Changes in the dynamical properties of the North Atlantic atmospheric circulation in the past 150 years

David Rodrigues, M Carmen Alvarez-Castro, Gabriele Messori, Pascal Yiou, Yoann Robin, Davide Faranda

Changes in the dynamical properties of the North Atlantic atmospheric circulation in the past 150 years

David Rodrigues[1], M. Carmen Alvarez-Castro[1], Gabriele Messori[2],

Pascal Yiou[1], Yoann Robin[1], Davide Faranda[1*]

[1]*Laboratoire de Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ,IPSL, Universite Paris-Saclay, 91191 Gif-sur-Yvette, France.*

[2]*Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden.*

[*]*Corresponding author address:* Laboratoire de Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ,IPSL, Universite Paris-Saclay, 91191 Gif-sur-Yvette, France.

E-mail: davide.faranda@lsce.ipsl.fr

ABSTRACT

It is of fundamental importance to evaluate the ability of climate models to capture the large-scale atmospheric circulation patterns. In the context of a rapidly changing climate, it is equally crucial to quantify the robustness of the modeled changes in the large-scale atmospheric dynamics. Here we approach this problem from an innovative point of view based on dynamical systems theory. We characterize the atmospheric circulation over the North Atlantic in the CMIP5 historical simulations (1851 to 2000) in terms of two instantaneous metrics: local dimension of the attractor and stability of trajectories. We then use these metrics to compare the models to the 20CR reanalysis over the same historical period. The comparison suggests that: i) all the models capture the mean attractor properties and models with finer grids perform better; ii) extremes of the dynamical systems metrics match the same large-scale patterns in most of the models; iii) changes in the attractor properties observed for the 20CR reanalysis - studied by dividing the 1851-2000 period into 3 subperiods of 50 years each - are not reproduced by the models; iv) some models present significant changes in the dynamical systems metrics over time but there is no agreement on the direction and on the intensity of the shifts.

1. Introduction

One of the main sources of uncertainty in determining the impact of climate change on extreme events is the forced response of atmospheric dynamics (Shepherd 2014; Field 2012). While changes in observables such as surface temperature are easily diagnosed, shifts in the mid-latitude atmospheric patterns have proved very difficult to quantify. Some advances have been made by focussing on specific features such as atmospheric blocking (Kay et al. 2015; Cassou and Cattiaux 2016; Faranda et al. 2016b), which in turn influence the occurrence of European cold spells and heat waves, but the broader appreciation of circulation changes is still unsatisfactory. Here we address this knowledge gap by using a dynamical systems framework. We illustrate the power of such an approach by considering the well-known Lorenz (1963) system, a conceptual model of atmospheric convection consisting of three differential equations:

$$\dot{x} = \sigma(y-x) \quad \dot{y} = rx - y - xz \quad \dot{z} = xy - bz \tag{1}$$

where x, y, z represent respectively the convection strength, the temperature difference between the surface and the top of the troposphere and the asymmetry of the convection cells. The parameters σ, r are the Prandtl and the Rayleigh numbers, while b is a ratio of critical parameters. A trajectory of the Lorenz (1963) attractor is shown in blue in figure 1. The figure consists of 2000 points obtained by iterating the Lorenz equations with $\Delta t \simeq 0.035$, $\sigma = 28$, $r = 10$, $b = 8/3$ with a Runge Kutta scheme of order 4.

To study the effects of an external forcing, we increase σ by 2% with respect to the classical value. Figure 1 shows a trajectory at $\sigma = 28.5$ in magenta. The magenta trajectory favours higher values of the variable z, but the changes relative to the original trajectory depend on the point being considered: some points are not displaced, while some others are mapped elsewhere.

Assuming no knowledge of the system other than the trajectories' paths, how could we determine whether they both belong to the Lorenz attractors with two different forcings? To answer this question we would need: i) to measure the dynamical properties of an ensemble of trajectories representing the two configurations; ii) to estimate the distance between the obtained trajectories and determine if the shift has changed the properties of the points in a detectable way.

The atmospheric equivalent of a point on the Lorenz attractor is the ensemble of instantaneous fields describing the atmosphere at a time t. To study the atmospheric circulation over the North Atlantic we will focus on a single field: the sea-level pressure (SLP) over this region. The SLP field reflects the major modes of variability affecting the North Atlantic (Hurrell 1995; Moore et al. 2013) and can further be used to diagnose a wealth of other atmospheric features, ranging from teleconnection patterns to storm track activity to atmospheric blocking e.g. (Rogers 1997; Comas-Bru and McDermott 2014). The trajectories of our dynamical systems are the succession of daily SLP fields from 27 CMIP5 models and the 20CR reanalysis over the period 1851 to 2000. In order to measure changes in the systems, one must be able to specify at each point (each day) the local (daily) dynamical properties and track their evolution. Recent contributions to dynamical systems analysis have proven that local properties of the trajectories are characterized by two quantities: the local dimension and stability of the field considered (Lucarini et al. 2016; Faranda et al. 2017). They correspond respectively to the rarity and the typical persistence of the configuration. Faranda et al. (2017) have also shown that these two metrics can be connected to the predictability of a given atmospheric state and that their extremes match climate extremes.

In this work we will first assess whether the models and reanalysis present similar attractor properties over the full time period considered. To do this, we compute daily values for the dimension

and stability of the SLP fields and study their average and extreme properties. Next, we study the attractors for three sub-periods of 50 years each and quantify their changes across the periods. We then compare the changes seen in the models to those observed in the reanalysis. Finally, we use statistical mechanics arguments to attribute the changes observed in the reanalysis to greenhouse forcing.

2. Data & Methods

We use daily model output from the historical simulations of 27 CMIP5 models (see Table 1). The data is publicly available from the CMIP5 archive (Taylor et al. 2012). We then compare these to the 20th Century Reanalysis (20CRv2c) ensemble mean dataset (Compo et al. 2011). The analysis focuses on the region $22.5°N - 70°N$ and $80°W - 50°E$.

In order to compute the dynamical systems metrics, we combine the statistical tools of extreme value theory with the results obtained by Freitas et al. (2010) for Poincaré recurrences. The parameters mentioned in the introduction (local dimension d and stability θ) are computed for the points ζ on the attractor obtained as sequence of states of the system. The dynamical indicators are linked to the probability $\mathscr{P}$ that a trajectory $x(t)$ of the distances of a trajectory emerging from a sphere of center ζ and diameter 2ε, i.e. the recurrence rate of the configuration ζ. We briefly outline the physical meaning of these quantities and the way they are computed below.

(i) Local Dimensions: The Freitas et al. (2010) theorem and its modification in Lucarini et al. (2012) states that the probability $\mathscr{P}$ for chaotic attractors is a generalized Pareto distribution (Pickands III 1975). We first compute the distance δ between the SLP field ζ and all other observations along the trajectory. We then weight the time series of the distance:

$$g(x(t)) = -\log(\delta(x(t), \zeta)).$$

The reason for taking the logarithm is explained by Collet and Eckmann (2009): in the dynamical system set-up the logarithm increases the discrimination of small values of $\delta(x,y)$ which correspond to large values of $g(x(t))$. The generalized Pareto distribution then reduces to:

$$\mathscr{P}(g(x(t)) > q, \zeta) \simeq \exp(-[x - \mu(\zeta)]/\beta)$$

namely an exponential law whose parameters μ and σ depend on the point ζ chosen on the attractor. Remarkably, $\beta(\zeta) = 1/d(\zeta)$, where $d(\zeta)$ is the local dimension around the point ζ. This result has recently been applied to SLP fields in Faranda et al. (2017).

(ii) Local Stability: θ, the inverse of the residence time within a neighborhood of the configuration, is exactly the extremal index introduced in extreme value theory to measure clustering (Freitas et al. 2012; Faranda et al. 2016a). As in the extreme value theory, θ varies between 0 and 1. The value $\theta = 0$ corresponds to a stable fixed point of the dynamics where the observation ζ is repeated infinite times (as for a pendulum left in its equilibrium position). This is of course never observed in the atmospheric dynamics. A value of $\theta = 1$ indicates a point immediately leaving the neighborhood of ζ. Since θ is the inverse of a persistence time, it depends on the Δt used. If Δt is too large, the time dependence structure is hidden and θ will to be close to 1. If Δt is too small, θ is close to zero. In Faranda et al. (2017) it has been observed that θ for SLP fields over the North Altantic is between 0.3 and 0.5, when $\Delta t = 1$ day. In this work we use the same Δt.

Figure 2 illustrates of the meaning of the indicators: the local dimension d is the number of degrees of freedom needed to describe the dynamics of the system linearized around the state ζ and it is therefore proportional to the number of possible states resulting from ζ. The inverse of the persistence time θ is linked to the probability that the trajectory follows a path where each field resembles those of the previous and subsequent days.

Before beginning the analysis, it is necessary to outline how the method of recurrence deals with changes in the attractor. There are few theoretical results on non-stationary statistics of dynamical systems, as well as on non-stationary extreme value theory. Luckily, the recurrence technique also allows to bypass most of the technical difficulties linked to non-stationarity because the dynamical properties are measured with respect to each single state of the attractor. If the change affects the neighbourhood of a state, it will change its dynamical properties d and θ. If most of the states are affected by the changes in the dynamics, then the average dimension of the attractor and the average persistence will change accordingly.

In order to test this idea, we again consider the two Lorenz (1963) systems discussed in the Introduction and perform two sets of 30 realizations (trajectories) at $\sigma = 28$ and $\sigma = 28.5$, each of them consisting of 50000 points. These values correspond respectively to the size of the CMIP5 ensemble (about 30) and to the number of days in the period 1850-2000 (about 50000). If the method of recurrence is capable to distinguish between the $\sigma = 28$ and $\sigma = 28.5$ Lorenz attractors, then the d and θ distributions of the two sets of realizations should be significantly different.

First, we have to define a metric to compare (d, θ) distributions. The simplest idea is to compute the median of the d, θ distribution for each realization and verify that the clouds of median centroids for $\sigma = 28$ and for $\sigma = 28.5$ are separated. This is shown in the top panel of Figure 3. As a further test, we can compute the empirical probability density functions $\Lambda(\sigma_{28})$ and $\Lambda(\sigma_{28.5})$ of the pairwise distances between the d, θ median-based centroids. These distributions should be significantly different from the $\Lambda(\sigma_{28})/\Lambda(\sigma_{28.5}))$ distribution obtained by mixing together the two sets of realizations. The visual inspection of Λ distributions (Figure 3-bottom panel) suggests that distances computed when mixing the $\sigma = 28$ and $\sigma = 28.5$ realizations are generally higher than those computed by realizations of the same attractor. This claim is statistically supported by the Kolmogorov-Smirnov test (Von Storch 1999) results reported in Table 2.

Table 2 also contains the results of the Kolmogorov-Smirnov test obtained when considering the Wasserstein distances $\mathscr{W}_2$ (Villani 2008) between the full d, θ distributions. As described by Robin et al. (2017), the Wasserstein distance is the proper tool to measure distance in multivariate set-ups so that our median-based indicator should be tested against $\mathscr{W}_2$. The test results suggest that the median-based centroids are good proxies of the Wasserstein distances. We will see in the next section that this result also hold for the CMIP5 ensemble. Since in the climate system we do not have distinct trajectories (one before and one after climate change) we will divide the dataset in three different periods imagining three separate trajectories obtained under different (greenhouse) forcings.

3. Aggregate analysis of model and reanalysis attractors

We begin the analysis of the daily SLP fields from 1851 to 2000 by presenting the scatterplot of d versus θ for the 20CR reanalysis (Figure 4). The average of d is proportional to the

number of degrees of freedom needed to represent the systems' dynamics while the average of θ is the inverse of the mean persistence time. Maxima (minima) of d correspond to the most complex (simple) trajectories of the system. Maxima (minima) of θ correspond to the most unstable (stable) trajectories (Messori et al. 2017). Such extremes are associated to specific weather patterns that closely resemble the canonical North-Atlantic weather regimes. The top panels in Figure 4 show the composite SLP anomalies for days beyond the 0.98 and 0.02 quantiles of the d and θ distributions. Maxima of θ reproduce an Atlantic Ridge pattern, while minima of θ correspond to a negative North Atlantic Oscillation (NAO) phase. Similarly, maxima of d correspond to a Blocking pattern and minima to a positive NAO (Faranda et al. 2017).

We next compare the (d, θ) bivariate histograms obtained for the 20CR with those computed for the CMIP5 models (Figure 5-right). Two different behaviors emerge: some of the models (e.g. CMCC-CMS reported in Figure 5-centre) yeld a single mode distribution resembling that obtained for the 20CR ; other models show bimodal distributions (e.g. the IPCC-CM5A) histograms reported in Figure 5-bottom. We find the different behaviors to be related to the seasonal cycle: in Figure 5-left, we plot the (d, θ) diagrams for the same models by coloring each point according to the month of the year. In the 20CR and the CMCC-CMS model, the different seasons are spread across the cloud, although maxima of θ mostly occur in winter. The IPSL-CM5A displays a much stronger seasonal discrimination, with two distinct (d, θ) clouds for the winter and for the summer seasons corresponding to the different modes of the bivariate histograms. This implies that both the bulk statistics and the extremes are modified by the seasonal cycle.

Given the variety of the possible behaviors, we will analyze separately the mean and the extreme behavior of the dynamical properties. We report the aggregate analysis in Table 3 and in Figure 7. The dots correspond to the centres of the ellipses and represent the median values for each model; the semiaxes correspond to the standard deviation of the mean. Models are numbered as in Tables 1, 3 and are ordered by increasing horizontal resolution. In Table 3 we also provide the distances $\delta(d)$ and $\delta(\theta)$ from the median of the 20CR, the relative distances $R(d) = \delta(d)/\max\delta(d)$ and $R(\theta) = \delta(\theta)/\max\delta(\theta)$ with respect to the farthest model and a global score $R_{tot} = (R(d) + R(\theta))/2$. To check the validity of this global score, we compare R_{tot} with the Wasserstein distance $\mathscr{W}_2$ between 20CR and the CMIP5 ensemble, computed as described in Robin et al. (2017). The results are displayed in Figure 6. The two indicators are so similar (Pearson coefficient: $r_{pear} = 0.90$ and Spearman coefficient $r_{spear} = 0.85$ (Von Storch 1999)) that we will use the simpler R_{tot} when discussing our results. R_{tot} further indicates the direction of the changes (larger or smaller d, θ) while $\mathscr{W}_2$ only provides this information if the transport plan is computed (Villani 2008). the latter would be particularly complex to compute for the dataset analysed here.

By using the R_{tot} metrics, we find that all the models are within one standard deviation of the 20CR ensemble mean. At the same time, most models display median values in d and θ which are statistically different from those found in 20CR. The results of a Wilcoxon ranksum test (Von Storch 1999) are reported in Table S1. We remark that both Figure 7 and Table S1 indicate that models with a higher horizontal resolution have median values generally closer to those of the 20CR.

Different models have different spreads in d and θ, making it worth to investigate the extremes of these quantities and their relation with the weather regimes found in 20CR. Figures 8 and 9 display the composite SLP anomalies for the d and θ extremes - computed as in Figure 4 - for the three closest and the three farthest models (in terms of R_{tot}) relative to the 20CR. The models display similar composites independently their R_{tot} score. A quantitative analysis is reported in Table S2 using the Root Mean Square Error (RMSE) between 20CR and CMIP5 SLP composite anomalies. In general, we find the NAO- and NAO+ patterns to have higher RMSE whereas Atlantic Ridge and Blocking pattern are better represented.

The aggregate analysis shows that many CMIP5 models provide a dynamical picture coherent with that of the 20CR reanalysis. At the same time, most models reproduce statistically different medians of the two dynamical systems metrics relative to what found in the reanalysis.

4. Changes in the attractor properties

We next investigate whether the SLP's dynamical indicators have changed as an effect of past greenhouse forcing by separating the results into three periods: 1851-1900, 1901-1950 and 1951-2000. To provide a visually immediate depiction of the changes, we compute the joint histograms of (d, θ). The left-hand side and middle columns in Figure 10 display these for CCSM4 (the model with the highest horizontal resolution among those analysed) and 20CR. The median values are highlighted by magenta lines. The right-hand side panels show the differences between 20CR and CCSM4. Results for all the other models are shown in Figures S1-S26. In the reanalysis, both the median of the local dimension d and the median of θ increase from the first to the second and the third period. These increases, although small relative to the metrics' spread, are significant at the 95% level under a Wilcoxon ranksum test

(Table S1). For the models, the sign and significance of the changes depend on the period considered and are often not significant (Table S1). As a general feature, models tend to reproduce the dynamical properties better for the period 1951-2000 (see Figure 10 and Figures S1-S26).

We also perform a simple analysis of extremes in d and θ for the three different periods. We first compute the quantiles 0.02 and 0.98 for the d and θ distributions of the whole dataset. For the 20 CR these quantiles correspond to the black lines of Figure 4. Then we compute the number of days falling beyond the 0.02 and 0.98 quantiles in each of the three different periods. The results are shown in Figure 11. From top to bottom, we report the four extremes corresponding to Atlantic Ridge (maxima of θ), NAO- (minima of θ), Blocking (maxima of d) and NAO+ (minima of d). Again, the changes observed for the 20CR are generally larger than those observed for the models. None of the models displays changes coherent with those observed in the reanalysis. The results are stable with respect to reasonable changes in the quantiles.

5. Discussion and Conclusions

We have computed the instantaneous dynamical properties of the SLP fields for the 20CR and the CMIP5 historical runs, over the period 1851-2000. The goal of our analysis was to assess whether different models with different physics and resolutions quantitatively represent the *same* dynamical system and therefore possess attractors with similar characteristics. The metrics we used are the local dimension d and the inverse of the persistence time θ. As described in Faranda et al. (2017), these two quantities give a complete characterization of the attractor of the system.

When the whole analysis period is considered, we find that the models successfully capture some of the dynamical systems features identified in the reanalysis. For example, the range and

variability of the dynamical metrics are consistent across the datasets. At the same time, some models exagerate the effects of the seasonal cycle on the dynamical indicators, and the statistical agreement in the median values of the metrics is generally poor. Models with higher horizontal resolutions tend to perform better. The SLP fields corresponding to extremes in d and θ are mostly similar across the models and reanalysis. The main differences are found for the minima of θ and d, which in 20CR correspond to the NAO- and NAO+ patterns.

To detect the changes in the attractor propertieswith time, we have then divided the results into three periods: 1851-1900, 1901-1950 and 1951-2000. We have analysed the joint histograms of the d, θ variables and compared them to those obtained for the 20CR. The reanalysis shows significant increases in d and θ throughout the time period analysed. These changes also reflect in the number of days having extreme d and/or θ values. Days with Blocking and Atlantic Ridge patterns increase in frequency with time, while days with NAO patterns are decreasing. This is coherent with the results obtained by Alvarez-Castro et al. (2017). We note that the decrease in the frequency of NAO patterns does not imply a more negative NAO index. Indeed, the frequency of the NAO- pattern decreases more than that of the NAO+ pattern suggesting that – if anything – the positive NAO phase becomes more dominant towards the end of the analysis period, in agreement with SLP measurements (Hurrell et al. 2001). None of the models show comparable changes: changes in d are mostly not significant and the shifts in θ are significantly smaller than in the reanalysis. This is reflected in the fact that the median values of the two variables in the models do not show a clear upward trend trhoughout the three sub-periods considered. As a caveat we note that our analysis does not attempt to separate the forced variability from natural low-frequency oscillations and that, especially during the first two periods analysed, it is unclear whether the greenhouse forcing can be clearly discerned above the background

278 ”climate noise” (Paeth et al. 1999; Lyu et al. 2015). We must therefore take into account the possi-
279 bility that the model’s internal variability dominates over the trends for the time period considered.
280
281 As a further caveat we note that several studies (Krueger et al. 2013; Ferguson and Villarini
282 2012, 2014) have questioned the consistency of the 20CR dataset. So, how much can we trust
283 the results obtained when investigating the three separate periods? Most of the observations used
284 to constrain 20CR in the first part of the dataset are located in Europe or eastern North America
285 (Cram et al. 2015); the North Atlantic sector can therefore be expected to perform better than
286 elsewhere. The dataset has a sufficiently high horizontal resolution to obtain a good estimate of
287 the local dimension distribution (Faranda et al. 2017). However, the fact that the 20CR data is
288 increasingly constrained to follow the SLP observations as the time approaches the present day
289 causes a decrease of the ensemble spread with time, since the system is more closely pinned to
290 a specific manifold (the observations), without the possibility of exploring the full phase space.
291 This may explain the changes in the 20CR local dimension with time.
292
293 If the results obtained for 20CR do not depend on the quality of the dataset but have a real phyis-
294 cal meaning, the increase in dimension with time could be explained by using the results obtained
295 in Faranda et al. (2013) for simple dynamical systems. In these conceptual models, the dimension
296 increases with the *temperature* of the system defined – following the Einstein model for Brownian
297 motion – as the variance of a stochastic noise term added to the deterministic dynamics. Schubert
298 and Lucarini (2015) have demonstrated that this modeling approach is relevant for climate and, in
299 general, for any multiscale system. Faranda et al. (2013) have further shown that $\theta \to 1$ when the
300 temperature is increased, coherently with what observed in the 20CR.

Acknowledgments. P.Yiou and D. Faranda were supported by ERC grant No. 338965, M.C. Alvarez-Castro was supported by Swedish Research Council grant No. C0629701 and G. Messori was supported by a grant from the Department of Meteorology of Stockholm University.

References

Alvarez-Castro, M., D. Faranda, and P. Yiou, 2017: Changes in the atmospheric dynamics leading to west-european heatwaves since 1900. *Environmental Reserch Letters*, **in review**.

Cassou, C., and J. Cattiaux, 2016: Disruption of the european climate seasonal clock in a warming world. *Nature Climate Change*.

Collet, P., and J.-P. Eckmann, 2009: *Iterated maps on the interval as dynamical systems*. Springer Science & Business Media.

Comas-Bru, L., and F. McDermott, 2014: Impacts of the ea and sca patterns on the european twentieth century nao–winter climate relationship. *Quarterly Journal of the Royal Meteorological Society*, **140 (679)**, 354–363.

Compo, G. P., and Coauthors, 2011: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137 (654)**, 1–28.

Cram, T. A., and Coauthors, 2015: The international surface pressure databank version 2. *Geoscience Data Journal*, **2 (1)**, 31–46, doi:10.1002/gdj3.25, URL http://dx.doi.org/10.1002/gdj3.25.

Faranda, D., M. C. Alvarez-Castro, and P. Yiou, 2016a: Return times of hot and cold days via recurrences and extreme value theory. *Climate Dynamics*, 1–13, doi:10.1007/s00382-016-3042-6.

Faranda, D., J. M. Freitas, V. Lucarini, G. Turchetti, and S. Vaienti, 2013: Extreme value statistics for dynamical systems with noise. *Nonlinearity*, **26 (9)**, 2597.

Faranda, D., G. Masato, N. Moloney, Y. Sato, F. Daviaud, B. Dubrulle, and P. Yiou, 2016b: The switching between zonal and blocked mid-latitude atmospheric circulation: a dynamical system perspective. *Climate Dynamics*, **47 (5-6)**, 1587–1599.

Faranda, D., G. Messori, and P. Yiou, 2017: Dynamical proxies of north atlantic predictability and extremes. *Scientific Reports*, **7**, 41 278.

Ferguson, C. R., and G. Villarini, 2012: Detecting inhomogeneities in the twentieth century reanalysis over the central united states. *Journal of Geophysical Research: Atmospheres*, **117 (D5)**, n/a–n/a, doi:10.1029/2011JD016988, URL http://dx.doi.org/10.1029/2011JD016988, d05123.

Ferguson, C. R., and G. Villarini, 2014: An evaluation of the statistical homogeneity of the twentieth century reanalysis. *Climate Dynamics*, **42 (11)**, 2841–2866, doi:10.1007/s00382-013-1996-1, URL http://dx.doi.org/10.1007/s00382-013-1996-1.

Field, C. B., 2012: *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.

Freitas, A. C. M., J. M. Freitas, and M. Todd, 2010: Hitting time statistics and extreme value theory. *Probability Theory and Related Fields*, **147 (3-4)**, 675–710.

Freitas, A. C. M., J. M. Freitas, and M. Todd, 2012: The extremal index, hitting time statistics and periodicity. *Advances in Mathematics*, **231 (5)**, 2626–2665.

Hurrell, J. W., 1995: Decadal trends in the north atlantic oscillation: Regional temperatures and precipitation. *Science*, **269 (5224)**, 676–679, doi:10.1126/science.269.5224.

676, URL http://science.sciencemag.org/content/269/5224/676, http://science.sciencemag.org/content/269/5224/676.full.pdf.

Hurrell, J. W., Y. Kushnir, and M. Visbeck, 2001: The north atlantic oscillation. *Science*, **291 (5504)**, 603–605.

Kay, J., and Coauthors, 2015: The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, **96 (8)**, 1333–1349.

Krueger, O., F. Schenk, F. Feser, and R. Weisse, 2013: Inconsistencies between long-term trends in storminess derived from the 20cr reanalysis and observations. *Journal of Climate*, **26 (3)**, 868–874, doi:10.1175/JCLI-D-12-00309.1.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20 (2)**, 130–141.

Lucarini, V., D. Faranda, G. Turchetti, and S. Vaienti, 2012: Extreme value theory for singular measures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **22 (2)**, 023 135.

Lucarini, V., and Coauthors, 2016: *Extremes and recurrence in dynamical systems*. John Wiley & Sons.

Lyu, K., X. Zhang, J. A. Church, and J. Hu, 2015: Quantifying internally generated and externally forced climate signals at regional scales in cmip5 models. *Geophysical Research Letters*, **42 (21)**, 9394–9403, doi:10.1002/2015GL065508, URL http://dx.doi.org/10.1002/2015GL065508, 2015GL065508.

Messori, G., R. Caballero, and D. Faranda, 2017: A dynamical systems approach to studying mid-latitude weather extremes. *Submitted: Geophysical Research Letters*.

Moore, G., I. A. Renfrew, and R. S. Pickart, 2013: Multidecadal mobility of the north atlantic oscillation. *Journal of Climate*, **26 (8)**, 2453–2466.

Paeth, H., A. Hense, R. Glowienka-Hense, S. Voss, and U. Cubasch, 1999: The north atlantic oscillation as an indicator for greenhouse-gas induced regional climate change. *Climate Dynamics*, **15 (12)**, 953–960, doi:10.1007/s003820050324, URL http://dx.doi.org/10.1007/s003820050324.

Pickands III, J., 1975: Statistical inference using extreme order statistics. *the Annals of Statistics*, 119–131.

Robin, Y., P. Yiou, and P. Naveau, 2017: Detecting changes in forced climate attractors with wasserstein distance. *Nonlinear Processes in Geophysics Discussions*, **2017**, 1–19, doi:10.5194/npg-2017-5, URL http://www.nonlin-processes-geophys-discuss.net/npg-2017-5/.

Rogers, J. C., 1997: North atlantic storm track variability and its association to the north atlantic oscillation and climate variability of northern europe. *Journal of Climate*, **10 (7)**, 1635–1647, doi:10.1175/1520-0442(1997)010⟨1635:NASTVA⟩2.0.CO;2.

Schubert, S., and V. Lucarini, 2015: Covariant lyapunov vectors of a quasi-geostrophic baroclinic model: analysis of instabilities and feedbacks. *Quarterly Journal of the Royal Meteorological Society*, **141 (693)**, 3040–3055.

Shepherd, T. G., 2014: Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, **7 (10)**, 703–708.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93 (4)**, 485–498.

Villani, C., 2008: *Optimal transport: old and new*, Vol. 338. Springer Science & Business Media.

Von Storch, H., 1999: Misuses of statistical analysis in climate research. *Analysis of Climate Variability*, Springer, 11–26.

LIST OF TABLES

402 TABLE 1. List of CMIP5 Models Analysed and 20CR reanalysis from 1851 to 2001. Models are ordered by
403 increasing in resolution.

No.[1]	Model	Institution/ID	Country	Resolution[2]
1	20CRv2c	NOAA-CIRES	USA	2 x 2
2	CMCC-CESM	Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	3,75x3,75
3	CanESM2	Canadian Centre for Climate Modelling and Analysis, CCCMa	Canada	2.81x 2.79
4	MIROC-ESM-CHEM	MIROC[4]	Japan	2,81x 2,79
5	MIROC-ESM	MIROC[4]	Japan	2,81x 2,79
6	BCC-CSM1-1	Beijing Climate Center	China	2,81x 2,79
7	IPSL-CM5B-LR	Institute Pierre Simon Laplace, IPSL	France	3,75x1,89
8	NorESM1-M	Norwegian Climate Center	Norway	2,5x1,89
9	FGOALS-2	Institute of Atmospheric Physics, Chinese Academy of Sciences	China	2,81x2,81
10	MPI-ESM-P	Max Planck Institute for Meteorology, MPI	Germany	1,87x1,87
11	MPI-ESM-LR	Max Planck Institute for Meteorology, MPI	Germany	1,87x1,87
12	CSIRO-MK3-6-0	CSIRO-BOM[5]	Australia	1,87x1,87
13	CMCC-CMS	Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	1,87x1,87
14	MPI-ESM-MR	Max Planck Institute for Meteorology, MPI	Germany	1,87x1,87
15	IPSL-CM5A-MR	Institute Pierre Simon Laplace, IPSL	France	2,5x1,26
16	INM-CM4	Institute for Numerical Mathematics, INM	Russia	2x1,5
17	ACCESS 1-0	CSIRO-BOM[5]	Australia	1,87x1,25
18	MIROC5	MIROC[4]	Japan	1,40x1,40
19	CNRM-CM5	CNRM-CERFACS[3]	France	1,40x1,40
20	MRI-ESM1	Meteorological Research Institute, MRI	Japan	1,125x1,125
21	BCC-CSM1-M	Beijing Climate Center	China	1,125x1,125
22	MRI-CGCM3	Meteorological Research Institute, MRI	Japan	1,125x1,125
23	EC-EARTH	Danish Meteorological Institute, DMI	Denmark	1,125x1,125
24	CESM1-FASTCHEM	Community Earth System Model Contributors, NCAR	USA	1,25x0,94
25	CESM1-CAM5	Community Earth System Model Contributors, NCAR	USA	1,25x0,94
26	CESM1-BGC	Community Earth System Model Contributors, NCAR	USA	1,25x0,94
27	CCSM4	National Center for Atmospheric Research, NCAR	USA	1,25x0,94

[1] Order by horizontal resolution (Decreasing)
[2] Longitude x Latitude (°)
[3] Centre National de Recherches Meteorologiques - Centre Europeen de Recherche et de Formation Avance en Calcul Scientifique
[4] Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
[5] Commonwealth Scientific and Industrial Research Organisation(CSIRO), Bureau of Meteorology(BOM)

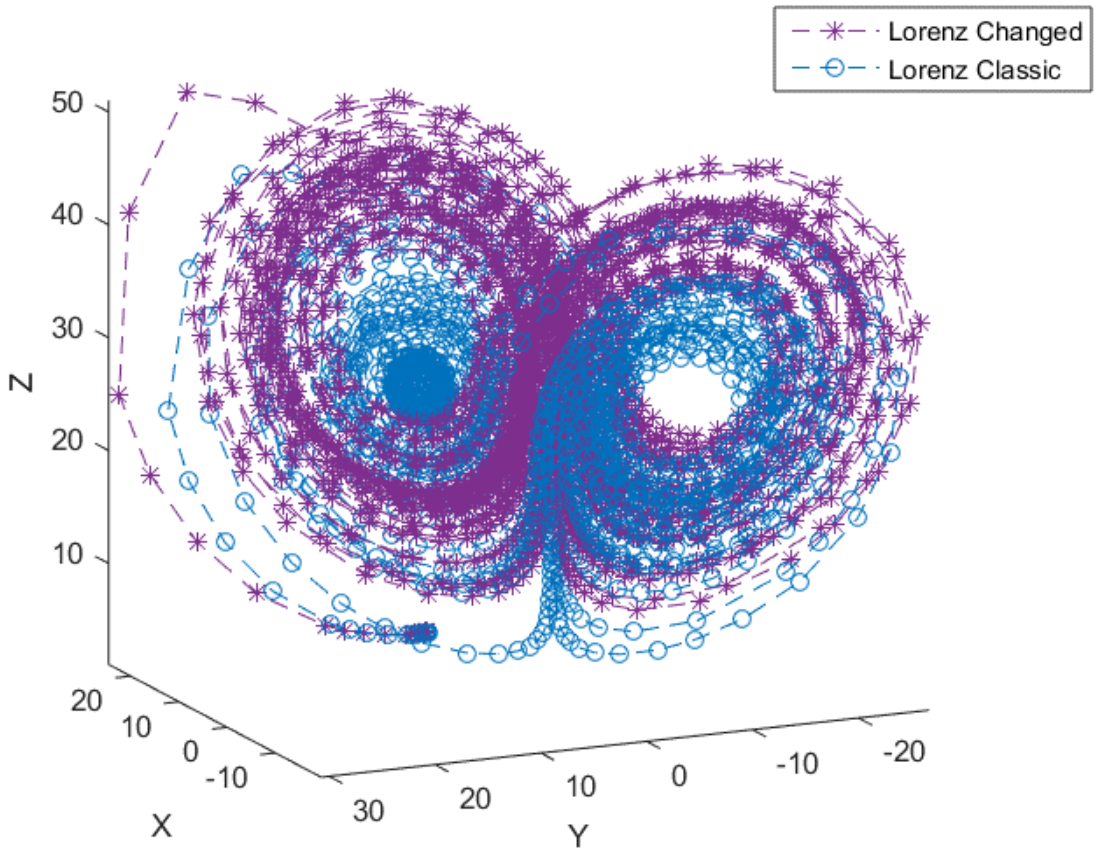
TABLE 2. Two sample Kolmogorov Smirnov (KS) tests results and p-values for the distribution of distances between realizations of the Lorenz system for two different metrics: (W_2) Wasserstein distance between full d, θ distributions, (m) distances between the medians of the d, θ distributions. The null hypothesis is that data in $\Lambda(1)$ and $\Lambda(2)$ are from the same continuous distribution, with a significance level of 95%. The null hypothesis is always rejected.

	$\Lambda(\sigma_{28})/\Lambda(\sigma_{28/28.5})$	$\Lambda(\sigma_{28.5})/\Lambda(\sigma_{28/28.5})$
KS(W_2)	0.16	0.22
p-value(W_2)	$5 \cdot 10^{-10}$	$2 \cdot 10^{-20}$
KS(m)	0.29	0.4
p-value(m)	$3 \cdot 10^{-33}$	$1 \cdot 10^{-63}$

TABLE 3. List of median values for dimension d and inverse of persistence time θ, distances $\delta(d)$ and $\delta(\theta)$ from the median of the 20CR , relative distances $R(d) = \delta(d)/\max\delta(d)$ and $R(\theta) = \delta(\theta)/\max\delta(\theta)$ with respect to the farthest model and global score $R_{tot} = (R(d) + R(\theta))/2$.

N.	Model	median(d)	$\delta(d)$	$R(d)$	median(θ)	$\delta(\theta)$	$R(\theta)$	R_{tot}
1	20CRv2c	11,56	–	–	0.5	–	–	–
2	CMCC-CESM	12,22	0,67	0,54	0,51	0,01	0,13	0,33
3	CanESM2	11,99	0,43	0,35	0,51	0	00.2	0,19
4	MIROC-ESM-CHEM	12,54	0,98	0,8	0,47	0,04	0,75	0,77
5	MIROC-ESM	12,48	0,92	0,75	0,47	0,04	0,76	0,76
6	BCC-CSM1	12,12	0,57	0,46	0,51	0,01	0,12	0,29
7	IPSL-CM5B	12,73	1,17	0,95	0,46	0,05	0,93	0,94
8	NorESM1-M	12,12	0,56	0,46	0,48	0,02	0,44	0,45
9	FGOALS-S2	11,63	0,07	0,06	0,45	0,05	1,00	0,53
10	MPI-ESM-P	12,17	0,61	0,5	0,51	0	0,06	0,28
11	MPI-ESM-LR	12,13	0,58	0,47	0,51	0,01	0,14	0,3
12	CSIRO-MK3-6-0	12,66	1,11	0,9	0,5	0	0,02	0,46
13	CMCC-CMS	11,95	0,39	0,32	0,52	0,01	0,22	0,27
14	MPI-ESM-MR	12,09	0,53	0,43	0,51	0	0,09	0,26
15	IPSL-CM5A	11,86	0,31	0,25	0,48	0,03	0,51	0,38
16	INM-CM4	12,79	1,23	1	0,47	0,04	0,70	0,85
17	ACCESS-1-0	11,74	0,19	0,15	0,49	0,02	0,31	0,23
18	MIROC5	12,58	1,02	0,83	0,49	0,02	0,33	0,58
19	CNRM-CM5	12,36	0,8	0,65	0,47	0,03	0,59	0,62
20	MRI-ESM1	11,72	0,16	0,13	0,51	0,01	0,12	0,13
21	BCC-CSM1-M	11,45	0,11	0,09	0,55	0,05	0,91	0,5
22	MRI-CGCM3	11,74	0,18	0,15	0,51	0	0,07	0,11
23	EC-EARTH	11,87	0,32	0,26	0,5	0	0,01	0,13
24	CESM1-FASTCHEM	11,56	0	0	0,51	0,01	0,18	0,09
25	CESM1-CAM5	11,88	0,32	0,26	0,51	0,01	0,13	0,2
26	CESM1-BGC	11,53	0,02	0,02	0,51	0,01	0,11	0,06
27	CCSM4	11,57	0,02	0,01	0,51	0,01	0,12	0,07

LIST OF FIGURES



457 FIG. 1. Two realizations of the Lorenz attractor. Blue: classic attractor $\Delta t \simeq 0.035$, $\sigma = 28$, $r = 10$, $b = 8/3$;
458 Violet: $\sigma = 28.5$

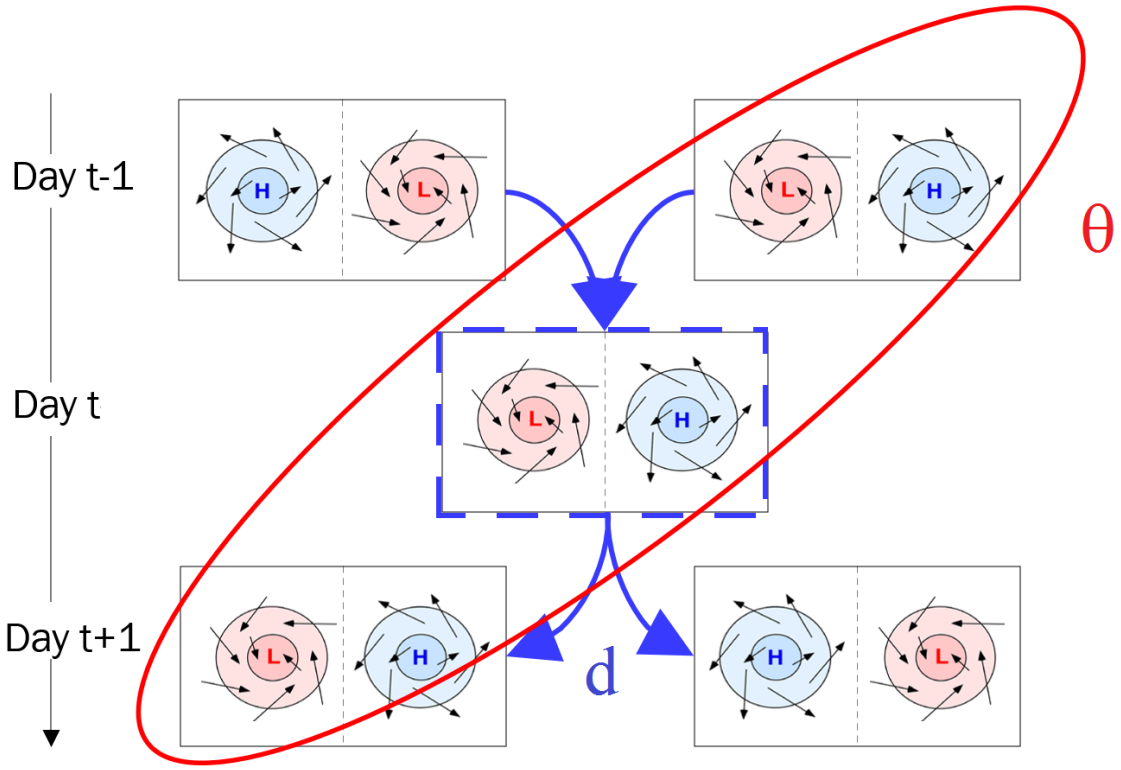


FIG. 2. Schematic representation of the dynamical indicators: the local dimension d is proportional to the number of possible configurations originating that of the day analyzed and resulting from it. θ, is the inverse of the persistence time of a certain configuration.

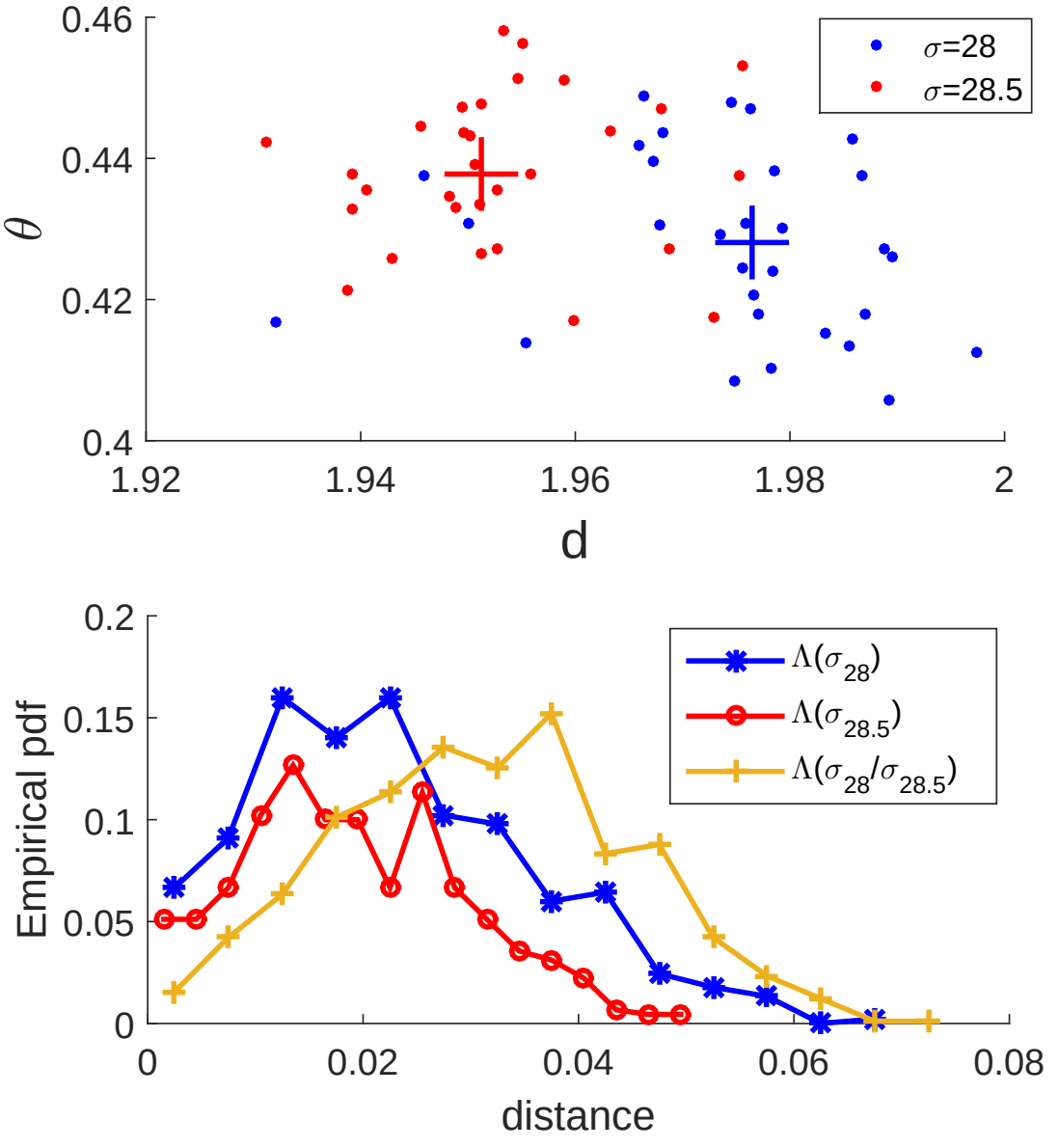


FIG. 3. Top: Medians of d and θ for 30 realizations of the Lorenz attractor with $\sigma = 28$ (blue) and $\sigma = 28.5$ (red); the crosses display the median of the ensemble of realizations. Bottom: Empirical probability density functions (pdf) of the pairwise distances among d and θ medians in $\sigma = 28$ realizations ($\Lambda(\sigma_{28})$), $\sigma = 28.5$ realizations ($\Lambda(\sigma_{28})$) and mixed realizations ($\Lambda(\sigma_{28}/\sigma_{28.5})$).

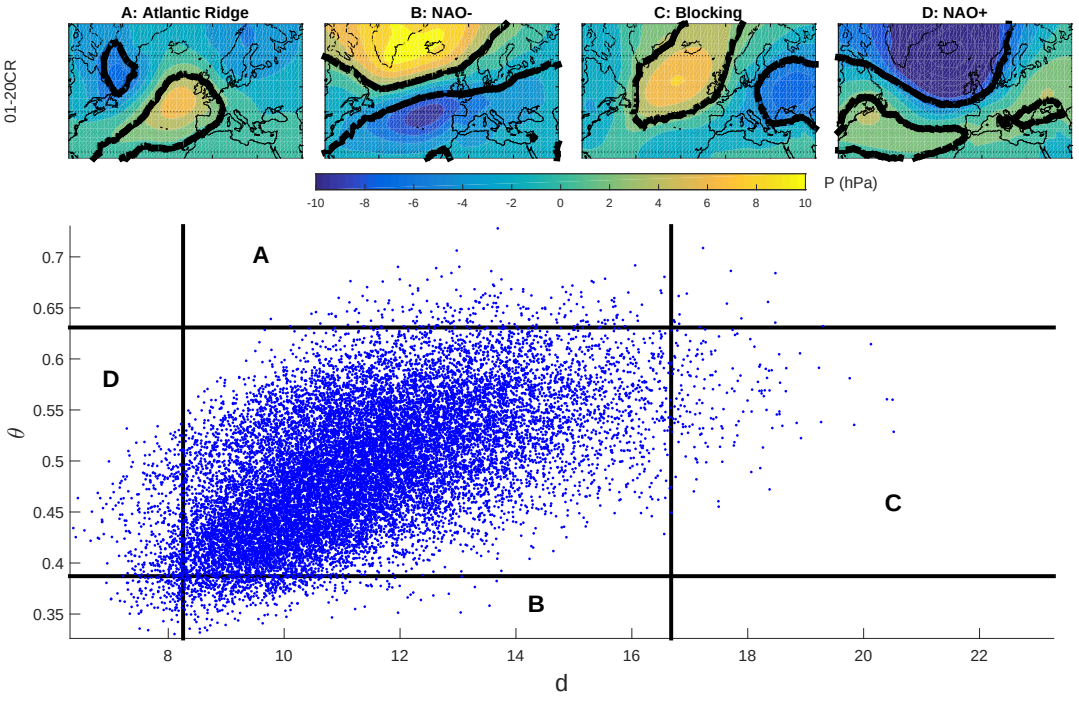


FIG. 4. The scatter plot displays the daily values of the instantaneous dimension d and the persistence θ of the field. The straight black lines mark the 0.02 and 0.98 quantiles of d and θ. The composite anomalies in SLP for the four regions delimited by the black lines are plotted as side panels and can be associated with known weather regimes: Atlantic Ridge (maxima of θ), NAO- (minima of θ), Blocking (maxima of d), NAO+ (minima of d). The black lines indicates regions where at least the 2/3 of the composite members display the same sign.

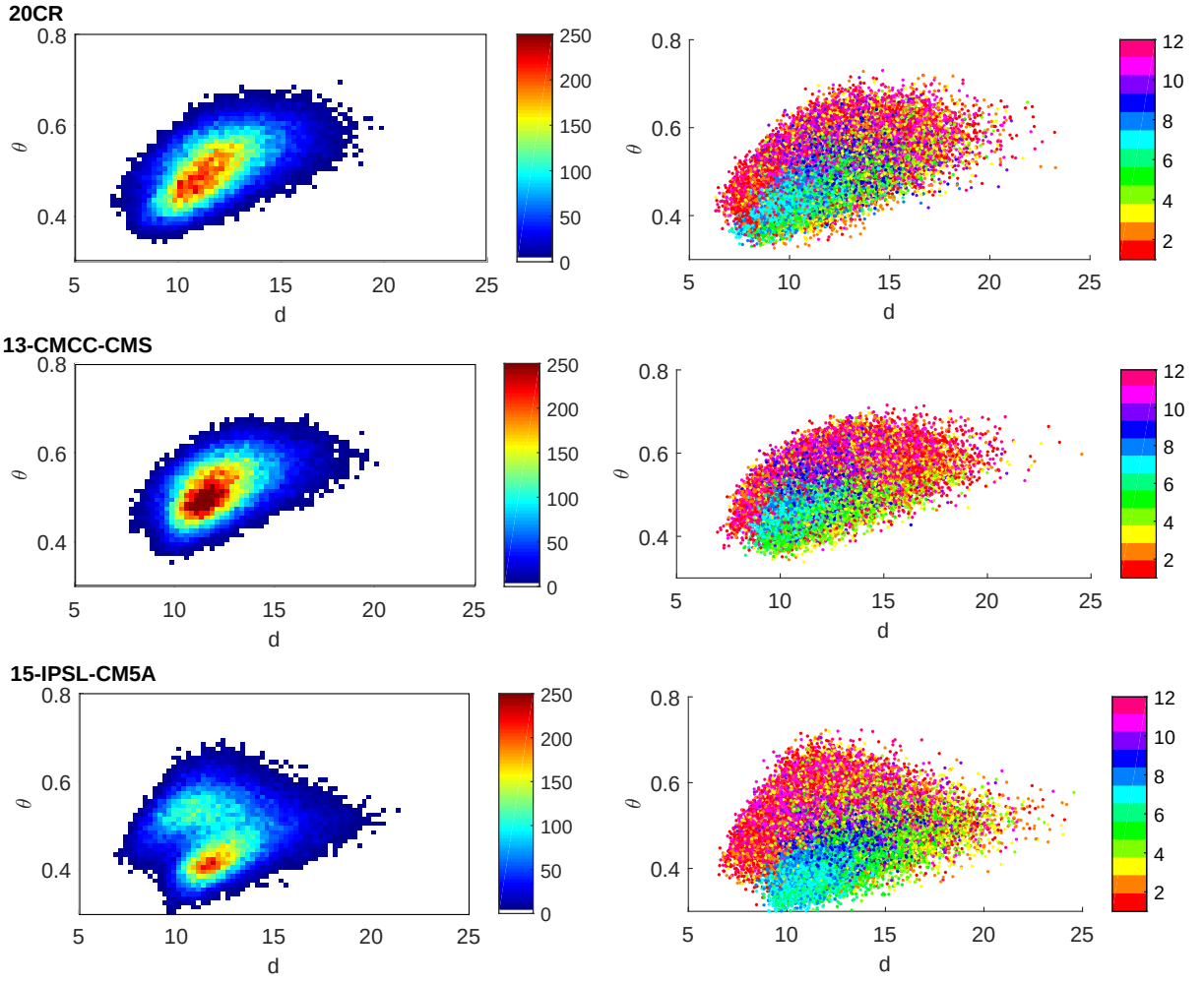
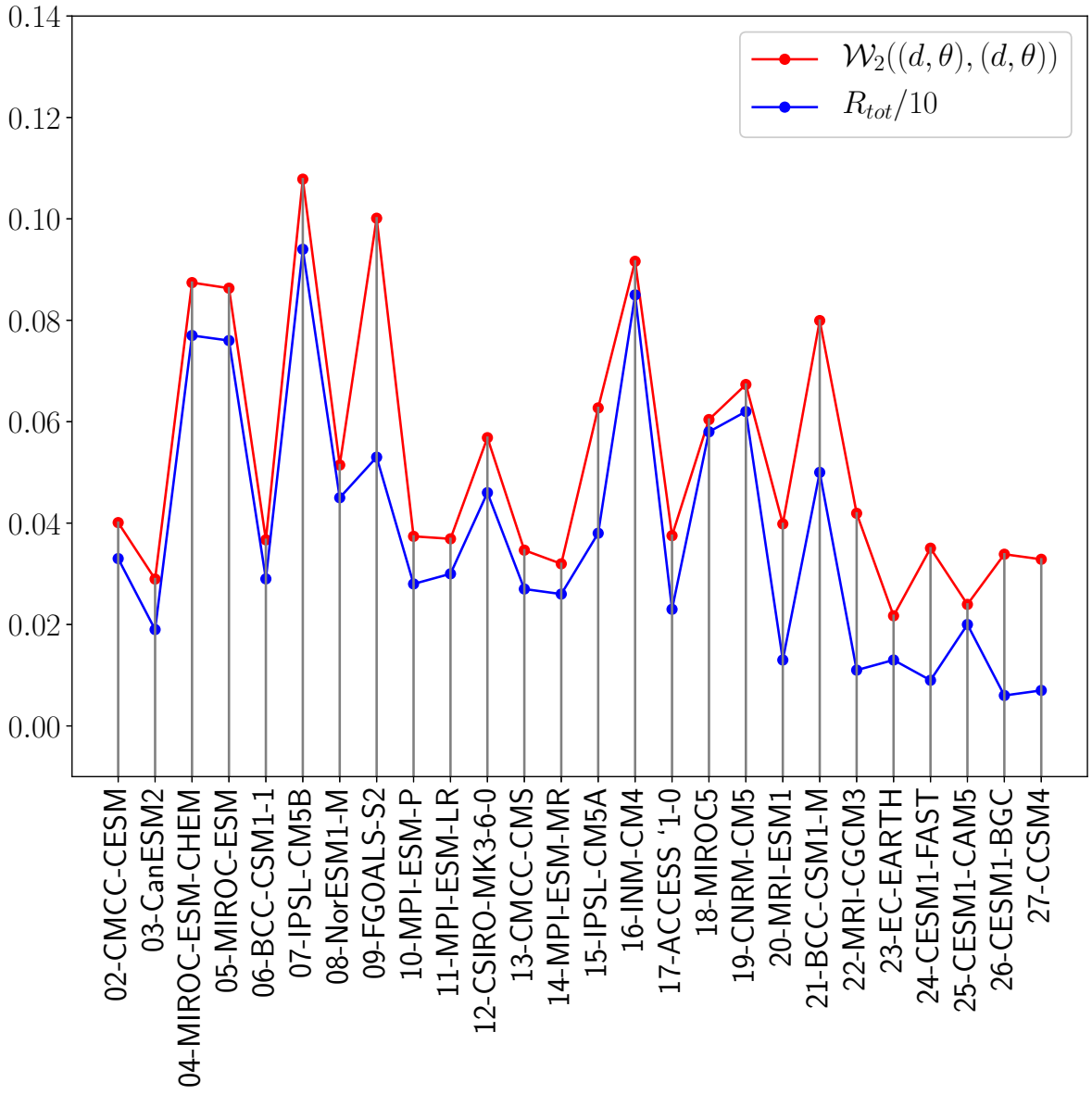


FIG. 5. (d, θ) bivariate histograms (left) and scatter-plots (right) for the 20 CR (top), the CMCC-CMS (center) and the IPSL-CM5A models(bottom). The color-scales on the left indicate the frequency of observations in number of days. The colorscales on the right indicate the month of the observation and show the dependence of the (d, θ) diagrams on the seasonal cycle.



475 FIG. 6. Comparison between R_{tot} values (blue) and Wasserstein distances $\mathscr{W}_2$ (red) computed with the proce-
476 dure described in Robin et al. (2017) between the (d,θ) 20CR and CMIP5 distributions.

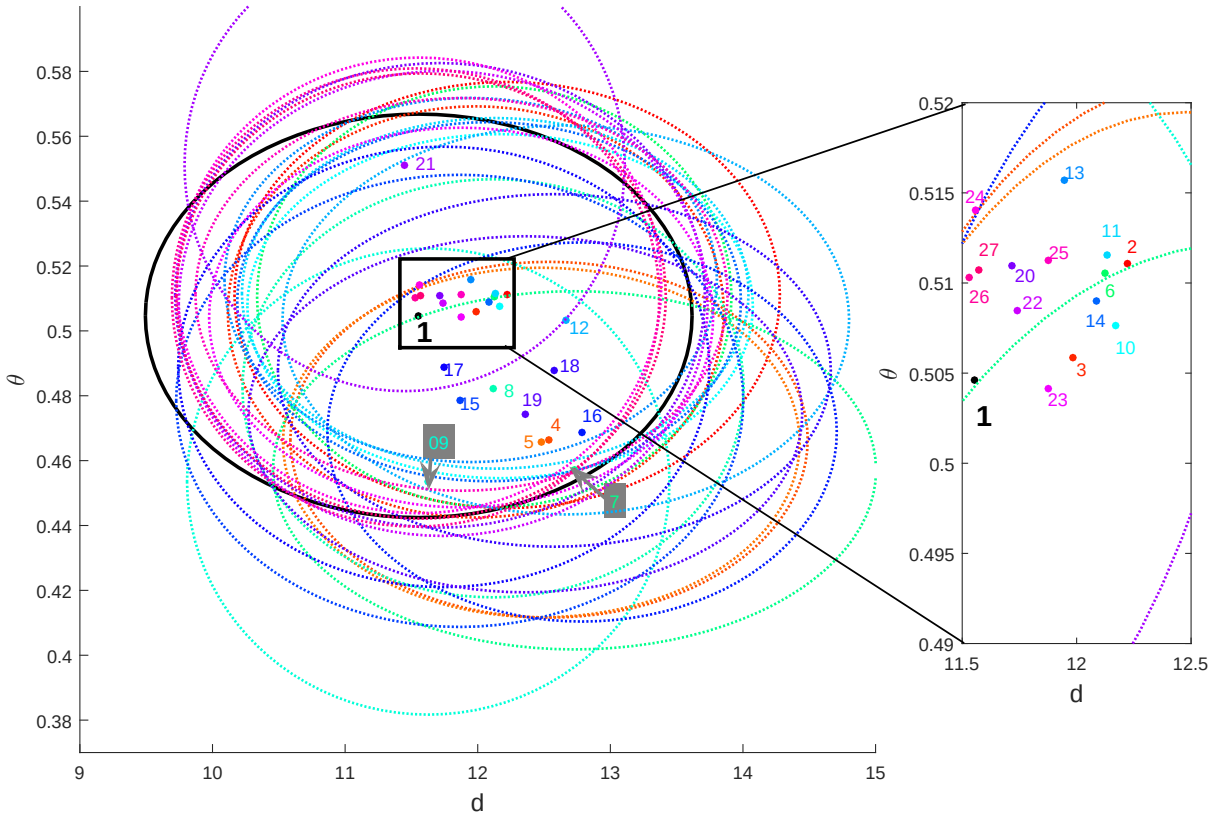


FIG. 7. Comparison between 20 CR median values of (d, θ) (denoted by 1) and all the other CMIP5 models (progressive numbers 2-27, see table 1 for the details). The semiaxes of each ellipse represent one standard deviation of d and θ. Models 07 and 09 are highlighted by arrows and a grey background to increase their visibility.

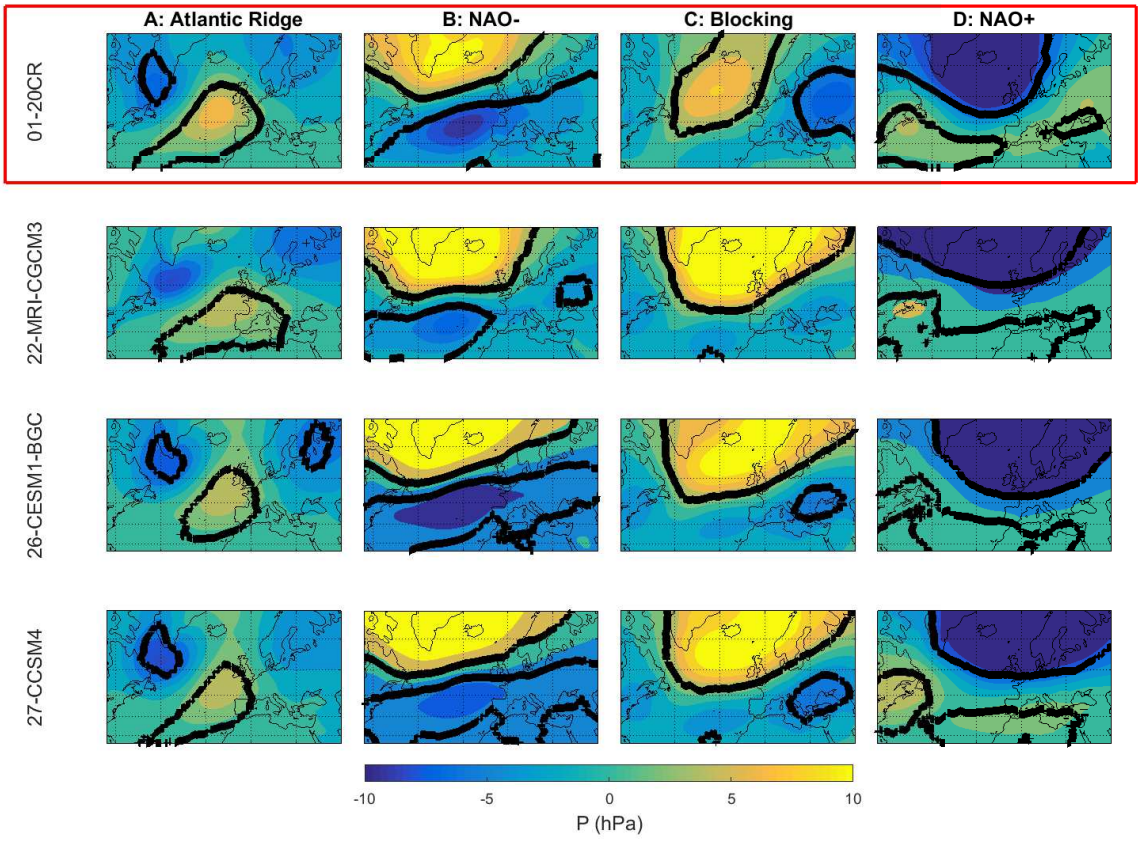
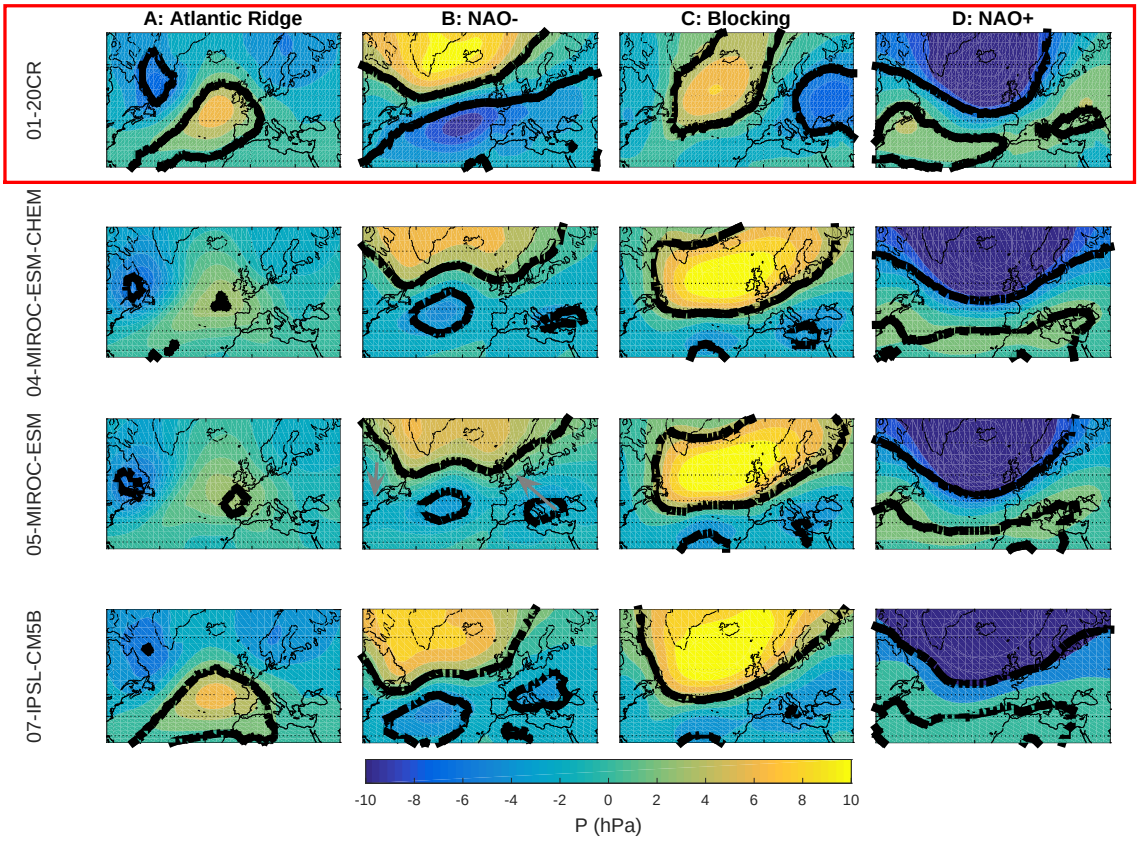


FIG. 8. Comparison between 20CR and the best 3 models (according to the metric R_{tot}, a measure of distance from the median values of d, θ) of the composite anomalies in SLP for the four regions of the (d, θ) diagram, highlighted in Fig. 4. The black lines indicate regions where at least the 2/3 of the composite members display the same sign.



485 FIG. 9. Comparison between 20CR and the worst 3 models (according to the metric R_{tot}, a measure of distance
486 from the median values of d, θ) of the composite anomalies in SLP for the four regions of the (d, θ) diagram,
487 highlighted in Fig. 4. The black lines indicate regions where at least the 2/3 of the composite members display
488 the same sign.

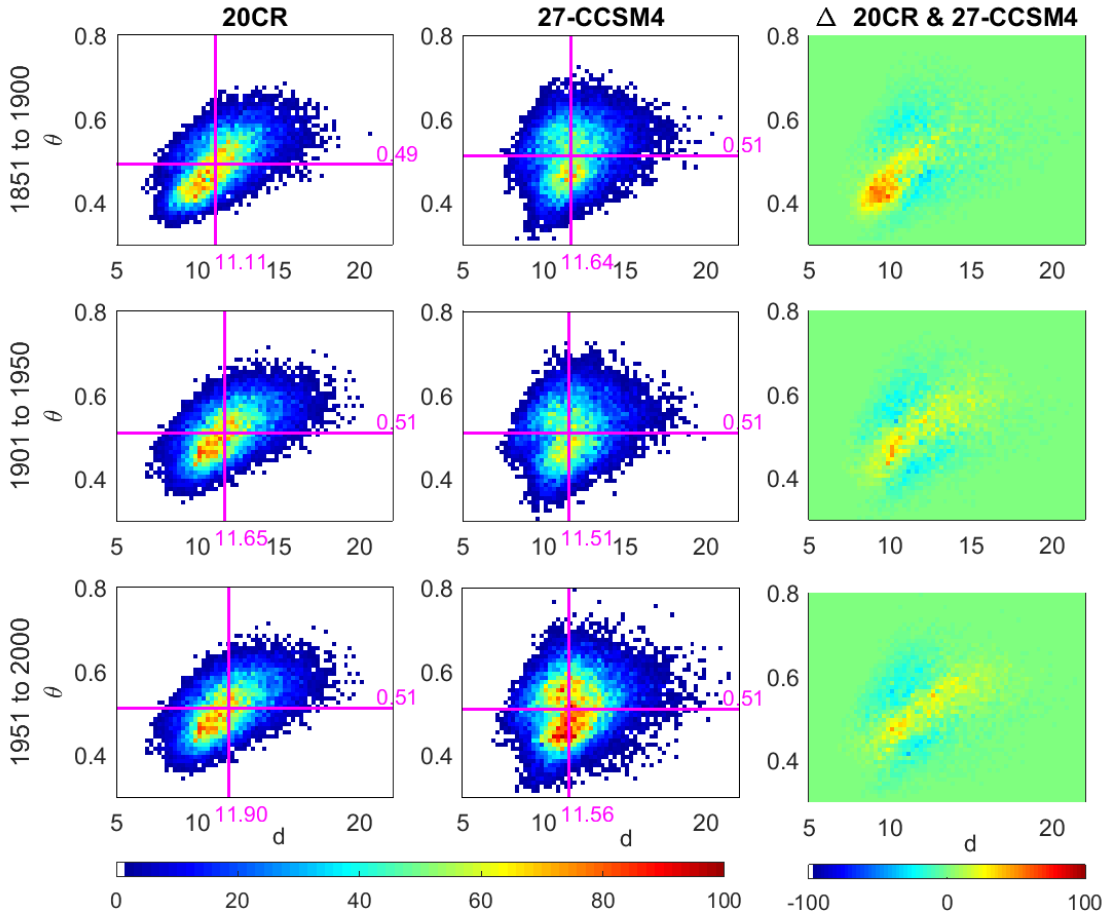


FIG. 10. Bivariate histograms of (d, θ) for 20CR (left), the model CCSM4 (centre) and their difference Δ (right) for three periods: 1851-1900 (top), 1901-1950 (centre), 1951-2000 (bottom). The violet lines indicate the median values. The colorbars represent the frequency of joint d, θ observations in number of days.

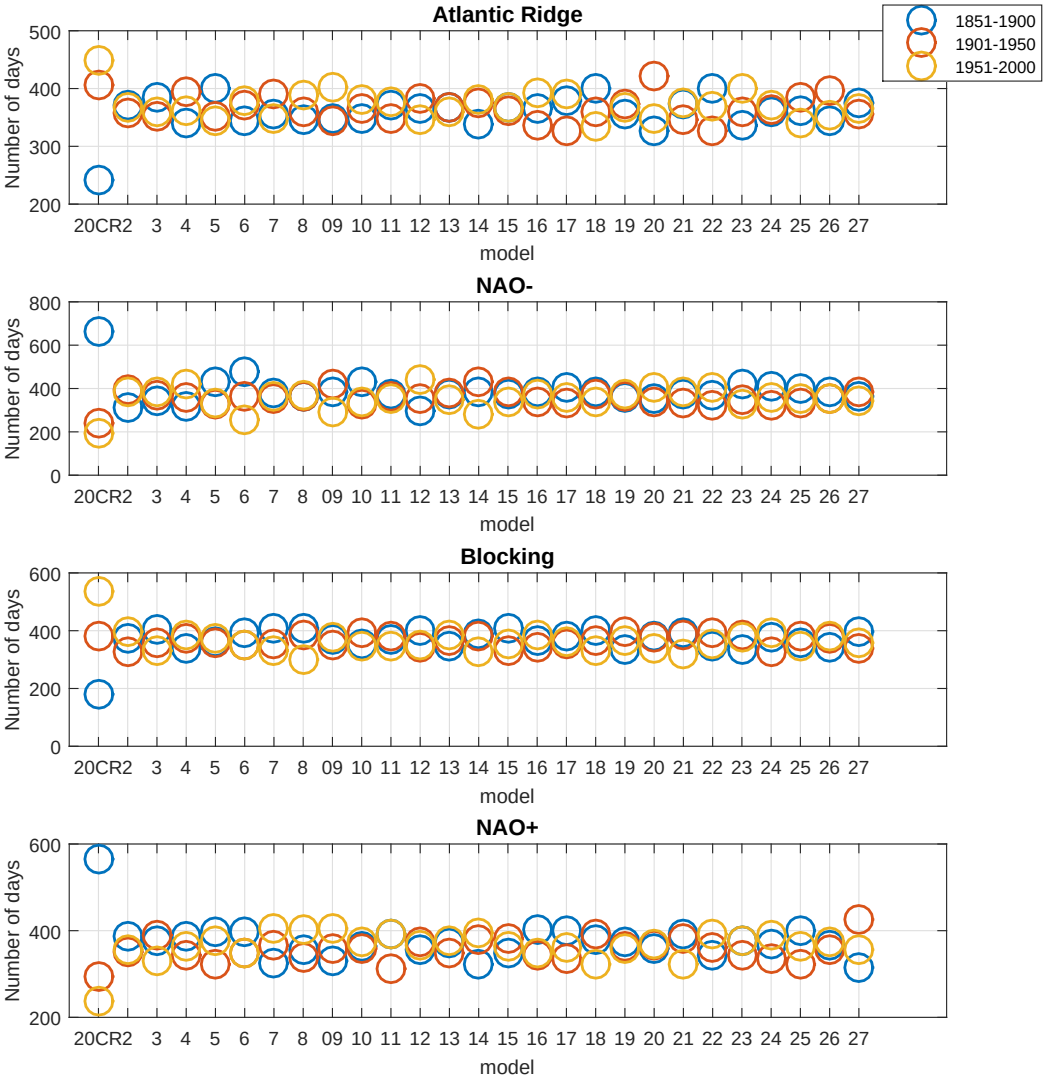


FIG. 11. Change in the number of extreme days, computed using the 0.98 and 0.02 quantiles of the full time series for each model, for three different sub-periods. From top to bottom different regimes: Atlantic Ridge (maxima of θ), NAO- (minima of θ), Blocking (maxima of d), NAO+ (minima of d)