Analysis, characterization,
811 prediction, and attribution of extreme atmospheric events with machine learning and deep
812 learning techniques: a review. *Theoretical and Applied Climatology*, 155(1):1–44, Jan. 2024.
813 ISSN 1434-4483. doi: 10.1007/s00704-023-04571-5. URL [https://doi.org/10.1007/
814 s00704-023-04571-5](https://doi.org/10.1007/s00704-023-04571-5).
- 815 A. A. Scaife, T. Woollings, J. Knight, G. Martin, and T. Hinton. Atmospheric Blocking and Mean
816 Biases in Climate Models. *Journal of Climate*, 23(23):6143–6152, Dec. 2010. ISSN 0894-
817 8755, 1520-0442. doi: 10.1175/2010JCLI3728.1. URL <https://journals.ametsoc.org/>

818 view/journals/clim/23/23/2010jcli3728.1.xml. Publisher: American Meteorological
819 Society Section: Journal of Climate.

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

829 N. C. Stenseth, G. Ottersen, J. W. Hurrell, A. Mysterud, M. Lima, K. Chan, N. G. Yoccoz,
830 and B. Ådlandsvik. Studying climate effects on ecology through the use of climate indices:
831 the North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proceedings of the*
832 *Royal Society of London. Series B: Biological Sciences*, 270(1529):2087–2096, Oct. 2003.
833 doi: 10.1098/rspb.2003.2415. URL [https://royalsocietypublishing.org/doi/abs/](https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2003.2415)
834 [10.1098/rspb.2003.2415](https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2003.2415). Publisher: Royal Society.

835 N. C. Swart, J. N. S. Cole, V. V. Kharin, M. Lazare, J. F. Scinocca, N. P. Gillett, J. Anstey,
836 V. Arora, J. R. Christian, S. Hanna, Y. Jiao, W. G. Lee, F. Majaess, O. A. Saenko, C. Seiler,
837 C. Seinen, A. Shao, M. Sigmond, L. Solheim, K. von Salzen, D. Yang, and B. Winter. The
838 Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12
839 (11):4823–4873, Nov. 2019. ISSN 1991-959X. doi: 10.5194/gmd-12-4823-2019. URL [https://](https://gmd.copernicus.org/articles/12/4823/2019/)
840 gmd.copernicus.org/articles/12/4823/2019/. Publisher: Copernicus GmbH.

841 H. Tatebe, T. Ogura, T. Nitta, Y. Komuro, K. Ogochi, T. Takemura, K. Sudo, M. Sekiguchi, M. Abe,
842 F. Saito, M. Chikira, S. Watanabe, M. Mori, N. Hirota, Y. Kawatani, T. Mochizuki, K. Yoshimura,
843 K. Takata, R. O’ishi, D. Yamazaki, T. Suzuki, M. Kurogi, T. Kataoka, M. Watanabe, and
844 M. Kimoto. Description and basic evaluation of simulated mean state, internal variability, and
845 climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7):2727–2765, July 2019.

846 ISSN 1991-959X. doi: 10.5194/gmd-12-2727-2019. URL [https://gmd.copernicus.org/
847 articles/12/2727/2019/](https://gmd.copernicus.org/articles/12/2727/2019/). Publisher: Copernicus GmbH.

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

854 G. J. van Oldenborgh, S. Drijfhout, A. van Ulden, R. Haarsma, A. Sterl, C. Severijns, W. Hazeleger,
855 and H. Dijkstra. Western Europe is warming much faster than expected. *Climate of the Past*,
856 5(1):1–12, Jan. 2009. ISSN 1814-9324. doi: 10.5194/cp-5-1-2009. URL [https://cp.
857 copernicus.org/articles/5/1/2009/](https://cp.copernicus.org/articles/5/1/2009/). Publisher: Copernicus GmbH.

858 R. Vautard. Multiple Weather Regimes over the North Atlantic: Analysis of Pre-
859 cursors and Successors. *Monthly Weather Review*, 118(10):2056–2081, Oct. 1990.
860 ISSN 1520-0493, 0027-0644. doi: 10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;
861 2. URL [https://journals.ametsoc.org/view/journals/mwre/118/10/1520-0493_
862 1990_118_2056_mwrotn_2_0_co_2.xml](https://journals.ametsoc.org/view/journals/mwre/118/10/1520-0493_1990_118_2056_mwrotn_2_0_co_2.xml). Publisher: American Meteorological Society Sec-
863 tion: Monthly Weather Review.

864 R. Vautard, J. Cattiaux, T. Happé, J. Singh, R. Bonnet, C. Cassou, D. Coumou, F. D’Andrea,
865 D. Faranda, E. Fischer, A. Ribes, P. Yiou, and S. Sippel. Heat extremes in Western Europe are
866 increasing faster than simulated due to missed atmospheric circulation trends, Jan. 2023a. URL
867 <https://hal.science/hal-03937057>.

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

- 874 J. M. Wallace and P. V. Hobbs. *Atmospheric science: an introductory survey*, volume 92. Elsevier,
875 2006.
- 876 K. Wehrli, B. P. Guillod, M. Hauser, M. Leclair, and S. I. Seneviratne. Assessing the
877 Dynamic Versus Thermodynamic Origin of Climate Model Biases. *Geophysical Re-*
878 *search Letters*, 45(16):8471–8479, 2018. ISSN 1944-8007. doi: 10.1029/2018GL079220.
879 URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL079220>. eprint:
880 <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL079220>.
- 881 V. Weilhhammer, J. Schmid, I. Mittermeier, F. Schreiber, L. Jiang, V. Pastuhovic, C. Herr, and
882 S. Heinze. Extreme weather events in europe and their health consequences – A systematic
883 review. *International Journal of Hygiene and Environmental Health*, 233:113688, Apr. 2021.
884 ISSN 1438-4639. doi: 10.1016/j.ijheh.2021.113688. URL [https://www.sciencedirect.](https://www.sciencedirect.com/science/article/pii/S1438463921000018)
885 [com/science/article/pii/S1438463921000018](https://www.sciencedirect.com/science/article/pii/S1438463921000018).
- 886 T. Ziehn, M. A. Chamberlain, R. M. Law, A. Lenton, R. W. Bodman, M. Dix, L. Stevens, Y.-P. Wang,
887 J. Srbinovsky, T. Ziehn, M. A. Chamberlain, R. M. Law, A. Lenton, R. W. Bodman, M. Dix,
888 L. Stevens, Y.-P. Wang, and J. Srbinovsky. The Australian Earth System Model: ACCESS-
889 ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, 70(1):193–214, Aug. 2020.
890 ISSN 2206-5865, 2206-5865. doi: 10.1071/ES19035. URL [https://www.publish.csiro.](https://www.publish.csiro.au/es/ES19035)
891 [au/es/ES19035](https://www.publish.csiro.au/es/ES19035). Publisher: CSIRO PUBLISHING.