



**HAL**  
open science

# Dynamical proxies of North Atlantic predictability and extremes

Davide Faranda, Gabriele Messori, Pascal Yiou

► **To cite this version:**

Davide Faranda, Gabriele Messori, Pascal Yiou. Dynamical proxies of North Atlantic predictability and extremes. *Scientific Reports*, 2017, 10.1038/srep41278 . hal-01340301

**HAL Id: hal-01340301**

**<https://hal.science/hal-01340301>**

Submitted on 30 Jun 2016

**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.

# Dynamical proxies of North Atlantic predictability and extremes

Davide Faranda<sup>1\*</sup>, Gabriele Messori<sup>2</sup> & Pascal Yiou<sup>1</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>2</sup>Met Office Hadley Centre, Exeter, UK

\* Correspondence to: Davide Faranda, email: [davide.faranda@lsce.ipsl.fr](mailto:davide.faranda@lsce.ipsl.fr)

## 1 **Main**

2 **Atmospheric flows are characterized by chaotic dynamics and recurring large-scale pat-**  
3 **terns . These two characteristics point to the existence of an atmospheric attractor defined**  
4 **by Lorenz as: “the collection of all states that the system can assume or approach again**  
5 **and again, as opposed to those that it will ultimately avoid”. The average dimension  $D$**   
6 **of the attractor corresponds to the number of degrees of freedom sufficient to describe**  
7 **the atmospheric circulation. However, obtaining reliable estimates of  $D$  has proved chal-**  
8 **lenging . Moreover,  $D$  does not provide information on transient atmospheric motions,**  
9 **which lead to weather extremes . Using recent developments in dynamical systems theory**  
10 **, we show that such motions can be classified through instantaneous rather than aver-**  
11 **age properties of the attractor. The instantaneous properties are uniquely determined by**  
12 **instantaneous dimension and stability. Their extreme values correspond to specific atmo-**  
13 **spheric patterns, and match extreme weather occurrences. We further show the existence**

14 **of a significant correlation between the time series of instantaneous stability and dimen-**  
15 **sion and the mean spread of sea-level pressure fields in an operational ensemble weather**  
16 **forecast at steps of over two weeks. We believe this method provides an efficient and prac-**  
17 **tical way of evaluating and informing operational weather forecasts.**

18

19 Dynamical systems analyses led to the crucial notion that atmospheric motions are chaotic and  
20 settle on an attractor [1]. Estimates of the average dimensions  $D$  of atmospheric attractors  
21 were produced from times series of various meteorological variables [2] because this quan-  
22 tity roughly indicates the numbers of degrees of freedom sufficient to describe the flow in its  
23 average, time-stationary configuration. However, many weather phenomena of great societal  
24 and economic relevance such as extratropical storms, heatwaves and cold-spells are linked to  
25 transient metastable states of the atmosphere, whose dynamical properties depend on the instan-  
26 taneous rather than average properties of the attractor [3]. Such local properties are uniquely  
27 determined by two quantities: the local dimension and stability of the state being considered  
28 [4].

29 The concept of instantaneous dimension is intuitive: for a state  $\zeta$  of the attractor (an atmo-  
30 spheric configuration), the instantaneous dimension  $d(\zeta)$  measures the density of neighbouring  
31 points (similar configurations). This implies that  $d$  can be related to both the entropy and the  
32 predictability of nearby trajectories [5]. The stability of the state  $\zeta$  is measured by the stickiness  
33  $\theta(\zeta)$  defined as the inverse of the average persistence time of trajectories around  $\zeta$ . If  $\zeta$  is a fixed  
34 point of the dynamics,  $\theta(\zeta) = 0$ . For a point that leaves the neighbourhood of  $\zeta$  immediately,  
35  $\theta = 1$ . In general, the stickier the point  $\zeta$ , the longer the previous and subsequent states of the  
36 system will resemble  $\zeta$ .

37 These instantaneous properties have not been previously computed for atmospheric flows for  
38 two main reasons: i) until recently the length of high-frequency, geographically extensive me-

39 teorological time series has been insufficient to allow reliable estimates ii) the methodologies  
40 used to compute the average dimension of the attractor were not suited to the purpose. In this  
41 paper we compute the distribution in phase space of the instantaneous dimensions  $0 < d < \infty$   
42 and the stickiness  $0 < \theta < 1$  for daily sea-level pressure (SLP) fields in the North Atlantic,  
43 by applying a novel methodology based on the universal behavior of the Poincaré recurrences  
44 in chaotic systems [4]. This methodology is general, and thus applicable to a wide range of  
45 dynamical systems beyond climate science. There are two key advances relative to previous  
46 attempts: first, our methodology removes major uncertainties associated with past estimates  
47 [6, 2] (see Methods). Second, it yields the full probability distribution of the instantaneous  
48 dimension of the attractor. The validity of our approach has been successfully tested on the  
49 idealized Lorenz system [7] and on a number of synthetic fields (Extended Data Figs A.1 and  
50 A.2).

51 We chose SLP as a representative field for the large-scale atmospheric circulation over the  
52 North Atlantic and Europe (see Methods). To verify the robustness of our results we analyses  
53 two distinct daily SLP timeseries: the NCEP/NCAR [8] and ERA-Interim data [9]. The average  
54 dimensions  $D$  obtained by averaging  $d$  over all  $\zeta$ , including tests performed on coarse-grained  
55 NCEP/NCAR data, are shown in Fig. 1-a. If we interpret the resolution, which is of order  
56  $10^4$  grid-points for ERA-Interim and of order  $10^3$  grid-points for NCEP/NCAR, as an upper  
57 bound for  $D$ , our results point to the existence of a low dimensional attractor. The values of  $D$   
58 are comparable across all resolutions, except when the coarse graining degrades the resolution  
59 to the point where large-scale SLP low and high centers become unrecognizable (resolution  
60  $20^\circ$ ) and the phase space itself is shrinking. We find that the distribution of  $d(\zeta)$  for the  $2.5^\circ$   
61 NCEP/NCAR reanalysis ranges from as low as 3 to as high as 20 (Fig. 1-b). The average value  
62  $D = 13$  is only representative of a limited number of daily pressure fields.

63 Our estimate of  $D$  is not the dimension of the global climate attractor: here we focus on a

64 specific region and we analyse SLP structures with a well defined radius. The large spread  
65 in the distribution of instantaneous dimensions ( $3 < d(\zeta) < 20$ ) explains why deterministic  
66 low dimensional models are unable to reproduce the transients between metastable states of  
67 the atmospheric circulation, such as the transitions between zonal and blocked phases of the  
68 mid-latitude flow [1, 3].

69 Figure 1-c displays the time series of  $d(\zeta)$ . A seasonal cycle is identifiable both from the whole  
70 time series and from the inset which shows the last three years of instantaneous dimension  
71 estimates. The troughs occur in summer and the peaks in winter, when the temporal variability  
72 of the instantaneous dimension is also high. The corresponding results for ERA-Interim are  
73 shown in Extended Data Fig. A.3.

74 We now use  $d(\zeta)$  and  $\theta(\zeta)$  as probes to investigate the large-scale dynamics and associated  
75 weather extremes of the North Atlantic. Here we only discuss the results for the highest reso-  
76 lution NCEP/NCAR data (see Extended Data Figs. A.4 and A.5 for the ERA-Interim results).  
77 The analysis of idealised systems, such as the one proposed by Lorenz[7], suggests that ex-  
78 tremes in  $d$  and  $\theta$  trace extremes in phase space (see Extended Data Fig. A.1 and Methods). We  
79 begin by isolating all the days  $t$  such that the instantaneous dimension  $d$  and persistence  $\theta$  are  
80 beyond the 0.02 and 0.98 quantiles of the respective distributions. The results are insensitive to  
81 the exact choice of quantile. The North Atlantic SLP experiences a strong seasonal cycle with  
82 significant changes in the principal atmospheric patterns and modes of variability between the  
83 winter and summer seasons (e.g. ref. [10]). If extremes in the instantaneous properties have  
84 a direct correspondence to the large-scale circulation, a similarly pronounced seasonal cycle in  
85 their occurrence might be expected. Indeed, the dimensional extremes occur almost exclusively  
86 during the extended boreal winter period, as do the maxima of  $\theta$  (Fig. 2). The  $\theta$  minima, on the  
87 opposite, occur from early summer into winter. Since these are the most persistent patterns, the  
88 summertime episodes might be linked to the more staid character of the stable dynamics during

89 this season. Fig. 3 displays the pairs  $(\theta, d)$  for each day.

90 The SLP anomaly patterns corresponding to the extremes in the instantaneous dimension  $d(\zeta)$   
91 and persistence  $\theta(\zeta)$  are shown in the four side panels of Fig. 3. As expected, they correspond to  
92 very large SLP anomalies. More surprisingly, all four composites resemble well-known large-  
93 scale weather regimes [11, 12, 13]. The persistence of the patterns ranges between three days  
94 and just over one day ( $0.3 < \theta < 0.8$ ). This should not be compared directly to the persistence  
95 of the traditional weather regimes, as the requirement that the flow does not leave the neighbour-  
96 hood of the state  $\zeta$  is a more restrictive condition than that imposed by clustering algorithms.

97 The minima in daily dimension correspond to positive North Atlantic oscillation (NAO) condi-  
98 tions, which favour the occurrence of cyclones across the North Atlantic [14], and destructive  
99 surface winds over continental Europe [15]. Indeed, we find a statistically significant match  
100 between these low instantaneous dimensions and historical storms: the names and dates of storms  
101 corresponding to minima of  $d(\zeta)$  are reported in Fig. 4 (see also methods). Conversely, minima  
102 of  $\theta(\zeta)$  match the NAO- regime. Both NAO phases have a strong impact on the downstream  
103 temperature extremes over Europe [16]. However, we note that there are differences between  
104 the impacts of the traditional weather regimes and the patterns we identify here (see Extended  
105 Data Fig. A.6 and Methods). From the dynamical systems perspective, the negative NAO has  
106 a systematically longer persistence than the positive phase. Both the largest and the smallest  
107 NAO values mostly display below-average  $d(\zeta)$ , thus affording good predictability in accor-  
108 dance with recent analyses of ensemble prediction systems [17]. The maxima of  $\theta(\zeta)$  and  $d(\zeta)$   
109 are both associated with a blocked zonal flow and resemble the Atlantic Ridge and the Block-  
110 ing regimes, respectively. These are linked to European temperature extremes, although our  
111 patterns again display some novel connections (see Extended Data Fig. A.7 and Methods).

112 Blocked zonal flows are notoriously difficult to forecast [17]. This is due to the transition from  
113 persistent low instantaneous dimensions to less persistent, higher dimensions as the atmosphere

114 shifts from a zonal NAO-type flow to a blocked configuration.

115 We argue that these results can be used in an operational forecast system. If our dynamical  
116 indicators are indeed linked to predictability, given a certain state they should correlate with  
117 the skill of the model in predicting the future states of the field. We test this by correlating  
118 the values of  $d$  and  $\theta$  for the period 2000-2015 with the output of the 2nd Generation NOAA  
119 Global Ensemble Reforecast (GER) data set [18]. This is consistent with the current operational  
120 NCEP Global Ensemble Forecast System (GEFS), and consists of 11 ensemble members. The  
121 ensemble spread at different forecast hours is defined as:  $\langle\sigma_{SLP}\rangle$ , where angular brackets indi-  
122 cate spatial averages over the region considered in the analysis.. The correlation coefficient  $R$  is  
123 significant at all forecast steps at the 0.05 significance level, reaching  $R(\theta, \langle\sigma_{SLP}\rangle) = 0.42$  and  
124  $R(d, \langle\sigma_{SLP}\rangle) = 0.2$  (see Extended Data Fig. A.10-a for the complete analysis). While these  
125 values are not high for forecasts a short time ahead, their persistence at steps of over two weeks  
126 is noteworthy. Figure 5a), b) shows the bivariate histograms of  $(d, \langle\sigma_{SLP}\rangle)$  and  $(\theta, \langle\sigma_{SLP}\rangle)$  for  
127 for the forecast step of 384 hours ahead initialized on all days in the data set. There is a strong  
128 linear relationship between  $\theta$  and  $\langle\sigma_{SLP}\rangle$ , although the histogram suggests that the distribution  
129 is not uniform and peaks at either at low ( $\langle\sigma_{SLP}\rangle \simeq 400$  Pa) or high ( $\langle\sigma_{SLP}\rangle \simeq 800$  Pa) values.  
130 This latter feature might be dependent on the reforecast product. The analysis further suggests  
131 that  $\theta$  is a better proxy of ensemble spread  $\langle\sigma_{SLP}\rangle$  than  $d$ . The ensemble spread is a generic  
132 measure of the dispersion of the trajectories whereas  $d$  and  $\theta$  provide different but complemen-  
133 tary information on how the trajectories spread.  $d$  is linked to the entropy and therefore to the  
134 maximum divergence of the trajectories whereas  $\theta$  indicates how the trajectories stick together.  
135 A high stability implies that only few members strongly deviate from the bulk so that  $\langle\sigma_{SLP}\rangle$   
136 is linked to  $\theta$  (in most of the cases). On the rarer cases when most members strongly deviate  
137 from each other, we observe a positive correlation between  $\langle\sigma_{SLP}\rangle$  and  $d$  and a weakened link  
138 with  $\theta$ . For illustration, we report two examples of both situations in Fig. 5-c (low spread in the

139 ensemble reforecast), and Figure 5-d (high spread in the ensemble reforecast).  
140 The daily dimension and persistence therefore have an immediate practical use as proxies for  
141 predictability, and could be fruitfully used as a time-effective way in which to evaluate and in-  
142 form operational forecast systems. For example, one could imagine a system which determines  
143 the resolution and ensemble size for a given initialization step based on the values of  $d$  and  $\theta$ .  
144 Moreover, the visualization of the trajectory for a given season (see the supplementary video)  
145 can be used to provide a day-by-day tracking of weather extreme events.

## 146 **Acknowledgments**

147 D. Faranda and P. Yiou were supported by ERC grant No. 338965-A2C2. DF acknowledges V  
148 Lucarini, B Dubrulle, F Daviaud, A Jezequel, MC Alvarez-Castro and for useful discussions.  
149 ECMWF data were accessed at <http://apps.ecmwf.int/datasets/>. NCEP/NCAR  
150 data was obtained at: [http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.](http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml)  
151 [shtml](http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml).

## 152 **Author contributions**

153 DF conceived and designed the study. DF and GM performed the dynamical systems analysis,  
154 and interpreted the results. DF further tested the technique on idealized systems and compiled  
155 the storms database. GM performed the analysis on extreme weather. All authors contributed  
156 to the interpretation of the results and the editing of the manuscript.

## 157 **References and Notes**

158 [1] Lorenz, E. N. Attractor sets and quasi-geostrophic equilibrium . *J. Atmos. Sci.* **37**, 1685–  
159 1699 (1980).



- 160 [2] Grassberger, P. Do climatic attractors exist? . *Nature* **323**, 609–612 (1986).
- 161 [3] Vautard, R. & Ghil, M. Singular spectrum analysis in nonlinear dynamics, with applica-  
162 tions to paleoclimatic time series . *Physica D* **35**, 395–424 (1989).
- 163 [4] Lucarini, V. *et al.* *Extremes and Recurrence in Dynamical Systems*. Pure and Applied  
164 Mathematics: A Wiley Series of Texts, Monographs and Tracts (Wiley, 2016). URL  
165 <https://books.google.fr/books?id=ebTOoQEACAAJ>.
- 166 [5] Young, L.-S. Dimension, entropy and Lyapunov exponents . *Ergod. Theor. Dyn. Syst.* **2**,  
167 109–124 (1982).
- 168 [6] Takens, F. *Detecting strange attractors in turbulence* (Springer, 1981).
- 169 [7] Lorenz, E. N. Deterministic nonperiodic flow . *J. Atmos. Sci.* **20**, 130–141 (1963).
- 170 [8] Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project . *Bull. Am. Meteorol. Soc.*  
171 **77**, 437–471 (1996).
- 172 [9] Dee, D. *et al.* The ERA-Interim reanalysis: Configuration and performance of the data  
173 assimilation system . *Quart. J. Roy. Meteor. Soc.* **137**, 553–597 (2011).
- 174 [10] Folland, C. K. *et al.* The summer North Atlantic Oscillation: past, present, and future . *J.*  
175 *Clim.* **22**, 1082–1103 (2009).
- 176 [11] Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and  
177 precipitation . *Science* **269**, 676–679 (1995).
- 178 [12] Yiou, P. & Nogaj, M. Extreme climatic events and weather regimes over the North At-  
179 lantic: When and where? . *Geophys. Res. Lett.* **31(7)** (2004).

- 180 [13] Moore, G., Renfrew, I. A. & Pickart, R. S. Multidecadal mobility of the North Atlantic  
181 oscillation . *J. Clim.* **26**, 2453–2466 (2013).
- 182 [14] Gómara, I., Rodríguez-Fonseca, B., Zurita-Gotor, P. & Pinto, J. G. On the relation between  
183 explosive cyclones affecting Europe and the North Atlantic Oscillation . *Geophys. Res.*  
184 *Lett.* **41**, 2182–2190 (2014).
- 185 [15] Messori, G. & Caballero, R. On double Rossby wave breaking in the North Atlantic . *J.*  
186 *Geophys. Res.* **120(21)** (2015).
- 187 [16] Rogers, J. C. North Atlantic storm track variability and its association to the North Atlantic  
188 Oscillation and climate variability of northern Europe . *J. Clim* **10**, 1635–1647 (1997).
- 189 [17] Ferranti, L., Corti, S. & Janousek, M. Flow-dependent verification of the ECMWF en-  
190 semble over the Euro-Atlantic sector . *Quart. J. Roy. Meteor. Soc.* **141**, 916–924 (2015).
- 191 [18] Hamill, T. M. *et al.* NOAA’s second-generation global medium-range ensemble reforecast  
192 dataset . *Bull. Am. Meteorol. Soc.* **94**, 1553–1565 (2013).
- 193 [19] Grassberger, P. & Procaccia, I. Dimensions and entropies of strange attractors from a  
194 fluctuating dynamics approach . *Physica D* **13**, 34–54 (1984).
- 195 [20] Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. & Shraiman, B. I. Fractal  
196 measures and their singularities: the characterization of strange sets . *Phys. Rev. A* **33**,  
197 1141 (1986).
- 198 [21] Lorenz, E. N. Dimension of weather and climate attractors . *Nature* **353**, 241–244 (1991).
- 199 [22] Freitas, A. C. M., Freitas, J. M. & Todd, M. Hitting time statistics and extreme value  
200 theory . *Probab. Theory Rel.* **147**, 675–710 (2010).

- 201 [23] Lucarini, V., Faranda, D. & Wouters, J. Universal behaviour of extreme value statistics for  
202 selected observables of dynamical systems . *J. Stat. Phys.* **147**, 63–73 (2012).
- 203 [24] Pickands III, J. Statistical inference using extreme order statistics . *Ann. Stat.* 119–131  
204 (1975).
- 205 [25] Grassberger, P. & Procaccia, I. Characterization of strange attractors . *Phys. Rev. Lett.* **50**,  
206 346 (1983).
- 207 [26] Comas-Bru, L. & McDermott, F. Impacts of the EA and SCA patterns on the European  
208 twentieth century NAO–winter climate relationship . *Quart. J. Roy. Meteor. Soc.* **140**,  
209 354–363 (2014).
- 210 [27] Cassou, C. Intraseasonal interaction between the Madden–Julian oscillation and the North  
211 Atlantic oscillation . *Nature* **455**, 523–527 (2008).

## 212 **Methods**

213 Attractors are geometrical sets which hosts all the trajectories of a system. To characterize an  
214 attractor, one wants to know how often the state  $\zeta$  occurs over a certain time interval and how  
215 long the dynamics stick to  $\zeta$  before leaving its neighbourhood. If one is able to specify such  
216 properties for all the points of the attractor, then the behaviour of the system is entirely known.  
217 The general problem one faces in reconstructing an attractor is the limited number of trajectories  
218 that can be observed or simulated [2]. In the case of climate observations we have just one single  
219 trajectory  $x(t)$  (here represented by the time series of SLP daily fields) that we can exploit to  
220 reconstruct the attractor properties.

221 The purpose of our methodology it is to use just a long trajectory  $x(t)$  of systems states to  
222 reconstruct the salient properties of the attractor. The method is based on the link between

223 extreme value theory (where the extremes are the recurrences of the points  $\zeta$  with respect to  
224 all the possible states of the system) and the Poincaré theorem of recurrence. The idea is that  
225 each state of the system  $x(t)$  approximates a point  $\zeta$  on the attractor and its neighbours are  
226 all the states whose distance with respect to  $x(t)$  is small. So at each time  $t$  and for each  
227 state  $x$  observed we can define instantaneous properties: the instantaneous dimension  $d$  and the  
228 stickiness  $\theta$ . For the theoretical details, demonstrations and examples on dynamical systems see  
229 [4]. These properties are instantaneous because they change at each instant  $t$ , but they are also  
230 local, because states observed at different times but close in the phase space will have similar  
231 instantaneous properties. We refer to instantaneous dimensions rather than to local dimension  
232 to avoid ambiguity with the notion of local to indicate a geographic region.

### 233 **Instantaneous Dimensions**

234 The distribution of instantaneous dimension of the attractor of a dynamical system gives use-  
235 ful information on the predictability of observed states because it is related to the Lyapunov  
236 exponents [5]. Therefore, estimating the dimension distribution in phase space helps charac-  
237 terizing the overall dynamics of the system. The embedding methods developed in the 1980's  
238 [19, 20] do not provide instantaneous dimensions but only the average dimension of the attrac-  
239 tor. Moreover, such computations have proved to be problematic in systems with large numbers  
240 of degrees of freedom and have given controversial results when applied to atmospheric flows  
241 [2, 21].

242 The method we adopt in the present study results from the application of extreme value theory  
243 to the field of dynamical systems [22, 4]. In this approach, the points on the attractor are fully  
244 characterized by parameters of extreme value laws: if one fixes an arbitrary point  $\zeta$  on a chaotic  
245 attractor and consider the probability  $P$  that a trajectory  $x(t)$  returns within a sphere of radius  
246  $\epsilon$  centered on the point  $\zeta$ , then the Freitas-Freitas-Todd theorem [22] modified in [23] states

247 that such probability is a generalized Pareto distribution [24]. The time series of the distance  
248 between  $\zeta$  and the other observations along the trajectory is defined by:

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

249  $\delta(x, y)$  is a distance function between two vectors, which tends to zero when  $x$  and  $y$  are close to  
250 each other. Taking the logarithm increases the discrimination of small values of  $\delta(x, y)$  which,  
251 as described below, correspond to large values of  $g(x(t))$ . The probability of logarithmic returns  
252 can then be expressed as:

$$P(g(x(t)) > q, \zeta) \simeq \exp\left[-\frac{x - \mu(\zeta)}{\sigma(\zeta)}\right]$$

253 namely an exponential law whose parameters  $\mu$  and  $\sigma$  depend on the point  $\zeta$  chosen on the  
254 attractor. Remarkably,  $\sigma(\zeta) = 1/d(\zeta)$ , where  $d(\zeta)$  is the dimension around the point  $\zeta$ . This  
255 result has been proved theoretically and verified numerically in several studies collected in [4].  
256 In the above equation,  $q$  is a high threshold, and is linked to the radius  $\epsilon$  via  $q = g^{-1}(\epsilon) =$   
257  $\exp(-\epsilon)$ . In other words, requiring that the trajectory falls within a sphere around the point  $\zeta$  is  
258 equivalent to asking that the series of  $g(x(t))$  is over the threshold  $q$ , which can be simply set  
259 as a percentile of the series itself. If this approach is iterated for several different  $\zeta$  points, the  
260 attractor dimension is then obtained as:

$$D = \overline{d(\zeta)}$$

261 where the overbar means averaging over all  $\zeta$ . This is a powerful result because it provides a  
262 direct way to compute dimensions on the attractor without the need for embedding.

## 263 **Stickiness (persistence) in phase space**

The previous results hold when the state  $\zeta$  considered is not in the vicinity of a fixed point of the attractor. Fixed points are such that  $x(t+1) = x(t)$ , for all  $t$ , i.e. the system is stuck in the same state for an infinite time. In most natural systems, fixed points are unstable: a trajectory passing close to a fixed point spends a finite amount of time in its vicinity before leaving. Such time can also be computed by introducing a further parameter in the previous law. This parameter, known as extremal index, is indicated with  $\theta$  and is such that:

$$P(g(x(t)) > q \simeq \exp \left[ -\theta \left( \frac{x - \mu(\zeta)}{\sigma(\zeta)} \right) \right]$$

264 We can interpret  $\theta$  as a measure of the residence time of consecutive iterations in the small  
265 sphere around the point  $\zeta$ . In this interpretation, the inverse of  $\theta$  is precisely the mean residence  
266 time within the sphere.

## 267 **Some idealized examples**

268 We illustrate the procedure to compute the instantaneous dimension described above by apply-  
269 ing it to the Lorenz system [7]. This system and its attractor (often referred to as the Lorenz  
270 butterfly) has been studied extensively in the literature, and therefore allows us to compare the  
271 results of our approach to those of standard techniques in dynamical systems analysis. We be-  
272 gin by generating a trajectory  $\vec{x}(t)$ , using a time step of 0.025. Next, we select approximately  
273 75,000 locations along the trajectory as our  $\zeta$  points on the attractor.

274 For each  $\zeta$ : i) the series  $g(\vec{x}(t), \zeta)$  is computed, ii) a high threshold  $q$  is selected (here the 98th  
275 percentile of the series  $g$ ), iii) a Generalized Pareto distribution is fitted to the observations  
276 exceeding the threshold  $q$ , iv) an instantaneous dimension of the attractor  $d(\zeta)$  is then obtained.  
277 Extended Data Fig. A.1) displays the values of the instantaneous dimension at all points along  
278 our trajectory, while panel (b) displays the corresponding histogram. It is interesting to observe

279 that the minima and maxima of the instantaneous dimension track the extremes of the Lorenz [7]  
280 attractor. Maxima of the dimension have a non-trivial structure and are found where recurrences  
281 are rare — namely in the wings of the butterfly — and where the trajectories diverge the most —  
282 namely between the two wings. The minima correspond to the centre of the butterfly wings, i.e.  
283 the fixed points of the Lorenz 63 system. The average value of all the  $d(\zeta)$  is, by definition, the  
284 attractor dimension  $D$ . The value we find:  $D = 2.06$ , corresponds exactly to the value proposed  
285 by Grassberger and Procaccia [25]. For this specific example, any  $q$  larger than the 95<sup>th</sup> quantile  
286 of  $g$  yields the same results. The stickiness, measured in terms of  $\theta$  is instead shown in Extended  
287 Data Fig. A.1). The Lorenz attractor consists of three unstable fixed points: two at the center of  
288 the wings and one at the origin of the axes. The three points are well captured by the statistics  
289 of  $\theta$ .

290 The embedding methodologies adopted in the 1980s were unable to estimate high attractor  
291 dimensions [2], thus providing artificially low values for complex systems. To verify that our  
292 methodology does not suffer from the same bias, we have applied it to test fields of the same  
293 grid-size as the NCEP data set. For completely random fields, we have obtained estimates for  
294  $D$  larger than 80. Conversely, estimates of  $D$  for fields with coherent structures resembling  
295 pressure centres were lower than 5 (see Extended Data Fig. A.2).

## 296 **Sea-level pressure data**

297 In this study we adopt sea-level pressure (SLP) as the meteorological variable to describe the  
298 North Atlantic circulation. The major modes of variability affecting the North Atlantic are  
299 often defined in terms of the empirical orthogonal functions of SLP [13, 11], and a wealth  
300 of other atmospheric features, ranging from teleconnection patterns to storm track activity to  
301 atmospheric blocking can be diagnosed from the SLP field [26, 16].

302 We base our study on NCEP/NCAR reanalysis data [8] over the period 1948-2015, with a hori-

303 zontal resolution of  $2.5^\circ$  and ERA-Interim data [9] over 1979-2011 with horizontal resolutions  
304 of  $0.75^\circ$  and  $1.5^\circ$  respectively. We consider a domain spanning the North Atlantic and Europe  
305 ( $80^\circ \text{ W} \leq \text{Long.} \leq 50^\circ \text{ E}$ ,  $22.5^\circ \text{ N} \leq \text{Lat.} \leq 70^\circ \text{ N}$ ). Further tests show that the results are  
306 linearly insensitive to the exact boundaries chosen (Extended Data Fig. A.8).

307 For the NAO, we use daily values computed by NCEP's Climate Prediction Center. The values  
308 are based on NAO patterns which vary on a monthly basis, and cover the full year. The data is  
309 freely available from: <http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml>

### 310 **Statistical significance and robustness of the results**

311 For general problems where the value of  $D$  is not known a priori, the appropriateness of the  
312 value of  $q$  for a given  $\zeta$  can be tested using a number of statistical approaches. In the present  
313 study we use the Anderson-Darling test [28] to reject the hypothesis that all  $g(x(\vec{t}), \zeta) > q$   
314 come from a generalized Pareto distribution. The test can be repeated for each  $\zeta$  to obtain a  
315 statistical confidence level for the chosen  $q$ . All the results displayed in the paper use  $q = 0.98$ ,  
316 which satisfies the Anderson-Darling test at the 0.05 confidence level for more than 95% of  
317 the chosen  $\zeta$ . We further performed a visual inspection of the results, and found them to be  
318 stable for  $0.99 \geq q \geq 0.975$ . In fact, if a too high threshold is selected, the number of values  
319 exceeding it is insufficient to successfully fit the generalized Pareto distribution.

320 The results presented in Fig. 2 and 3 in the main text for the NCEP/NCAR data have been  
321 repeated for the ERA-Interim reanalysis, and are shown in Extended Data Figs. A.4 and A.5  
322 for both the  $0.75^\circ$  and  $1.5^\circ$  resolutions. It can be seen that both the histograms and the seasonal  
323 variability of the instantaneous dimensions are very similar across the three data sets.



## 324 **Storm database**

325 We present a database of historical storms (see Supplementary Information) which affected  
326 Europe between 1948 and 2015. This database is largely based on the Lamb [29] and the  
327 Roberts et al. [30] catalogues. Additional storms have been integrated because of their relevance  
328 in terms of human losses, damages or their profile in the media. This results in a total of 336  
329 storms. The database is organized in four columns: 1) the day of occurrence in the format  
330 yyyyymmdd, 2) the name(s) of the storm, 3) the countries or the region affected and 4) a reference  
331 to a peer-reviewed article, a report or a press article describing the importance of the storm.  
332 As a caveat to our methodology, we note that the increasing coverage of both meteorological  
333 instruments and technological means of information results in an increasing number of storms  
334 with time (whereas the minima of the instantaneous dimensions are equally distributed over  
335 time).

336 We use the database to evaluate the correspondence between the minima of the instantaneous  
337 dimensions, which equates to positive NAO-like SLP anomalies, and storminess. A day is said  
338 to match a storm if it falls within 2 days of the date specified in the database. Among the 481  
339 minima (corresponding to the quantile 0.02) identified in the NCEP/NCAR data, 17% (namely  
340 82 days) match historical storms recorded in our database. This is a very high percentage: if  
341 an equal number of random days is selected (from the extended winter period over which the  
342 dimensional extremes occur), statistically only 2.5% will match a historical storm. Examples  
343 of well-known storms matching dimensional extremes include Dirk and Herta (see Fig. 4).

## 344 **Temperature and precipitation extremes**

345 Weather regimes can explain a large part of the statistical distribution for surface variables, and  
346 have been linked to anomalies in the frequencies of extreme weather events [12]. Here we show  
347 that the extremes in phase space we discuss in the main text have a direct link to the occurrence

348 of extreme weather events. At the same time, we note that there are some differences relative  
349 to previous analyses based on the traditional weather regimes. Extreme events are defined by  
350 the 0.98 quantile of the distribution of the anomalies for each gridbox and variable during an  
351 extended winter season (September-April). The latter matches the period of occurrence of the  
352 vast majority of the phase space extremes. A value of 1 means that the selected days are not  
353 discriminating for extremes, while 0 means that there is no chance of an extreme occurring on  
354 those days. A value of 2 indicates that the extreme is twice as likely to occur relative to the  
355 wintertime climatology. Similarly, a value of 0.5 corresponds to a halved frequency. Statistical  
356 significance is evaluated at the 5% level using a montecarlo method. This is indicated in the  
357 figures by the grey shading.

358 Extended Data Fig. A.6 displays the relative changes in the frequency of extreme cold events for  
359 (a) high and (b) low  $d(\zeta)$ . High dimensions show enhanced cold spell frequencies over Scandi-  
360 navia and northern Europe and the Mediterranean basin. The large signal over Scandinavia is  
361 not typically associated with blocking [12], and is possibly linked to the more eastern location  
362 of the high in our pattern. The low dimensions shows very large increases in cold extremes  
363 over the western North Atlantic and Greenland and south-eastern Europe, matching closely the  
364 temperature anomaly footprint of the positive NAO. The link with rainfall extremes (Extended  
365 Data Fig. A.6-c,d is less clear. For the high dimensions, significant changes in Europe are lim-  
366 ited to regional decreases in extreme wet days over the British Isles and Scandinavia. The low  
367 dimensions display instead regional increases over Scandinavia and continental Europe.

368 Extended Data Fig. A.7 displays the relative changes in the frequency of extreme cold events for  
369 (a) high and (b) low  $\theta$ . High  $\theta$  events correspond to frequency increases over the Mediterranean  
370 region, in agreement with a similar analysis by Cassou [27]. The low  $\theta$  match instead a strongly  
371 enhanced likelihood of cold extremes over the British Isles and Scandinavia, similarly to the  
372 impact of the negative NAO phase. Extended Data Figs. A.7 c, d displays the results for wet

373 extremes, which again display less significant links. The most robust feature over Europe is a  
374 decrease in the frequency of extreme wet days over the British Isles and Northern France for  
375 low  $\theta$  values. In contrast, previous analyses have highlighted the role of a positive NAO in  
376 driving wet extremes over these areas[12, 27] .

## 377 **Methods References**

378 [29] Anderson, T. W. & Darling, D. A. Asymptotic theory of certain goodness of fit criteria  
379 based on stochastic processes. *Ann. Math. Stat.* 193–212 (1952).

380 [30] Lamb, H. & Frydendahl, K. *Historic Storms of the North Sea, British Isles and Northwest*  
381 *Europe* (Cambridge University Press, 1991).

382 [31] Roberts, J. et al. The XWS open access catalogue of extreme European windstorms from  
383 1979 to 2012 . *Nat. Hazards and Earth Syst. Sci.* **14**, 2487–2501 (2014).

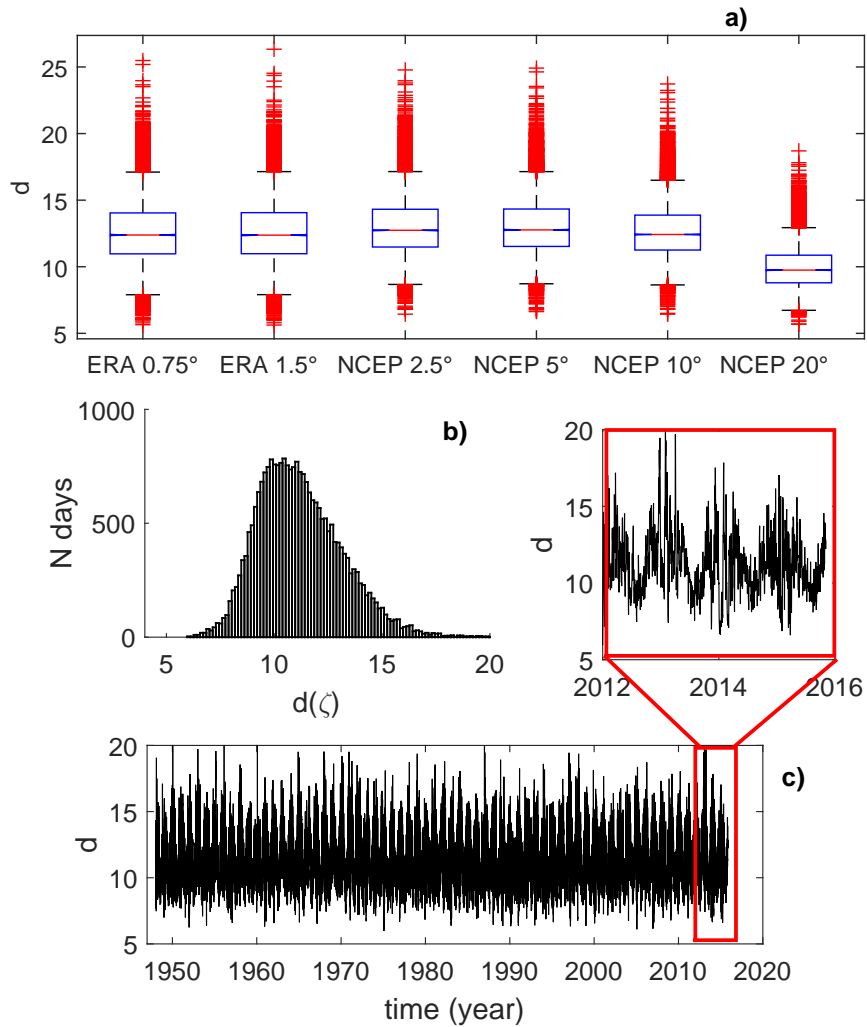


Figure 1: **Analysis of the distribution of instantaneous dimensions.** a) Box-plots of the distribution of  $d(\zeta)$  for different resolutions (in degrees of longitude and latitude). In each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers and outliers are plotted individually. b) Histogram of the daily dimension  $d(\zeta)$  for the NCEP reanalysis. c) Time series of the instantaneous dimensions and inset showing the last 3 years.

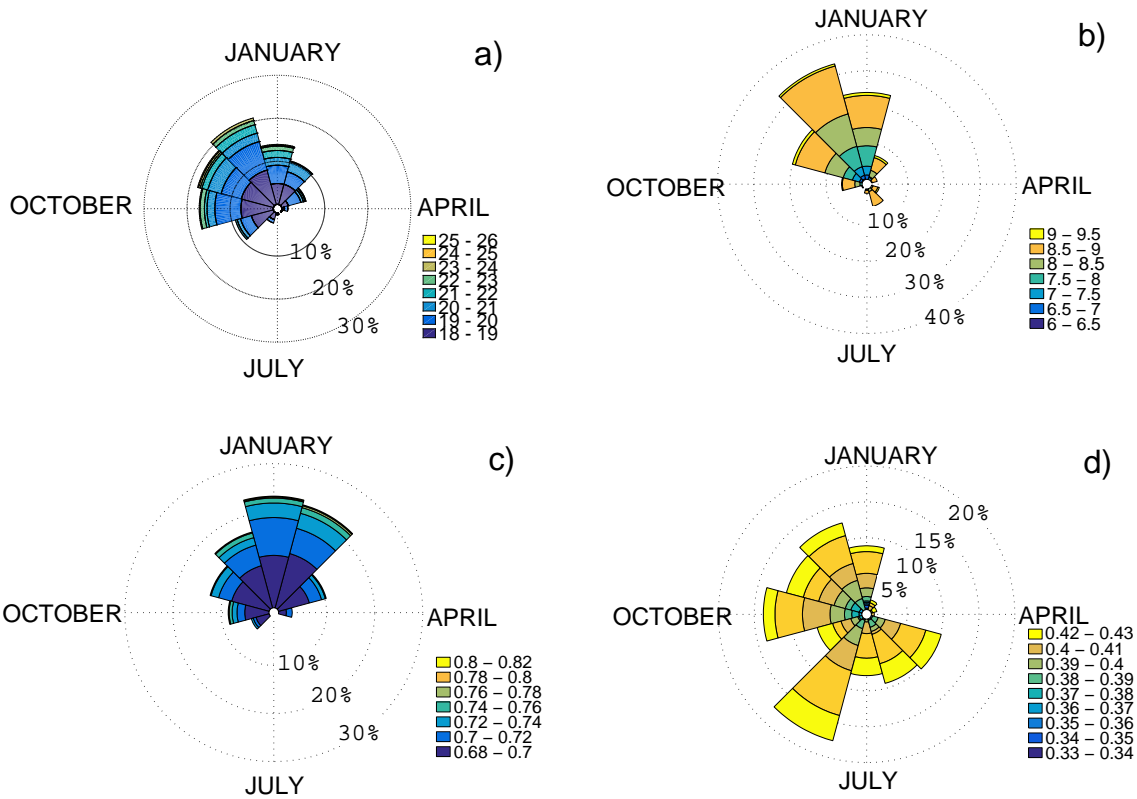


Figure 2: **Monthly distribution of the instantaneous properties exceeding the 0.02 and 0.98 quantiles of their respective distributions.** Maxima of a)  $d$  and (b)  $\theta$  and minima of (c)  $d$  and  $\theta$  (d) for the NCEP reanalysis. The percentage values indicate the occurrences in each month. The colourscale refers to the values of the quantities.

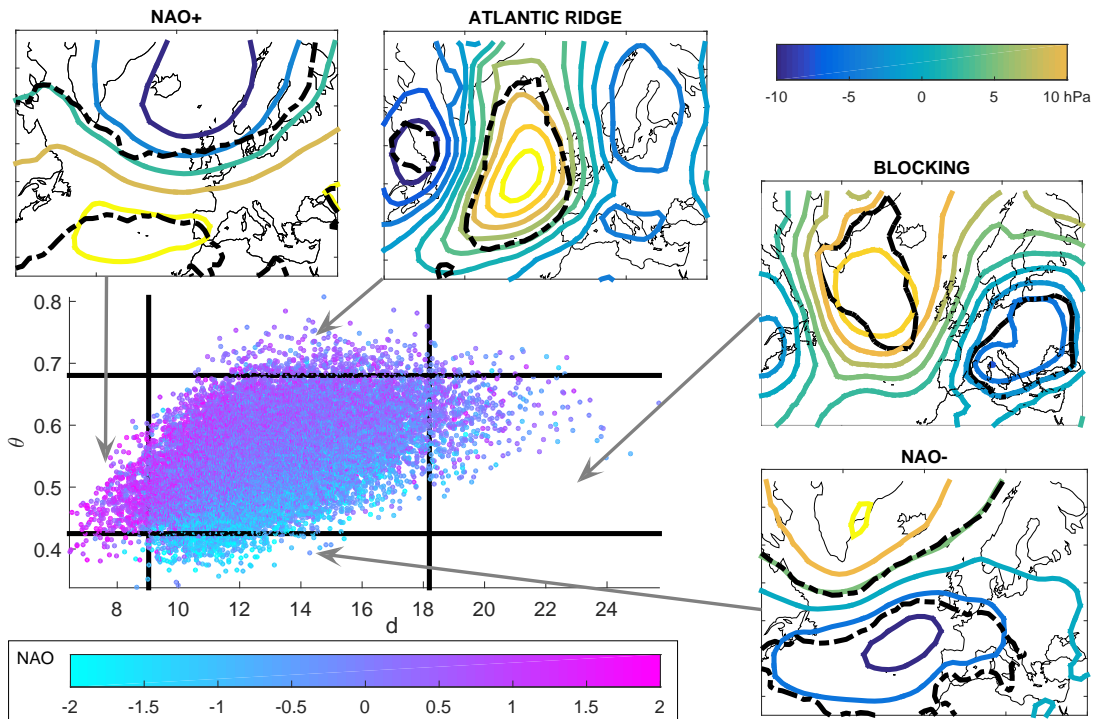


Figure 3: **Dynamical systems analysis for the NCEP reanalysis.** The scatter plot displays the daily values of the instantaneous dimension  $d$  and the persistence  $\theta$  of the field. The NAO value for that day is indicated by the colourscale (increment of four years). The black solid lines mark the 0.02 and 0.98 quantiles of the  $d$  and  $\theta$  distributions. 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: NAO+ (minima of  $d$ ), NAO- (minima of  $\theta$ ), Atlantic Ridge (maxima of  $\theta$ ), Blocking (maxima of  $d$ ). The black lines indicates the region where at least the 2/3 of extreme pressure anomalies have the same sign.

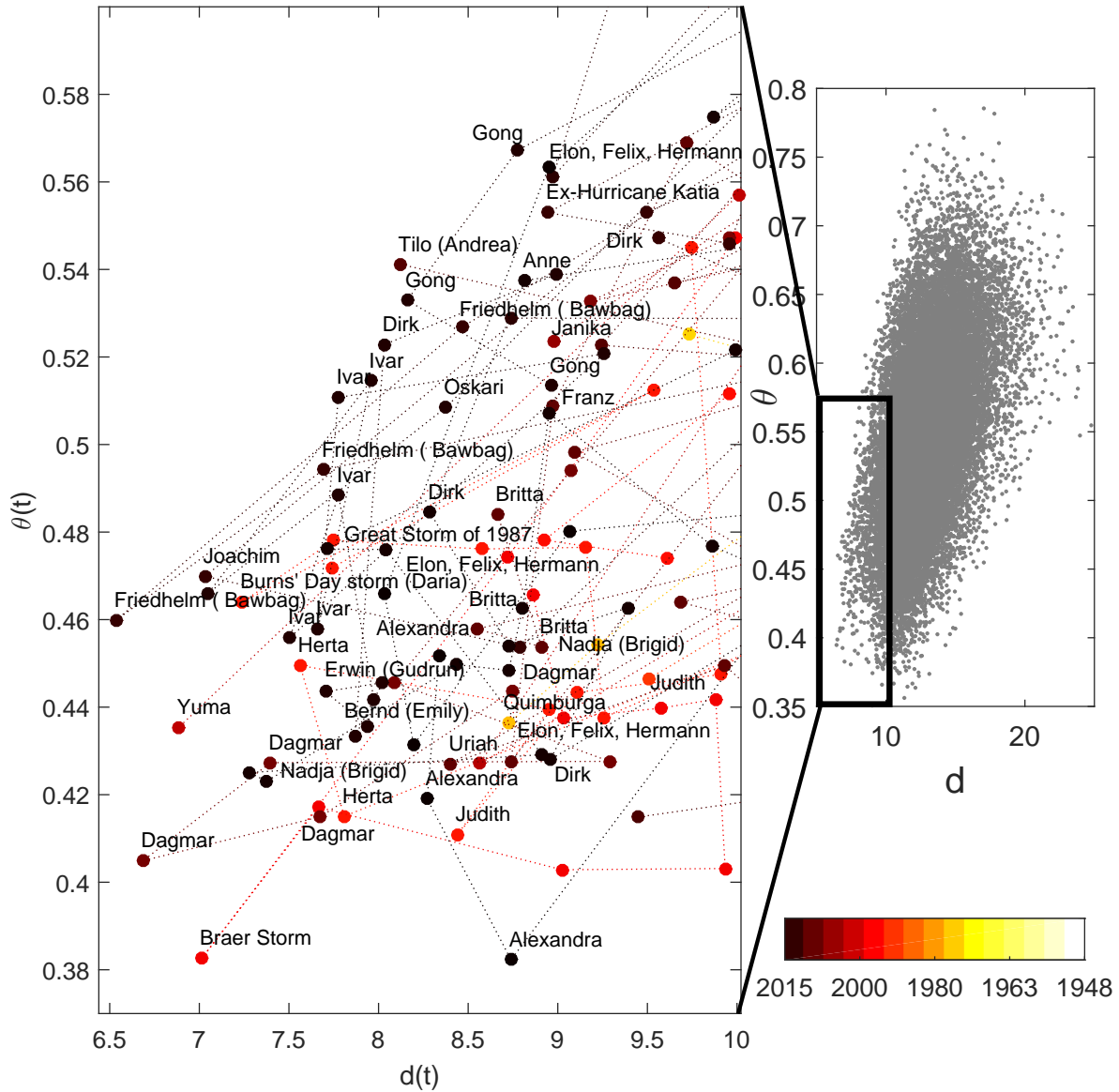


Figure 4: **Storms matching the minima of the instantaneous dimensions.** Instantaneous dimensions  $d$  (x-axis) and persistence  $\theta$  (y-axis) are plotted along with names and years (colourscale). The inset shows the full distribution of  $d, \theta$  values. Repeated names indicate storms which persist for several days (see supplementary storm database).

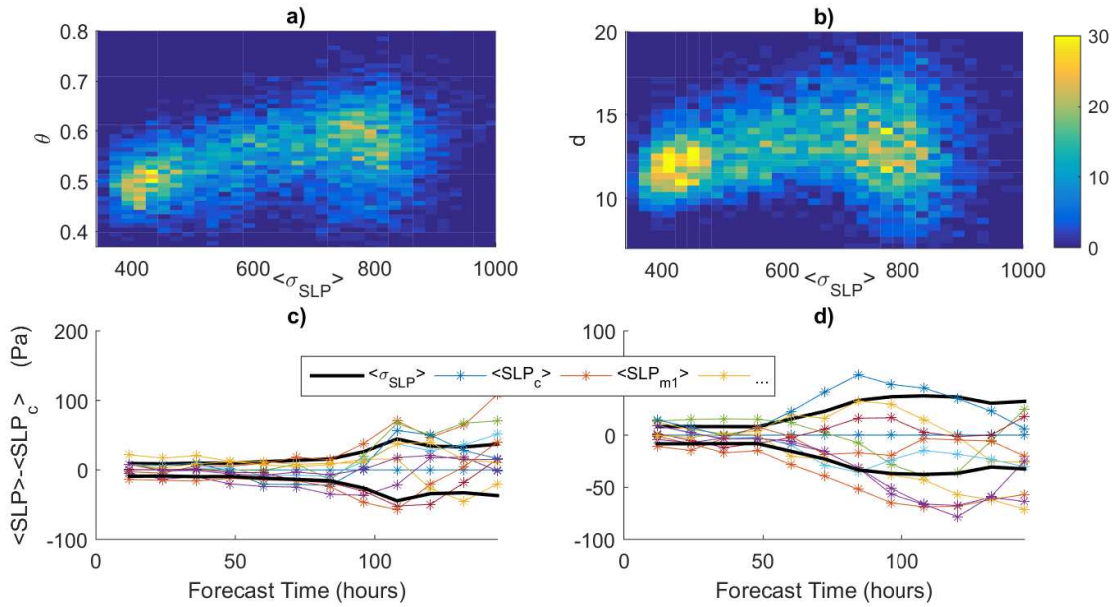


Figure 5: **Analysis of the relation between instantaneous properties and NOAA GER re-forecast.** a) and b) bivariate histograms of the ensemble spread  $\langle \sigma_{SLP} \rangle$  at step +384h as a function of the stability  $\theta$  (a) and the instantaneous dimension  $d$  (b) of the initialisation field. for the period 2000-2015. Colourscale indicates the number of days with the same pair of parameters. c) and d): case studies on how the Reforecast trajectories diverge from the control run  $\langle SLP_c \rangle$  and their relation with ensemble spread,  $d$  and  $\theta$ . c) Reforecast for the 21/01/2000 corresponding to high  $d \simeq 18$  with a moderate value of  $\theta = 0.56$ . d) Reforecast for the 12/01/2000 corresponding to high  $\theta = 0.73$  and moderate value of  $d=14$ .