

Evaluation of the HadGEM3-A simulations in view of detection and attribution of human influence on extreme events in Europe

Article

Accepted Version

Vautard, R., Christidis, N., Ciavarella, A., Alvarez-Castro, C., Bellprat, O., Christiansen, B., Colfescu, I., Cowan, T., Doblas-Reyes, F., Eden, J., Hauser, M., Hegerl, G., Hempelmann, N., Klehmet, K., Lott, F., Nangini, C., Orth, R., Radanovics, S., Seneviratne, S. I., van Oldenborgh, G. J., Stott, P., Tett, S., Wilcox, L. and Yiou, P. (2019) Evaluation of the HadGEM3-A simulations in view of detection and attribution of human influence on extreme events in Europe. *Climate Dynamics*, 52 (1-2). pp. 1187-1210. ISSN 0930-7575 doi: <https://doi.org/10.1007/s00382-018-4183-6> Available at <https://centaur.reading.ac.uk/76175/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1007/s00382-018-4183-6>

Publisher: Springer

including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

[Click here to view linked References](#)

1 **Evaluation of the HadGEM3-A simulations in view of**
2 **detection and attribution of human influence on extreme**
3 **events in Europe**

4
5 Robert Vautard (1), Nikolaos Christidis (2), Andrew Ciavarella (2), Carmen Alvarez-Castro (1), Omar
6 Bellprat (3), Bo Christiansen (4), Ioana Colfescu (5), Tim Cowan (5), Francisco Doblas-Reyes (3),
7 Jonathan Eden (6), Mathias Hauser (7) Gabriele Hegerl (5), Nils Hempelmann (1), Katharina Klehmet
8 (8), Fraser Lott (2), Cathy Nangini (1), René Orth (7) , Sabine Radanovics (1), Sonia I. Seneviratne (7),
9 Geert Jan van Oldenborgh (6), Peter Stott (2), Simon Tett (5), Laura Wilcox (9), Pascal Yiou (1)

10
11 (1) Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Université
12 Paris-Saclay, Gif sur Yvette, France

13 (2) UK Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK

14 (3) Barcelona Supercomputing Center, Barcelona, Spain

15 (4) Danish Meteorological Institute, Copenhagen, Denmark

16 (5) School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK

17 (6) Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

18 (7) Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

19 (8) Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

20 (9) Department of Meteorology, University of Reading, P. O. Box 243, Earley Gate, Reading RG6 6BB, UK

21

22 Abstract

23

24 A detailed analysis is carried out to assess the HadGEM3-A global atmospheric model skill in simulating
25 extreme temperatures, precipitation and storm surges in Europe in the view of their attribution to
26 human influence. The analysis is performed based on an ensemble of 15 atmospheric simulations
27 forced with observed Sea Surface Temperature of the 54 year period 1960–2013. These simulations,
28 together with dual simulations without human influence in the forcing, are intended to be used in
29 weather and climate event attribution. The analysis investigates the main processes leading to extreme
30 events, including atmospheric circulation patterns, their links with temperature extremes, land-
31 atmosphere and troposphere-stratosphere interactions. It also compares observed and simulated
32 variability, trends and generalized extreme value theory parameters for temperature and precipitation.

33 One of the most striking findings is the ability of the model to capture North-Atlantic atmospheric
34 weather regimes as obtained from a cluster analysis of sea level pressure fields. The model also
35 reproduces the main observed weather patterns responsible for temperature and precipitation
36 extreme events. However, biases are found in many physical processes. Slightly excessive drying may
37 be the cause of an overestimated summer interannual variability and too intense heat waves, especially
38 in central/northern Europe. However, this does not seem to hinder proper simulation of summer
39 temperature trends. Cold extremes appear well simulated, as well as the underlying blocking frequency
40 and stratosphere-troposphere interactions. Extreme precipitation amounts are overestimated and too
41 variable. The atmospheric conditions leading to storm surges were also examined in the Baltics region.
42 There, simulated weather conditions appear not to be leading to strong enough storm surges, but
43 winds were found in very good agreement with reanalyses. The performance in reproducing
44 atmospheric weather patterns indicates that biases mainly originate from local and regional physical
45 processes. This makes local bias adjustment meaningful for climate change attribution.

46

47

48 1. Introduction

49

50 In recent years attribution of changing likelihood of weather events has motivated an outstanding
51 effort of the climate science community (Stott et al., 2016). While detecting trends in odds of extreme
52 events (eg. as characterized by the exceedance of a threshold) can draw solely on observational data,
53 formal attribution to human activities requires comparing statistics in a “current climate” world and in
54 a world where human activities have not occurred. This requires model simulations with different sets
55 of assumptions concerning external forcing. This also requires that the models used are able to
56 simulate the changes in likelihood of extremes by comparing with observations, which is often difficult
57 in practice due to the short length and lack of homogeneity of observational data sets. A simplification
58 is often made with the assumption that the anthropogenic effect is included in surface variables such
59 as SST, sea ice (Pall et al., 2011) or soil moisture (Hauser et al., 2016), and in atmospheric composition,
60 and that extreme events respond to this influence through processes linking surface and atmosphere.
61 In contrast, attribution of observed trends to causes relies on analysis of the observed change with the
62 help of climate models, hence is more directly anchored to the observed change (see NAS report, 2016;
63 Hegerl and Zwiers, 2011). In practice, anthropogenic forcing influence on temperature-related variables
64 is such that changes are found with a high consistency using both approaches for trends in mean and
65 extremes (Bindoff et al., 2013).

66 Attribution makes one unavoidable assumption: that dynamical and physical processes are correctly
67 represented in the climate model used for attribution. If all processes are well accounted for,
68 sensitivities to forcing changes should be realistic. Attribution of weather events therefore requires a
69 careful evaluation of processes involved in the build-up of the events. Evaluation also requires
70 examination of extreme events statistics, and if possible their change with increasing greenhouse gases
71 and other human-driven changes (Bellprat and Doblas-Reyes, 2016; Lott and Stott, 2016; Sippel et al.,
72 2016).

73 This study examines how the newly upgraded Hadley Centre Global Environmental Model version 3-
74 Atmosphere (HadGEM3-A) atmospheric model performs in view of event attribution in Europe, with a
75 focus on processes leading to extreme events. The earlier, lower resolution, version of the model was
76 employed in several attribution studies of extreme events including consecutive cold winters in the UK,
77 the Moscow heatwave in July 2010 (Christidis et al., 2013a), the severe East African drought of 2011
78 (Lott et al., 2013), the Eastern Australia floods of 2011 (Christidis et al., 2013b) and the cold spring of
79 2013 in the UK (Christidis et al., 2014). These analyses quantified the effect of anthropogenic influence
80 on the likelihood of the events as well as the associated uncertainty from limited number of available
81 simulations. Moreover, simple evaluation assessments were carried out to demonstrate that the model
82 was fit for purpose and able to realistically represent the type of extremes under consideration in the
83 region of interest. Angelil et al. (2016) compared the simulated extreme events with reanalyses
84 datasets at relatively high resolution and found mismatches among all sets (models and reanalyses and
85 among reanalyses themselves). This highlights observational difficulties when comparing sub-regional
86 trends using reanalyses, and emphasizes the need to not only evaluate statistical properties but also
87 physical mechanisms involved in the trends.

88 Here, the new ensemble of simulations is evaluated through comparison with available observations.
89 These simulations are now used in several attribution studies (e.g. Eden et al, 2016; van Oldenborgh et
90 al., 2017, in preparation; Philip et al., 2017, Hauser et al, 2017, Klehmet et al, 2017, in preparation,
91 Eden et al, 2017, Christiansen et al, 2017, in preparation, Wilcox et al., 2017), where evaluation is
92 carried out for the local case study. However, an overall evaluation of the model for Europe is necessary
93 in order to assess confidence in attribution results derived from this model.

94 This article addresses three main questions: (i) are the simulations correctly representing the statistics
95 of events for the historical period 1960-2013? (ii) Are the simulations correctly representing long-term
96 changes in extreme events and dynamics along the reference period? (iii) Are the simulations correctly
97 representing the key processes driving to extreme events?

98 The first and last issues are covered in detail in this document. The second one is a more difficult
99 question to address with 54-year long simulations. Trends, especially in extremes, have regional
100 patterns of response to human activities that are fairly uncertain due to long-term atmospheric
101 variability. Hence a single-realization observation is not expected to agree completely with model
102 simulations.

103 A last issue concerning the model ensemble is also whether the overall ensemble also captures the
104 natural variability well. This will however not be considered here to keep focus on processes. This
105 question was addressed in a theoretical framework to show that the consistency of the ensemble
106 spread can be measured by the notion of reliability (Bellprat and Doblas-Reyes, 2016; Lott and Stott,
107 2016). Ensemble reliability measures whether the probability to exceed a threshold (e.g. an extreme
108 event or a large model quantile) agrees with the frequencies of the same threshold in an observed
109 record. Correct reliability is therefore a necessary condition for the ensemble probabilities used in
110 event attribution studies not to be biased. A bias in ensemble reliability systematically affects the
111 fraction of attributable risk (Bellprat and Doblas-Reyes, 2016).

112 We focus here on a few types of events and processes to give an overview of the performance of the
113 HadGEM3-A system in Europe. The evaluation does not pretend to be exhaustive, as event-specific
114 evaluation will always be necessary. The selected events types are: heat waves, cold spells, droughts,
115 heavy precipitation events, and wind events leading to storm surges. These generally have a daily to
116 seasonal time scale. They were selected because the underlying weather variables have long
117 observational records.

118 In Section 2, we briefly describe the simulations. A more detailed description is given in a separate
119 article (Ciavarella et al., 2017, in preparation). We also describe the data sets used. Section 3 is devoted
120 to an overall assessment of the main biases in mean state, variability and extremes, as well as a
121 comparison between simulated and observed trends. In Section 4, an extreme value analysis is carried
122 out in order to investigate distribution tails. Section 5 is dedicated to an analysis of a few key processes

123 driving the extreme events. A final conclusion and discussion follow (Section 6) where tentative
124 conclusions for attribution are given.

125 **2. Simulations and observations**

126 **2.1 The HadGEM3-A simulations**

127 The simulations used in this work were generated by the Hadley Centre event attribution system
128 (Christidis et al., 2013a) that has facilitated numerous studies of different types of high-impact extreme
129 events. A typical attribution study involves pairs of large ensemble experiments with and without
130 anthropogenic forcings, from which the changing likelihood of extreme events under climate change
131 can be determined (Stott et al., 2016). The Hadley Centre system is built on the HadGEM3-A model
132 that was recently upgraded within the EUropean CLimate Event Interpretation and Attribution
133 (EUCLEIA) project (<http://eucleia.eu/>) and now features one of the highest resolution global models
134 used in global event attribution research. The model runs at N216 horizontal resolution, equivalent to
135 about 60 km at mid-latitudes, and comprises 85 vertical levels. The upgraded model also benefits from
136 a new atmospheric science package with an improved dynamical core, which leads to better numerical
137 stability (Williams et al., 2015).

138 An ensemble of 15 atmospheric simulations of the historical climate during the period 1960-2013 was
139 produced with the new model and is the basis of the evaluation assessments discussed in this paper. A
140 second ensemble of model runs without the effect of anthropogenic forcings was also generated and
141 employed in attribution analyses (Christidis et al., 2016; Burke et al., 2016; Wilcox et al., 2017), but is
142 not used here. The historical forcings in the model simulations include anthropogenic greenhouse gas,
143 aerosols, tropospheric and stratospheric ozone emissions, changing land use, as well as natural changes
144 in the solar output and volcanic aerosols (Jones et al., 2011). Ensemble members are generated by
145 implementing random parameter perturbations as well as a stochastic kinetic energy backscatter
146 scheme that accounts for energy sources on sub-grid scales (Christidis et al., 2013a). Monthly
147 observations of the sea surface temperature (SST) and sea-ice from the Hadley Centre Sea Ice and Sea

148 Surface Temperature (HadISST) v1 dataset (Rayner et al., 2003) provide boundary conditions for the
149 simulations of the historical climate. Building on the multi-decadal simulations, an operational
150 attribution system is currently being developed by firstly extending the model runs and increasing the
151 ensemble size and then by continuing to extend the simulations on a seasonal timescale in a similar
152 fashion to seasonal forecasting systems.

153 **2.2 Observations**

154 In this paper we use a number of observational data sets for the model simulations evaluation. In
155 general, the NCEP/NCAR 20th Century reanalysis 20CR re-analyses of sea level pressure have been used
156 for characterizing atmospheric circulations. Surface temperatures and precipitation are either taken
157 from CRUTS3.2 or from E-OBS data sets. Specific data sets have been used to study the land-
158 atmosphere interactions, described in Table 1.

159 **3. Model climatology, trends and variability**

160 **3.1 Mean states and trends**

161 In this section we review the main statistics of the model climate and compare it to observations. The
162 mean state, time evolution, and interannual variability of metrics of mean and daily extremes in the
163 15-member HadGEM3-A N216 ensemble are compared to a variety of observational datasets (primarily
164 CRUTS3.23 and E-OBS) for the June to August (JJA) and December to February (DJF) seasonal means.
165 Spatial patterns have been considered over the European domain, and time series have been
166 considered over three regions: Europe (35–70°N, 10°W–40°E); Northern Europe (50–60°N, 10°W–
167 25°E); and Southern Europe (35–45°N, 10°W–25°E). Where regional means are considered, they only
168 include model grid cells with a land fraction over 75%, as observations are only available over land.

169 In general, HadGEM3-A represents the spatial pattern of mean near-surface temperature well, but does
170 not reproduce the regional pattern of the trends. In summer, the model underestimates warming in
171 southern Europe (in line with coupled models, van Oldenborgh et al, 2009; Kirtman and Power, 2013,

172 Box 11.2), and overestimates it in northern and eastern Europe (Figure 1a-b,e-f). Such trend
173 discrepancies are not due to atmospheric internal long-term variability as they are found in most
174 members (Figure 1f). However, when averaging over Europe, the model trend of $0.36\pm 0.05\text{K/decade}$ is
175 compatible with the linear trends from CRUTS3.24 ($0.33\pm 0.08\text{K/decade}$), E-OBS v14.0
176 ($0.35\pm 0.08\text{K/decade}$), and CRUTEM4.5 ($0.32\pm 0.08\text{K/decade}$). In the above numbers, the 95% confidence
177 interval is provided. The histogram of the rank of the observations in the ensemble also shows an
178 overall reliability of simulated temperatures at the continental scale (Figure 1i, see van Oldenborgh et
179 al., 2013 for the reliability rank histograms calculations). When averaged over Northern Europe, the
180 model slightly overestimates the positive trend in near-surface temperature ($0.36\pm 0.05\text{K/decade}$
181 compared to $0.28\text{-}0.31\text{K/decade}$ in observations), and underestimates the positive trend when
182 averaged over Southern Europe ($0.36\pm 0.06\text{K/decade}$ compared to $0.41\text{-}0.44\text{K/decade}$ in observations).
183 Similar findings are obtained for the trends in daily minimal and maximal temperatures (T_{min} and T_{max}),
184 yearly maximum of daily maxima and minima (TXx and TNx) (not shown). Least squares linear trends,
185 as calculated above, were taken for the period 1960-2013, and should be interpreted with a degree of
186 caution, due to the nonlinear nature of the time series evolution (see also Figure 3).

187 HadGEM3-A also represents the spatial pattern of mean summer precipitation, and trend patterns
188 match the observed dipole, with some discrepancies (Figure 1c-d,g-h), and a general underestimation
189 of precipitation trend in the ensemble members (Figure 1h). Positive precipitation trends over
190 Scandinavia and negative trends over France and Eastern Europe are found. However, the model fails
191 to capture the observed increase in precipitation over the UK and drying over Spain, and does not
192 simulate drying over the full longitudinal extent of the Alps, as is seen in observations. The imbalance
193 toward systematic trend underestimation is also shown in the rank histograms when considering the
194 whole continent (Fig. 1j). The simulated trend over southern Europe is $-0.023\pm 0.021\text{mm/day/decade}$,
195 while it is -0.042 and $-0.034\text{mm/day/decade}$ in E-OBS v14 and CRUTS3.23 respectively. In Northern
196 Europe, trends are found in observations (0.052 and $0.046\text{ mm/day/decade}$) however they are not
197 significant. HadGEM3-A also shows no significant trend here (see also van Haren et al., 2013).

198 In the winter season (Figure 2), mean states are again well simulated, but regional trend patterns are
199 not well reproduced either. Over Scandinavia, the pattern of the near-surface temperature mean state
200 is also well-represented by the model, but the model is too cold (Figure 2a-b). Observed temperature
201 trends show significant warming over most of Europe at the 5% or 10% level, with the greatest warming
202 over Scandinavia and the Baltics, but HadGEM3-A generally underestimates the magnitude and
203 significance of the trends (Fig. 2e-f). However these trends discrepancies can be due to long-term
204 atmospheric variability, as seen from Fig.2f and the rank histogram of Figure 2i, and no major
205 incompatibility with the observation is found. HadGEM3-A simulates the pattern of the mean states
206 and interannual variability in T_{min} , T_{max} , T_{NX} , and T_{Xx} well, but it does not reproduce the observed
207 trends (not shown).

208 The pattern of the wintertime precipitation mean state is strongly tied to orography in both the model
209 and observations. However, the model overestimates precipitation over the Pyrenees, Massif Central,
210 Alps, and Greece, and underestimates it over the UK and Ireland (Figure 2c-d). Observed trends in
211 precipitation have a strong dipole pattern, with drying in southern Europe, and increasing precipitation
212 in the north resembling trends associated with a tendency towards positive NAO (see Deser et al.,
213 2016). There is a hint of this pattern in the ensemble mean model trend, but the magnitude is much
214 weaker than observed (Figure 3), and the ensemble fails to capture the main contrasts (Figure 2h and
215 2j). Patterns in the mean state and interannual variability in extreme precipitation values are well
216 represented in HadGEM3-A. Trends in these quantities are noisy in both the model output and
217 observations (not shown).

218 **3.2 Variability**

219 In general the interannual variability is reasonably well simulated, as seen in Figure 3 from time series
220 of individual members and superimposed observations. The model overestimates variability in
221 seasonal mean daily mean and maximal temperatures (Figure 3), for European average, but simulates

222 the variability in daily minimal temperatures fairly well (not shown). The overestimation in daily
223 maxima is more marked in Northern Europe than in Southern Europe.

224 In winter, HadGEM3-A reproduces the inter-annual standard deviation of near-surface temperature
225 over Europe as a whole, but shows a larger standard deviation in Southern Europe (Figure 3), and
226 appears to underestimate it in Northern Europe. Interannual variability in T_{max} and T_{min} is well
227 represented by HadGEM3-A in Europe, despite underestimates in the north, as for near-surface
228 temperature. In southern Europe, the model overestimates variability in T_{max} (not shown), but
229 underestimates it in T_{min} (Fig. 3). Variability in TNx and TXx is underestimated in all regions (not shown;
230 see also Section 4). Variability in seasonal precipitation amount, as well as in heavy precipitations (over
231 10 mm or 20 mm per day) is well represented by HadGEM3-A in general in both seasons (not shown).
232 However, it should be kept in mind that the model resolution does not allow a proper representation
233 of convective precipitation events.

234 **3.3 In summary**

235 HadGEM3-A generally shows reasonable performance in reproducing the observed mean-state,
236 variability, and trends in daily means and extremes when considering Europe as a whole. However
237 observed regional patterns of trends are not always well reproduced. For instance, the model fails to
238 reproduce the observed JJA and DJF drying in southern Europe. In JJA, the model also locates the
239 maximum in near-surface temperature trends too far east, so that the amplitude of warming over
240 southern Europe is underestimated. In winter, temperature variability is high making trends from
241 simulations and observations almost compatible despite a general tendency for the model to
242 underestimate warming. The model ensemble fails to reproduce positive trends in temperature
243 extremes (T_{min} , T_{max} , TNx, and TXx) throughout Europe, and also underestimates interannual variability
244 in TNx and TXx in winter. The amplitude of the dipole in precipitation trends in DJF is substantially
245 underestimated by HadGEM3-A in DJF, and to a lesser extent in JJA.

246 The correct simulation of trends in summer implies that their attribution should not be hindered by
247 model's climatological biases in this season. For temperature this means a realistic mean response to
248 external forcing and a potential for attributing temperature-related events. The differences in regional
249 patterns of trends are partly due to the relatively short length of observational data sets combined with
250 a chaotic atmosphere and weak SST dependence. It is also probably due to uncertainties in underlying
251 processes (see Section 5). In winter the too weak warming trend may potentially lead to
252 underestimation of likelihood reduction in winter cold spells. However, this discrepancy may also result
253 from the large interannual and variability in winter temperatures. Some of the 15 members do show
254 trends as observed in daily mean winter temperatures.

255 **4. Extreme value analysis**

256 A specific focus is given now on extremes of temperature and precipitation. The evaluation of the
257 model's representation of extremes was undertaken using extreme value analysis, based on annual
258 maxima of the historical runs in precipitation (rx1day) and maximum (TXx) and minimum daily
259 temperature (TNn) discussed above. These were fitted to a stationary generalized extreme value (GEV)
260 distribution (Coles, 2001). The three parameters of the GEV distribution, namely the location
261 parameter μ (representing the mean values), scale parameter σ (representing the typical range of
262 values) and shape parameter ξ (describing whether the distribution is heavy tailed or not), were
263 evaluated alongside distributions fitted to the same extremes from E-OBS. Non-parametric
264 bootstrapping (1000 replications) was used to estimate the uncertainty margins. Comparisons are
265 made using the 0.5° regular grid E-OBS product, which represents the resolution closest to that of the
266 model.

267 For extreme maximum daily temperature (TXx), the location parameter is significantly under-estimated
268 in Northern Europe and over-estimated in much of Southern and Eastern Europe. As illustrated in
269 Figure 4, the model exhibits warm biases in hot events across Central, Eastern and, to a lesser extent,
270 Southern Europe, explaining the bias in the location parameter. The scale parameter is overestimated

271 somewhat across most of the continent, but underestimated in Britain, and the shape parameter is
272 overestimated somewhat over most of Northern Europe, indicating too heavy tail potentially related
273 to unrealistically high drying in summer in this model (see Section 5).

274 For extreme minimum daily temperature (TNn), regions of complex topography (including the Alps and
275 the western coastline of Scandinavia) are characterized by a clear under-estimation of the location
276 parameter. The cold bias to the south of the Alps is also apparent in the analysis of cold events in Figure
277 4, with similar spatial features evident in multiple ensemble members. The scale parameter is
278 reasonably well represented, but the shape parameter is much too large in Eastern Europe, where the
279 model simulates too extreme very cold events. By contrast, the shape parameter is too small in much
280 of Western Europe.

281 For extreme precipitation the broad coastal and topographical precipitation features are well-
282 reproduced by the model, but both the location and scale parameters are consistently larger than those
283 of observed extremes (Figure 5): the model generates too much rain in extremes with too much
284 variability. This is particularly the case in Mediterranean coastal regions and immediately south of the
285 Alps. This is the opposite of what one would intuitively expect: given the model's coarse resolution,
286 extremes in the simulated precipitation field should typically be smaller in magnitude than those
287 events occupying the same point of likelihood in the observed distribution.

288 **5. Process analysis**

289

290 **The ability of a model to simulate physical and dynamical processes leading to extremes is key for its**
291 **capacity to simulate their changes under human activities influence.** Extreme events generally result
292 from an ensemble of processes involving atmospheric dynamics, large-scale drivers, as well as regional
293 to local-scale processes which interact with one another. Here, we evaluate whether the model
294 captures the most important processes leading to extreme events. For the five types of events under
295 study (heat and cold waves, heavy precipitation events, drought and storm surges) we examine in

296 particular the role of large-scale circulation and a few key regional-to-local scale processes, such as
297 interaction with land surface.

298 In general, extreme weather events occur under specific types of weather patterns: heat waves,
299 droughts and cold spells relate to long persisting anticyclones sitting over a large area. In Europe, heavy
300 precipitation is associated either with summer convective episodes coming after a long warm period
301 with the arrival of frontal systems with cold air aloft destabilizing the troposphere, or in long-lasting
302 wintertime cyclonic episodes bringing in recurring storms. In each case typical atmospheric circulation
303 patterns are found. Then, extreme events also result from amplifying processes, which may dominate
304 in some cases, such as land-atmosphere interactions in particular in the case of heat waves and
305 droughts (Seneviratne et al. 2010), and also cold spells through the effect of snow cover (Orsolini et al.
306 2013). Stratosphere-troposphere interactions have also been shown to be important in the build-up of
307 cold spells (Baldwin and Dunkerton, 1999). Here we evaluate these processes in HadGEM3-A
308 simulations.

309 **5.1. Atmospheric weather patterns**

310
311

312 One way to evaluate whether the model correctly simulates the atmospheric circulation variability is
313 through the analysis of *weather regimes*. Weather regimes are usually defined as large typical clusters
314 of atmospheric flows that are observed. The concept of weather regimes is based on dynamical systems
315 theory analysis of atmospheric variability: certain phase-space areas may include slow-down of
316 trajectories, due to the vicinity of stationary solutions (Legras and Ghil, 1985), or quasi-stationary
317 solutions (Vautard and Legras 1988). Since then, a number of studies (e.g. Michelangeli et al., 1995;
318 Cassou et al. 2005) have characterized weather regimes using cluster analysis. Over the North-East
319 Atlantic and Europe, such an analysis usually finds four stable clusters from observations or reanalysis
320 of sea-level pressure or geopotential height.

321 Here, we compare clusters obtained by a *k*-means algorithm applied to the NCEP/NCAR reanalysis and
322 the HadGEM3-A simulations carried out over the same period (1960-2013). The same North-Atlantic
323 domain is used **both for model and observations** [-80°W-50°E, 22.5°N-70°N]. A separate analysis is
324 done for winter (DJF) and summer (JJA) seasons using sea-level pressure fields.

325 The centroids of the obtained clusters for the NCEP/NCAR reanalysis and the HadGEM3-A model 15
326 member ensemble, are shown for winter in Figure 6a-h and for summer in figure 6i-p. The HadGEM3-
327 A model weather regimes centroids are quite similar to the observed ones with slight shifts. For
328 instance, the “blocking” (BLO) regime is well represented in winter and summer but the “Atlantic ridge”
329 (AR) regime has differences that can be seen mainly in winter. However, this should not be a major
330 issue for European extremes of temperature and rain as this latter regime is generally not associated
331 with extremes. Cold spells are usually characterized in winter by either the negative North Atlantic
332 oscillation (NAO-) regime, as was the case for the winter of 2009-2010 (Cattiaux et al., 2010), or by the
333 BLO regime. **Mild** winters with persistent rainfalls over Western Europe are characterized by the “Zonal
334 flow” (ZO) regime as shown by Schaller et al. (2016).

335 Another important aspect for extremes is the frequency of occurrence of regimes. One expects that to
336 correctly simulate the statistics of extreme events, a model must simulate correct frequencies in the
337 weather regimes. **In order to compare similar clusters statistics, we used the NCEP/NCAR cluster**
338 **centroids SLP anomalies as reference and counted the number of SLP fields for which each centroid is**
339 **nearest, both for NCEP/NCAR and HadGEM3-A fields for a best comparison. HadGEM3-A weather**
340 **regime frequencies are well represented with respect to ones in NCEP/NCAR. BLO and NAO- regimes**
341 **are well represented** in both seasons while ZO (winter) and AL (summer) have slight differences to
342 NCEP/NCAR (lower and higher frequencies of occurrence respectively).

343 Table 2 shows the frequencies of nearest neighbors calculated in this way. It is quite remarkable how
344 well the frequencies match between observations and the model. We conclude that the HadGEM3-A

345 model simulates quite well the main weather patterns of the North East Atlantic with mean frequencies
346 that reproduce faithfully the observations.

347 **5.2. Atmospheric circulations associated with hot and cold events**

348
349 The previous analysis was made for weather patterns independently of extreme events. We now turn
350 to the evaluation of the capability of HadGEM3-A in representing the specific weather patterns
351 associated with hot and cold events in Central Europe (defined here as the average over 2°-15°E and
352 47°N-54°N). This analysis builds on Krueger et al. (2015) and is based on a composite analysis of
353 temperatures and circulation states (characterized by the geopotential heights at 500 hPa) for hot and
354 cold events. We show here results for hot extremes and cold events are shown in the supplement. The
355 temperature data was deseasonalized (using a 10-day filter for calculating the climatology); prior to
356 detecting hot and cold extremes the linear long-term trends over the analysis period were removed
357 from each gridpoint.

358 Hot and cold events with a time scale of five days were obtained as consecutive values above the 95th
359 and below the 5th daily temperature percentile for summer (JJA) and winter (DJF), respectively. These
360 moderate extremes should occur under broadly similar circulation conditions to stronger extremes, but
361 are well sampled (Krueger et al., 2015) and have been found useful (Alexander, 2016). Composites of
362 all such events were calculated for the 1960-2013 period which yields 143 heat waves and 137 cold
363 spells from the 20CR v2c in comparison to a range of 149-154 hot spells and 147-150 cold spells for the
364 model ensemble, respectively (note that the reanalysis shows slightly, but significantly, fewer hot and
365 cold spells). The associated circulation patterns are calculated as the composites of the 500 hPa
366 geopotential height found for each occurrence of a cold- or hot temperature event, following Krueger
367 et al. (2015). In contrast to Krueger et al. (2015), the composite analysis was performed for land-only
368 temperatures. The analysis for the model was performed for each of the 15 ensemble members
369 separately, with resulting composites then averaged to provide an ensemble mean value.

370 Figure 7 shows the temperature composites of hot events, and Figure 8 the circulation associated with
371 it. The differences between circulation composites are relatively large in both summer and winter even
372 though these are aggregated over events occurring over 54 years in each case. Larger differences
373 between the ensemble members are found for summer. For the circulation associated with extreme
374 hot events, there is high variability across the ensemble members while for the ensemble mean the
375 geopotential pattern resembles a classic omega blocking in 20CR with the eastern, negative center of
376 the blocking suppressed or moved in the average circulation of HadGEM3-A. The location of low
377 pressure anomalies and their magnitude varies across ensemble members for this 54 year average. The
378 spatial extent and intensity of heat waves varies across ensemble members consistent with the subtle
379 variations in circulation (for example, compare middle of the second to bottom left panel for figures 7
380 and 8). The observations lie within that large variability.

381 Results for cold events are similar (Supplementary Figure 1 and 2), with a strong pressure gradient
382 between a high and low in NW and SE Europe, respectively, causing cold spells, whose average intensity
383 and extent varies depending on the tilt of the pressure gradient, again exemplifying the important role
384 of atmospheric variability even on the long timescales averaged across here.

385 **5.3 Land-atmosphere interactions**

386

387 Land-atmosphere interactions are major processes in the development of many extremes and must
388 therefore be well represented in view of attribution studies. This is particularly important for heat
389 waves, which are expected to become more frequent with greenhouse gases increase (Seneviratne et
390 al. 2012), with potential severe impacts on society and economy (Rosenzweig et al. 2001, Corti et al.
391 2009, Blauhut et al. 2015, Zhao et al., 2016). The uncertainty of projections of future temperatures and
392 associated hot extremes is especially large in regions where a shift of the evapotranspiration regime is
393 expected, i.e. where evapotranspiration is radiation-limited in today's climate but will become soil-
394 moisture-limited in future climate. This is due to a large uncertainty in the representation of the land-
395 atmosphere coupling across state-of-the-art Earth System Models (ESMs) in present and future climate

396 (Seneviratne et al. 2016), and resulting fluxes (Stegehuis et al., 2013). This problem needs to be
397 addressed by validating and evaluating the involved modelled processes in present climate conditions
398 against observations. Thanks to recent advances in the development of reference datasets for land key
399 variables such as soil moisture (Orth and Seneviratne 2015) and evapotranspiration (Mueller et al.
400 2013), a comprehensive evaluation of the modelled land-atmosphere coupling became feasible.

401 We assess and evaluate the land-atmosphere coupling in the HadGEM3-A model in Europe by
402 considering all parts of the overall coupling separately (see Figure 9 of Seneviratne et al. 2010). In
403 particular we focus on (i) the coupling between soil moisture and evapotranspiration, (ii) the coupling
404 between evapotranspiration and temperature (extremes), and (iii) the (resulting) coupling between
405 precipitation and temperature (extremes). In terms of temperature we will focus on monthly mean
406 temperature, and to represent hot extremes we use TXx.

407 The relationship between the variables involved in each part modelled by HadGEM3-A is compared
408 with the corresponding observed interplay using state-of-the-art reference datasets of the
409 corresponding variables (Table 1). Here we focus on the time period 1960-2013, however, due to
410 limited availability of the reference datasets, the evaluation of evapotranspiration-related couplings is
411 constrained to 1989-2005, and the evaluation of soil moisture-related couplings is restricted to 1984-
412 2013. Note furthermore the different spatial resolutions between the employed reference datasets
413 (see Table 1), and of the HadGEM3-A output data. Model output has been masked whenever the
414 reference data was not available to ensure the same spatial and temporal basis of the analyses.

415 In order to focus on the highest coupling strengths, we perform all computations with monthly data
416 using only the hottest month of each year. In the case of soil moisture and precipitation we use the
417 previous month to capture their influence on subsequent temperature or evapotranspiration. For the
418 estimation of the considered coupling strengths we consider 3 European subregions, (i) Northern
419 Europe (NEU), (ii) Central Europe (CEU), and (iii) the Mediterranean (MED) as defined in Seneviratne et
420 al. (2012). For the Mediterranean region, however, we focus on latitudes between 35°N-45°N instead

421 of 30°N-45°N as in Seneviratne et al (2012) due to limited spatial availability of the reference datasets
422 (region hence denoted as MED*). Coupling strengths are expressed as monthly correlations.
423 Furthermore, we compare modelled versus reference distributions of the considered variables in the
424 considered months.

425 **Soil moisture - Evapotranspiration Coupling:** The HadGEM3-A coupling between preceding soil
426 moisture and evapotranspiration in the hottest month is compared with reference data in Figure 9.
427 Apart from the apparent bias in evapotranspiration in NEU and CEU, HadGEM3-A captures the
428 observed coupling well. Overall strength and the spatial pattern of the correlation between soil
429 moisture and evapotranspiration are also well represented. With few exceptions the HadGEM3-A
430 ensemble captures the observed coupling strength in all European regions. Only over the Iberian
431 Peninsula (underestimation) and in Ireland (no coupling) the model results do not agree with the
432 correlations across the reference datasets. Note the large spread of correlations between the individual
433 ensemble members suggesting strong variability of the modelled coupling.

434 **Evapotranspiration - Temperature Coupling:** The HadGEM3-A coupling between evapotranspiration
435 and temperature in the hottest month is compared with reference data in Figures 10 and 11. While the
436 overall strength and the north-south gradient in the correlation are represented in the model, its
437 simulated spatial coupling pattern agrees only partially with the reference datasets. The transition zone
438 with zero coupling strength between the positive coupling in NEU and the negative coupling in MED*
439 is too wide in the model, and it is shifted northward as compared to the reference datasets. This
440 contributes to the overestimation of hot temperature extremes by the HadGEM3-A model found in
441 Section 4. The underestimation of the evapotranspiration-temperature coupling between 50°N-65N°
442 also explains why the observed correlation is not contained in the HadGEM3-A ensemble in large parts
443 of this region. This occurs even though the spread of correlations between the ensemble members of
444 HadGEM3-A is large, as for the previous coupling (Figure 9 of Seneviratne et al., 2010). Results also
445 show a Northward extension of coupling region, potentially creating too warm hot periods, in

446 agreement with the extreme value analysis of Section 4. We find comparable results for mean and
447 extreme temperatures indicating almost no change of this coupling in heat waves.

448 We find a large spread of coupling strengths between the ensemble member simulations (not shown)
449 indicating large variability of the coupling. It remains unclear if this is a model-specific feature. This
450 could be tested by comparing the temporal variability of the coupling strength in the reference data
451 and in the model output using temporal subsets of the available data. However, this is beyond the
452 scope of this article. We note, however, that this variability could help to explain the offset in the spatial
453 patterns of coupling strengths between the reference datasets and the model.

454 **Spring preconditioning of heat waves:** We next investigate to what extent spring preconditioning of
455 soil matters for individual heat wave metrics (see e.g. Vautard et al., 2007; Hirschi et al., 2011). A metric
456 of European heat waves that targets impacts is used, based on maximum and minimum temperatures
457 exceeding the 90th percentile threshold for at least 3 days and 2 nights (Pezza et al., 2012; Cowan et al.
458 2017). This approach should be considered analogous to approaches using Excess Heat Factor (e.g.
459 Perkins et al., 2012) or hottest daily maximum temperature of the year (Hauser et al., 2016) and hence
460 relates to the index considered above. We tested the sensitivity of summer heat waves to preceding
461 wet and dry springs for different European sub-regions in E-OBS v14.0, and whether the HadGEM3-A
462 can capture this sensitivity. Heat wave composites were calculated over summers following the top
463 20% driest and wettest springs (i.e. for E-OBS this is equivalent to the 14 driest and 14 wettest springs,
464 for HadGEM3-A this corresponds to the 11 driest and 11 wettest springs per ensemble member) based
465 on 3-month Standardized Precipitation Index (SPI; McKee et al., 1993) for spring ending in May,
466 averaged over each region of interest (bounded regions in Figure 12). A non-parametric two-sample
467 Wilcoxon sign-ranked test (Hollander and Wolfe, 1999) was used in order to determine if the summer
468 heat wave metrics are distinguishably different between dry and wet spring cases at the 95%
469 confidence level (e.g. Cowan et al., 2017).

470 Figure 12 shows the composite patterns following the wet and dry springs for the heat wave duration
471 (HWD), which describes the longest seasonal heat wave. The model composites are based on 11 springs
472 each from the 15 historical ensemble members (165 springs in total). The patterns from E-OBS show
473 that dry springs across southern Europe are systemically followed by longer summer heat waves
474 compared to wet springs (Figure 12, left panels), with many Mediterranean regions exceeding 5.5 days
475 on average. This is consistent with the results for coupling strength shown above (Figure 10). Further
476 north into central and eastern continental Europe, this observed tendency becomes much weaker and
477 less significant. Across southern England and northern France, despite the lack of significance there is
478 a small increase in HWD following dry springs compared to wet springs. For western Scandinavia, longer
479 summer heat waves tend to emerge following wetter springs, consistent with a positive
480 evapotranspiration temperature coupling, which suggests that antecedent soil moisture conditions,
481 based on the SPI, are not a significant predictor of summer-time heat wave activity.

482 In general, HadGEM3-A shows a smaller effect of dry springs on HWDs across the western
483 Mediterranean, however, it captures the significant differences compared to the wet spring composites
484 (Figure 12; right panels). The model also appears to overestimate the dry-spring HWDs over the far
485 eastern Mediterranean including Romania. Further north, the model simulates a much weaker spring
486 SPI - summer HWD relationship, with strong positive biases over most of Scandinavia (compared to E-
487 OBS). Across southeastern England and northern France, the model suggests that spring drying has
488 significant control over heat wave activity (also seen in the simulated heat wave amplitude; not shown);
489 this signal is more pronounced in the model if upper layer soil moisture is used instead of the SPI (not
490 shown). Despite model biases, the patterns across southern Europe imply that dry springs and winters
491 do exert a strong influence on summer heat wave activity, confirming earlier observational studies
492 (Quesada et al., 2012, Vautard et al., 2007) and consistent with results for coupling strength shown
493 above (Figure 10). For central Europe, Scandinavia and the Baltic states, there is only a weak association
494 to spring conditions in both model and observations, although the model captures the strong spring
495 pre-conditioning across Eastern Europe. This is in agreement with a northward shift of the negative

496 coupling region as found above (Figure 10), and as such, the model appears to have a stronger response
497 to dry spring anomalies in Eastern Europe compared with observations. The results for both E-OBS and
498 HadGEM3-A are affected by sampling uncertainty, particularly for observations, and the fact that the
499 SPI is averaged over large domains with many different climates; thus care must be taken in interpreting
500 the spring-summer coupling. Furthermore, the SPI may not fully represent variations in the simulated
501 upper soil moisture over northern latitudes (e.g. Scandinavia), given low correlations (~ 0.1) in the
502 model, compared to 0.93 over western Mediterranean.

503 **5.4 Stratosphere-troposphere interactions**

504

505 A key process in cold spells development is the interaction between stratosphere and troposphere,
506 which must also be well represented in view of cold spell events attribution. In the extra-tropical NH
507 winter there is a tendency for anomalies to propagate from the stratosphere to the troposphere where
508 they disturb the NAO and the weather related to this dominating mode of variability. In particular, weak
509 stratospheric vortex events are followed by an increased probability of cold temperatures and cold
510 extremes in Europe. Although this coupling between the stratosphere and the troposphere on intra-
511 seasonal time-scales has been known for more than a decade (Baldwin and Dunkerton, 1999;
512 Christiansen, 2001) there still remain unanswered questions about how to represent the stratospheric
513 variability in order to optimally catch the coupling. Here, we evaluate the HadGEM3-A model's ability
514 to reproduce the observed connection between the stratosphere and the troposphere.

515 The downward propagation from the stratosphere to the troposphere can be demonstrated by lagged
516 correlations between zonal mean wind at 60°N , 10 hPa (a measure of the stratospheric vortex) and the
517 zonal mean wind at other vertical levels at 60°N .

518 Unfortunately, only monthly averaged stratospheric model data have been saved in the model
519 experiments while daily should be used. To partly overcome this we have interpolated the monthly
520 averages to daily values. To evaluate the soundness of this approach we compare them with
521 observations sub-sampled to monthly values and then interpolated back to daily values. In daily

522 observations the downward propagation is clearly seen with maximum correlations at the surface
523 lagging those in the stratosphere with about 2 weeks. In the model there is a similar connection
524 between the stratosphere and the troposphere but it appears less lagged. This is at least partly due to
525 the smoothing effect of dealing with monthly data (as seen in the top left panel of Fig. 13).

526 The fact that the stratosphere is leading the troposphere also in the model is more clearly seen in Fig.
527 13 which shows the correlations between the stratospheric vortex (zonal mean wind at 60°N 10 hPa)
528 and the NAO. The effect of a weak NAO on European temperatures are well known (Hurrell et al., 2003),
529 thus Figure 14d shows the model skill to simulate a key connection between stratosphere and the
530 circulation pattern present during cold spells occurrences.

531 Finally, Figure 14 shows the correlation between the anomaly of the stratospheric vortex, defined as
532 above, and surface temperatures, for observations and five model ensemble members. For the
533 observations we find a pattern that is consistent with the impact of the NAO: positive correlations in
534 the middle and Northern Europe and negative correlations in Southern Europe (although these
535 correlations are not statistically significant). For the model we find that the ensemble members agree
536 on the general pattern, as revealed by ensemble mean correlations although there are considerable
537 differences between ensemble members.

538 **5.5 Processes involved in storm surges**

539

540 Storm surges can occur in numerous places in Europe and driving processes are essentially the
541 interaction between winds, low pressure systems, seas dynamics, and waves. It would be a tremendous
542 task, well beyond the scope of this article, to assess the model's capacity to simulate weather
543 conditions conducive to storm surges everywhere in Europe. Only a case study is developed here as an
544 example, in the Baltic sea, a region that is known for witnessing severe surges due to the geometry of
545 the sea and weather conditions. These occur in particular when strong winds develop after the passage
546 of cyclones over the Baltic Sea, potentially inducing extreme variations in sea level resulting in storm

547 surges e.g. along the German Coast in the southwestern Baltic Sea region (Sztobryn et al., 2005;
548 Hünicke et al., 2015). During strong onshore winds, the sea level rises due to wind set-up.

549 In this section we assess the ability of using HadGEM3-A as atmospheric forcing data to drive the
550 regional ocean model TRIM-NP (Kapitza, 2008) for calculating water level of the Baltic Sea in 12.8km
551 spatial resolution. Results are summarized here as a parallel study assesses the attribution of these
552 events to climate change (Klehmet et al., 2017, in preparation). Dynamical downscaling of HadGEM3-
553 A data has been done with 7 ensemble members only (due to computing costs) for 1971-2010. To
554 obtain a gridded reference data for the evaluation, one model reconstruction of water level of TRIM-
555 NP has been performed using the CoastDat2 data (Geyer, 2014) as atmospheric forcing. CoastDat2 is a
556 regional atmospheric hindcast simulation for the European continent for 1948 to 2012 obtained with
557 the regional climate model COSMO-CLM (Rockel et al., 2008) using the global reanalysis data of NCEP-
558 R1 (Kalnay et al., 1996; Kistler et al., 2001) as forcing data. We then first directly compare the outputs
559 of HadGEM3A-TRIM-data with those of CoastDat-TRIM.

560 Maximal November water level anomalies for selected grid boxes representing locations co-located to
561 cities along the German coast (here : Warnemünde, Travemünde) for 1971-2010 relative to the 1971-
562 2010 mean of the HadGEM3A-TRIM-data underestimate extreme water levels as compared with
563 CoastDat-TRIM (Figure 15). The high water levels of both storm surge events in 1995 and 2006 shown
564 by CoastDat-TRIM, used as reference data, are not found in the time series of historical HadGEM3A-
565 TRIM ensemble simulations that represent actual climate with anthropogenic forcing.

566 However, extreme winds in the area are properly reproduced by the model. We compared the
567 simulated distribution of three simple wind indicators with ERA-Interim surface winds: the wind speed
568 itself, the wind speed conditional on winds in the North-East Quadrant, and the North-Easterly
569 component of the daily wind field calculated as $NEW = -U - V$, U and V being respectively the zonal and
570 meridional wind components. All indices were averaged over the area (10°E-18°E; 54°N-56°N), which
571 encompasses the South-Western Baltic Sea. Distributions are fairly well represented as shown in Figure

572 16, despite a minor wind underestimation by HadGEM3-A relative to ERA-Interim. This
573 underestimation is quite homogeneous irrespective of the wind speed, and reaches about 6% in the
574 extreme values, but cannot explain the too low water levels of storm surges in Figure 15. However,
575 ERA-Interim winds may themselves have biases and one should be prudent in the interpretation of
576 these results. Comparisons with winds over sea remains difficult as observation data are largely
577 missing. Therefore, at least for this Baltic Sea, we could not find any major HadGEM3-A simulation bias
578 hindering the attribution of storm surges.

579 6. Conclusion and discussion

580

581 In view of attribution of change of likelihood of extreme events to human activity, we have presented
582 a number of comparisons between an ensemble of 15 atmospheric simulations from the HadGEM3-A
583 model and various observations over Europe. We have presented an analysis of model mean and
584 extreme statistics, and an assessment of its capacity to simulate key processes involved in a few
585 extreme events development. Results presented here show that HadGEM3-A simulates the
586 atmospheric mean, variability and extremes in Europe fairly realistically. As for any climate model,
587 some biases are found but (i) the major regional patterns of the climatology of the main variables is
588 well simulated and (ii) dynamical weather patterns are faithfully simulated by the model. This provides
589 confidence in use for attribution. Concerning extreme values, too strong heat extremes and heavy
590 precipitation are found, but the parameters of distributions do not exhibit qualitatively different
591 behavior than in observations. However, simulations do not well capture the observed patterns and
592 amplitudes of trends in temperature and precipitation, which is partly due to a trend in circulation that
593 is different from the observed one and from other climate models. While for temperature our trend
594 analysis shows that these discrepancies can be due to internal atmospheric variability (especially in
595 winter), precipitation trends have slight, but systematic, biases across the ensemble, which remain
596 unexplained.

597 We then have examined some key atmospheric processes but found no major deficiencies. The
598 variability of circulation types is well simulated, both in terms of spatial patterns and occurrence
599 frequencies. Physical processes behind these statistics consistently demonstrate the ability of the
600 model to simulate extreme events. Here are the main consequences that could be drawn for each of
601 the five types of extremes that we considered in this study.

602 **Heat Waves**

603 Simulated weather patterns associated with hot events compare favorably with those shown in the
604 reanalysis, however, with significant internal variability in the representation of events between model
605 ensemble members. However, heat build-up is also amplified by land-atmosphere feedbacks. We
606 found that HadGEM3-A captures land-atmosphere interactions in present-day climate reasonably well.
607 We assessed the different parts of this coupling and find that especially the soil moisture-
608 evapotranspiration coupling is well represented, while the evapotranspiration-temperature coupling is
609 underestimated in regions between 50°N-65N°. The overall coupling is investigated by correlating
610 preceding precipitation with temperature in the hottest month where the correlations of the model
611 output and between the reference datasets are similar, but the spatial patterns are not entirely
612 captured. Consistently, observed heat wave metrics following wet springs are significantly different
613 from those following dry springs, particularly in Southern Europe and this process is reasonably well
614 captured in the model.

615 Too strong drying is taking place in the model with exaggerated evapotranspiration, in central and
616 northern Europe in the hottest month, a probable reason for too many and strong heat waves.
617 Simulated summer temperatures actually exhibit a too large interannual variability in these regions.
618 Whether all these phenomena are linked remains to be confirmed with further analyses, however, our
619 results are suggestive of a bias towards a too fast transition towards a soil-moisture limited regime in
620 Central/ Northern areas as found in many other models (Fischer et al., 2012, Bellprat et al., 2014). This
621 may explain the biases found in the shape and location parameters for hot extremes. The role of spring

622 preconditioning on heat wave metrics appears reasonably simulated, although findings are consistent
623 with the biases discussed above.

624 The consequences for attribution of these results remain difficult to evaluate. The overestimated
625 interannual variability, together with evapotranspiration overestimation in large parts of Europe
626 suggests that heat waves responses to atmospheric composition changes may be too large. However,
627 observed trends in summer temperatures themselves do not show evidence of such oversensitivity.
628 This indicates that biases may not have a major influence on the skill of the model to simulate the
629 overall change in odds of heat waves **or that some of the model errors compensate for current climate.**

630 **Droughts**

631 Droughts have not been investigated in detail. However, several remarks can be made. The above
632 results for temperatures and evapotranspiration should in principle translate in the model simulating
633 too strong summer droughts. In addition, interannual precipitation variability appears to be slightly
634 overestimated in summer, potentially leading to both drier and wetter summers. However a deeper
635 investigation is required to better understand biases of the model and whether these biases are
636 hindering attribution of drought in Europe. It should be noted that climate models have large
637 differences in trends in droughts in Central Europe.

638 **Cold spells**

639 The circulation associated with cold events in Central Europe is well captured by the model and
640 individual model ensemble members again show long-term variations in the extent and intensity of
641 average cold spells linked to atmospheric internal variability. Extreme value analysis of extreme cold
642 winter temperatures show a fairly good agreement between simulated and observed values. However,
643 the simulations are not free of biases in the frequency of cold spells. Weather regimes such as blocking
644 or negative NAO, which usually drive cold spells in Europe, are well simulated, although their trend
645 does not necessary match that in the model. Interactions and lagged correlations between the
646 stratospheric vortex and tropospheric NAO and European temperatures are similar in model and

647 observations. Therefore, we did not find any major process hindering the representation of cold spells.
648 However, the trends in circulation and temperature are not well-simulated. Due to high natural
649 variability it cannot be assessed how this translates to trends in cold extremes.

650 **Extreme precipitation**

651 Extreme daily precipitation are in large parts of Europe due to convective phenomena and thus local
652 by nature. Global climate models usually have difficulties in simulating such phenomena given their
653 coarse spatial resolution. HadGEM3-A has a wet bias in these extremes, associated with a too-large
654 variability, especially around the Mediterranean sea. **In this area we expect daily precipitation patterns**
655 **to have a smaller scale than the model resolution calling for prudence in attribution interpretation from**
656 **this model. However, it is noteworthy that** the spatial pattern of extreme precipitation distributions is
657 quite similar to observed. **Also, despite the biases, the simulations exhibit GEV parameters that are**
658 **quite consistent with observations, which could make the simulations** eligible for attribution of
659 precipitation extremes once the bias has been corrected.

660 **Storm surges**

661 Results for storm surges indicate an underestimation of the events amplitude when a regional ocean
662 model is driven by HadGEM3-A as compared to a regional atmospheric hindcast obtained by
663 downscaling the NCEP-R1 reanalysis. Comparisons of simulated winds with ERA-Interim reanalysis
664 show a good performance of the model for strong winds or strong North-Easterlies in the South-
665 Western Baltic Sea region where storm surges occur in Northern Germany, indicating that winds in the
666 investigated domain are actually not the main factor of underestimation. Thus HadGEM3-A model
667 simulations can a priori be used for storm surge attribution.

668

669

670 **Acknowledgements**

671

672 This study was part of the European CLimate and weather Events: Interpretation and Attribution
673 (EUCLEIA) FP7 SPACE project, Grant Agreement n° 607085, and concerned principally its Work Package
674 6 (Evaluation and diagnostics). NC and AC, FL and PS were also supported by the Joint BEIS/Defra Met
675 Office Hadley Centre Climate Programme (GA01101).

676 **References**

677

678 Alexander L. (2016) Global observed long-term changes in temperature and precipitation extremes: A
679 review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes*
680 Volume 11, March 2016, Pages 4–16.

681 Angéilil, O., Perkins-Kirkpatrick, S., Alexander, L. V., *et al.* Comparing regional precipitation and
682 temperature extremes in climate model and reanalysis products. *Weather and Climate Extremes*, 2016,
683 vol. 13, p. 35-43.

684 *Attribution of Extreme Weather Events in the Context of Climate Change* (National Academies, 2016).
685 [https://www.nap.edu/catalog/21852/attribution-of-extreme-weather-events-in-the-context-of-](https://www.nap.edu/catalog/21852/attribution-of-extreme-weather-events-in-the-context-of-climate-change)
686 [climate-change](https://www.nap.edu/catalog/21852/attribution-of-extreme-weather-events-in-the-context-of-climate-change)

687 Baldwin, M. P., & Dunkerton, T. J. (1999) Propagation of the Arctic Oscillation from the stratosphere to
688 the troposphere. *Journal of Geophysical Research: Atmospheres*, 104(D24), 30937-30946.

689 Bellprat, O., Kotlarski, S., Lüthi, D., & Schär, C. (2014) Physical constraints for temperature biases in
690 climate models. *Geophysical Research Letters*, 40(15), 4042-4047.

691 Bellprat, O., and F. Doblas-Reyes, (2016) Attribution of extreme weather and climate events
692 overestimated by unreliable climate simulations. *Geophys. Res. Lett.*, doi: 10.1002/2015GL067189.

693 Bindoff NL, Stott PA, AchutaRao KM, Allen MR, Gillett N, Gutzler D, Hansingo K, Hegerl G, et al. (2013).
694 *Chapter 10 - Detection and attribution of climate change: From global to regional*. In: *Climate Change*
695 *2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5*. Cambridge: Cambridge
696 University Press.

697 Blauhut, V., Gudmundsson L. & Stahl, K. (2015) Towards pan-European drought risk maps: quantifying
698 the link between drought indices and reported drought impacts. *Environ. Res. Lett.* 10, 014008.

699 Burke, C., P.A. Stott, Y. Sun, A. Ciavarella (2016) Wettest May in South-Eastern China for 40 years, In
700 "Explaining Extremes of 2015 from a Climate Perspective", *Bull. Amer. Meteor. Soc. Sup.*

701 Cassou, C., Terray, L. & Phillips, A. S. (2015) Tropical Atlantic influence on European heat waves. *J. Clim.*
702 18, 2805_2811.

703 Cattiaux, J., R. Vautard,, C. Cassou, P. Yiou, V. Masson-Delmotte and F. Codron (2010) Winter 2010 in
704 Europe: a cold extreme in a warming climate. *Geophys. Res. Lett.*, 37, L20704,
705 doi:10.1029/2010GL044613.

706 Christiansen, B. (2001) Downward propagation of zonal mean zonal wind anomalies from the
707 stratosphere to the troposphere: Model and reanalysis. *Journal of Geophysical Research: Atmospheres*,
708 106(D21), 27307-27322.

709 Christidis, N., P.A. Stott, A. Scaife, A. Arribas, G.S. Jones, D. Copey, J.R. Knight, W.J. Tennant (2013a) A
710 new HadGEM3-A based system for attribution of weather and climate-related extreme events, *J.*
711 *Climate*, 26, 2756-2783

712 Christidis, N., P.A. Stott, D.J. Karoly, A. Ciavarella (2013b) An attribution study of the heavy rainfall over
713 eastern Australia in March 2012, In "Explaining Extremes of 2012 from a Climate Perspective", *Bull.*
714 *Amer. Meteor. Soc. Supp.*

715 Christidis, N., P.A. Stott, A. Ciavarella (2014) The effect of anthropogenic climate change on the cold
716 spring of 2013 in the UK. In "Explaining Extremes of 2013 from a Climate Perspective". *Bull. Amer.*
717 *Meteor. Soc.*, 95(9), S79–S82

718 Christidis, N., M. McCarthy, A. Ciavarella, P.A. Stott (2016) Human contribution to the record sunshine
719 of 2014/15 in the United Kingdom, In "Explaining Extremes of 2015 from a Climate Perspective", *Bull.*
720 *Amer. Meteor. Soc.*, supp.

721 Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001) *An introduction to statistical modeling of extreme*
722 *values* (Vol. 208). London: Springer.

723 Corti, T., Muccione, V., Köllner-Heck, P., Bresch, D. & Seneviratne, S. I. (2009) Simulating past droughts
724 and associated building damages in France. *Hydrol. Earth Syst. Sci.* 13, 1739–1747.

725 Cowan, T., A. Purich, S. Perkins, A. Pezza, G. Boschat, and K. Sadler (2014) More Frequent, Longer, and
726 Hotter Heat Waves for Australia in the Twenty-First Century. *Journal of Climate*, 27, 5851–5871, doi:
727 10.1175/JCLI-D-14-00092.1.

728 Cowan, T., G. Hegerl, I. Colfescu, A. Purich and G. Boshcat (2017) Factors contributing to record-break-
729 ing heat waves over the Great Plains during the 1930s Dust Bowl, *Journal of Climate*, doi: 10.1175/JCLI-
730 D-16-0436.1 (in press).

731
732 Deser, C., J. W. Hurrell and A. S. Phillips (2016) The Role of the North Atlantic Oscillation in European
733 Climate Projections. *Clim. Dyn.*, doi: 10.1007/s00382-016-3502-z

734
735 Eden, J.M., Bellprat, O., Kew, S., Lenderink, G., Manola, I., Omrani, H. and Oldenborgh, G.J. van. (2017)
736 Extreme precipitation in the Netherlands: an event attribution case study, *Clim. Dynam.* (submitted).

737
738 Eden, J.M., Wolter, K., Otto, F.E.L. and Oldenborgh, G.J. van. (2016) Multi-method attribution analysis
739 of extreme precipitation in Boulder, Colorado, *Env. Res. Lett.*, 11, 124009. DOI:10.1088/1748-
740 9326/11/12/124009.

741
742 Fischer, E. M., Rajczak, J., & Schär, C. (2012) Changes in European summer temperature variability
743 revisited. *Geophysical Research Letters*, 39(19).

744
745 Geyer, B. (2014) High-resolution atmospheric reconstruction for Europe 1948–2012: coastDat2, *Earth*
746 *Syst. Sci. Data*, 6, 147-164

747 van Haren, R., van Oldenborgh, G. J., Lenderink, G., Collins, M., & Hazeleger, W. (2013) SST and
748 circulation trend biases cause an underestimation of European precipitation trends. *Climate dynamics*,
749 40(1-2), 1-20.

750 Hauser, M., R. Orth, and S. I. Seneviratne (2016) Role of soil moisture versus recent climate change for
751 the 2010 heat wave in Russia, *Geophysical Research Letters*, 43, 2819–2826,
752 doi:10.1002/2016GL068036.

753 Hauser, M., L. Gudmundsson, R. Orth, A. Jézéquel, K. Haustein, R. Vautard, G. J. van Oldenborgh and
754 S. I. Seneviratne, 2017. Methods and model dependency of extreme event attribution : the 2015
755 European drought. *Earth's Future*, submitted.

756

757 Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New. (2008) A European daily high-
758 resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res.* 113, D20119.

759 Hegerl, G., & Zwiers, F. (2011) Use of models in detection and attribution of climate change. *Wiley*
760 *Interdisciplinary Reviews: Climate Change*, 2(4), 570-591.

761 Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H.,
762 Orłowsky, B., Stepanek, P. (2011) Observational evidence for soil-moisture impact on hot extremes in
763 southeastern Europe. *Nature Geoscience*, 4(1), 17-21.

764 Hollander, M., and D. A. Wolfe (1999) *Nonparametric Statistical Methods*. John Wiley and Sons, 787 pp.

765 Hünicke B., Zorita E. et al., 2015: The BACC II Author Team, Second Assessment of Climate Change for
766 the Baltic Sea Basin, Regional Climate Studies, DOI 10.1007/978-3-319-16006-1_9.

767 Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M. (2003) *An overview of the North Atlantic oscillation*
768 (pp. 1-35). American Geophysical Union.

769 Jones, C.D., J.K. Hughes, N. Bellouin, S.C. Hardiman, G.S Jones, J. Knight, S. Liddicoat, F.M. O'Connor,
770 R.J. Andres, C. Bell, K.-O. Boo, A. Bozzo, N. Butchart, P. Cadule, K.D. Corbin, M. Doutriaux-Boucher, P.
771 Friedlingstein, J. Gornall, L.Gray, P.R. Halloran, G.Hurt, W.J. Ingram, J.-F. Lamarque, R.M. Law, M.
772 Meinshausen, S. Osprey, E.J. Palin, L. Parsons Chin, T. Raddatz, M.G. Sanderson, A.A. Sellar, A. Schurer,
773 P. Valdes, N. Wood, S. Woodward, M. Yoshioka, M.Zerroukat (2011) The HadGEM2-ES implementation
774 of CMIP5 centennial simulations, *Geosci. Model Dev.*, 4, 543–570

775 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G.,
776 Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang,
777 J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis
778 project, *B. Am. Meteorol. Soc.*, 77, 437–471.

779 Kapitza, H. (2008) Mops -a morphodynamical prediction system on cluster computers. In: High
780 performance computing for computational science - VECPAR 2008, J. M. Laginha, M. Palma, P.R.
781 Amestoy, M. Dayde, M. Mattoso, J. Lopez (Eds.), pp. 63-68. *Lecture Notes in Computer Science*, Springer
782 Verlag.

783 Kirtman, B., S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, A.M. Fiore,
784 M. Kimoto, G.A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J. van Oldenborgh, G. Vecchi and H.J.
785 Wang (2013) Near-term Climate Change: Projections and Predictability. In: *Climate Change 2013: The*
786 *Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the*
787 *Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J.
788 Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge,
789 United Kingdom and New York, NY, USA, pp. 953–1028, doi:10.1017/CBO9781107415324.023.

790 Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu,
791 M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M. (2001) The NCEP-NCAR 50-year reanalysis:
792 Monthly means CD-ROM and documentation, *B. Am. Meteorol. Soc.*, 82, 247–267.

793 Krueger, O., Hegerl, G. C., & Tett, S. F. (2015) Evaluation of mechanisms of hot and cold days in climate
794 models over Central Europe. *Environmental Research Letters*, 10(1), 014002.

795 Legras, B., and M. Ghil, (1985) Persistent anomalies, blocking and variations in atmospheric
796 predictability, *J. Atmos. Sci.*, 42, 433-471.

797 Lott, F., N. Christidis, P.A. Stott (2013) Can the 2011 East African drought be attributed to human-
798 induced climate change? *Geophys. Res. Lett.*, 40, 1177-1181

799 Lott, F. C., & Stott, P. A. (2016). Evaluating Simulated Fraction of Attributable Risk Using Climate
800 Observations. *Journal of Climate*, 29(12), 4565-4575.

801 McKee, T. B., N. J. Doesken, and J. Kleist (1993) The relationship of drought frequency and duration to
802 time scales. In Proceedings of the 8th Conference on Applied Climatology, American Meteorological
803 Society, Boston, MA, Vol. 17, 179–183.

804 Michelangeli, P.A., Vautard, R., Legras, B., (1995) Weather regimes: recurrence and quasi-stationarity,
805 *J. Atmos. Sci.*, 52, 1237-1256.

806 Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig,
807 F., Maignan, F., Miralles, D., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K. C., Wood, E. F., Y.
808 Zhang, Y. & Seneviratne, S. I. (2013) Benchmark products for land evapotranspiration: LandFlux-EVAL
809 multi-dataset synthesis. *Hydrol. Earth Syst. Sci.* 17, 3707-3720.

810 van Oldenborgh, G. J., Drijfhout, S., Ulden, A. V., Haarsma, R., Sterl, A., Severijns, C., W. Hazeleger &
811 Dijkstra, H. (2009) Western Europe is warming much faster than expected. *Climate of the Past*, 5(1), 1-
812 12.

813 van Oldenborgh, G. J., Reyes, F. D., Drijfhout, S. S., & Hawkins, E. (2013). Reliability of regional climate
814 model trends. *Environmental Research Letters*, 8(1), 014055.

815 Orsolini, Y. J., R. Senan, G. Balsamo, F. J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, and R. E.
816 Benestad, (2013) Impact of snow initialization on sub-seasonal forecasts. *Clim. Dyn.*, 41, 1969–1982,
817 doi:10.1007/s00382-013-1782-0.

818 Orth, R. & Seneviratne, S. I. (2015) Introduction of a simple-model-based land surface dataset for
819 Europe. *Env. Res. Lett.* 10, 044,012.

820 Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G. J., Lohmann, D., Allen, M. R. (2011)
821 Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*,
822 470(7334), 382-385.

823 Perkins, S. E., L. V. Alexander, and J. R. Nairn (2012) Increasing frequency, intensity and duration of
824 observed global heatwaves and warm spells. *Geophysical Research Letters*, 39, L20714, doi:
825 10.1029/2012GL053361.

826

827 Pezza, A. B., P. van Rensch, and W. Cai (2012) Severe heat waves in Southern Australia: synoptic
828 climatology and large scale connections. *Climate Dynamics*, 38, 209–224, doi:10.1007/s00382-011-
829 1016-2.

830 Philip, S., S. F. Kew, G. J. van Oldenborgh, E. Aalbers, R. Vautard, F. Otto, K. Haustein, F. Habets, R. Singh
831 and H. Cullen (2017) Validation of a rapid attribution of the May/June 2016 flood-inducing precipitation
832 in France to climate change, *Climate Dynamics*, submitted.

833 Quesada, B., Vautard, R., Yiou, P., Hirschi, M. & Seneviratne, S. I. (2012) Asymmetric European summer
834 heat predictability from wet and dry southern winters and springs. *Nature Clim. Change* 2, 736–741.
835

836 Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, A. Kaplan
837 (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the
838 late nineteenth century, *J. Geophys. Res.*, 108, doi:10.1029/2002JD002670

839 Reichstein, M. et al. (2013) Climate extremes and the carbon cycle. *Nature* 500, 287–295.

840 Rockel, B., A. Will, und A. Hense (2008) The Regional Climate Model COSMO-CLM (CCLM), Editorial,
841 *Meteorol. Z.*, Volume 12, Number 4, 347-348.

842 Rosenzweig, C., Iglesias, A. & Yang X (2001) Climate change and extreme weather events; implications
843 for food production, plant diseases, and pests. *Global Change & Hum. Health* 2, 90–104.

844 Schaller, N., A. L. Kay, R. Lamb, N. R. Massey, G.-J. van Oldenborgh, F. E. L. Otto, S. N. Sparrow, R. Vautard,
845 P. Yiou, A. Bowery, S. M. Crooks, C. Huntingford, W. Ingram, R. Jones, T. Legg, J. Miller, J. Skeggs, D.
846 Wallom, S. Wilson & M. R. Allen (2015) Human influence on climate in the 2014 Southern England
847 winter floods and their impacts. *Nature climate change*, doi:10.1038/nclimate2927.

848 Seneviratne, S.I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B. & Teuling, A. J.,
849 (2010) Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science*
850 *Reviews*, 99, 3-4, 125-161.

851 Seneviratne, S. I. et al. (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate*
852 *Change Adaptation* 109–230 (Cambridge Univ. Press, 2012).

853 Seneviratne, S.I., Donat, M., Pitman, A. J., Knutti, R. & Wilby, R. L. (2016) Allowable CO₂ emissions based
854 on regional and impact-related climate targets. *Nature* 529, 477-483.

855 Sippel, S., F. E. L. Otto, M. Forkel, M. R. Allen, B. P. Guillod, M. Heimann, M. Reichstein, S. I.
856 Seneviratne, K. Thonicke, and M. D. Mahecha (2016) A novel bias correction methodology for climate
857 impact simulations, *Earth Syst. Dyn.*, 7, 71–88, doi:10.5194/esd-7-71-2016.

858 Stegehuis, A., R. Vautard, P. Ciais, R Teuling, M. Jung, and P. Yiou, 2013: Summer temperatures in
859 Europe and land heat fluxes in observation-based data and regional climate model simulations.
860 *Climate Dynamics*, 41, 455-477.

861 Stott, P.A., N. Christidis, F. Otto, Y. Sun, J.-P. Vanderlinden, G.-J. van Oldenborgh, R. Vautard, H. von
862 Storch, P. Walton, P. Yiou, F.W. Zwiers (2016) Attribution of extreme weather and climate-related events,
863 *WIREs Clim. Change*, 7, 23-41.

864 Sztobryn M., Stigge H-J, Wiebliński D, Weidig B, Stanislawczyk, I, Kańska A, Krzysztofik B, Kowalska B,
865 Letkiewicz B, Mykita M, (2005) Storm Surges in the Southern Baltic Sea (Western and Central Parts),
866 *Berichte des Bundesamtes für Seeschifffahrt und Hydrographie* Nr. 39.

867 University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D. (2015) CRU TS3.23: Climatic
868 Research Unit (CRU) Time-Series (TS) Version 3.23 of High Resolution Gridded Data of Month-by-month
869 Variation in Climate (Jan. 1901- Dec. 2014). Centre for Environmental Data Analysis, 09 November 2015.
870 doi:10.5285/4c7fdfa6-f176-4c58-acee-683d5e9d2ed5. [http://dx.doi.org/10.5285/4c7fdfa6-f176-](http://dx.doi.org/10.5285/4c7fdfa6-f176-4c58-acee-683d5e9d2ed5)
871 [4c58-acee-683d5e9d2ed5](http://dx.doi.org/10.5285/4c7fdfa6-f176-4c58-acee-683d5e9d2ed5)

872 Vautard, R., and B. Legras (1988) On the source of midlatitude low-frequency variability. Part II:
873 Nonlinear equilibration of weather regimes. *J. Atmos. Sci.*, 45, 2845-2867.

874 Vautard, R., P. Yiou, F. D’Andrea, N. de Noblet, N. Viovy, C. Cassou, J. Polcher, P. Ciais, M. Kageyama, and
875 Y. Fan (2007) Summertime European heat and drought waves induced by wintertime Mediterranean
876 rainfall deficit, *Geophys. Res. Lett.*, 34, L07711, doi:10.1029/2006GL028001

- 877 Wilcox, L.J., P. Yiou, M. Hauser, F. C. Lott, G. J. van Oldenborgh, I. Colfescu, B. Dong, G. Hegerl, L. Shaffrey,
878 and R. Sutton (2017) Multiple perspectives on the attribution of the extreme European summer of
879 2012 to climate change, *Climate Dynamics*, First Online, doi:10.1007/s00382-017-3822-7
- 880 Williams, K.D., C.M. Harris, A. Bodas-Salcedo, J. Camp, R.E. Comer, D. Copesey, D. Fereday, T. Graham, R.
881 Hill, T. Hinton, P. Hyder, S. Ineson, G. Masato, S.F. Milton, M.J. Roberts, D.P. Rowell, C. Sanchez, A. Shelly,
882 B. Sinha, D.N. Walters, A. West, T. Woollings, P.K. Xavier (2015) The Met Office Global Coupled model
883 2.0 (GC2) configuration, *Geosci. Model Dev.*, 8, 1509-1524.
- 884 Zhao Y., B. Sultan, R. Vautard, P. Braconnot, H.J. Wang and A. Ducharne (2016) Potential escalation of
885 heat-related working costs with climate and socio-economic changes in China. *Proc. Nat. Acad. Sci.*,
886 113, 4640-4645.

887 **Figure Captions**

888

889 **Figure 1:** JJA mean near-surface temperature: (a) mean state (1960-2013) from CRUTS3.23; (b) mean
890 state (1960-2013) from HadGEM3-A; (c) linear trends (1960-2013) from CRUTS3.23; (d) the number of
891 HadGEM3-A ensemble members simulating a trend smaller than observed; (e) rank histogram over all
892 land grid points counting the probability of the observations falling in each bin between the ranked
893 simulated values. (c), (d), (g), (h), (j) are the equivalent plots for precipitation. Hatching in panels (e)
894 and (g) indicates where trends are significant at the 10% level ($p < 0.1$); cross-hatching indicates
895 significance at the 5% level ($p < 0.05$).

896 **Figure 2:** Same as Figure 1 but for the winter season (DJF).

897 **Figure 3:** Left, middle and right panels: Evolution of seasonal mean daily mean temperatures in Europe,
898 Southern Europe and Northern Europe; First row: JJA daily mean temperatures; Second row: JJA daily
899 max temperatures; Third row: DJF daily mean temperatures; Fourth row: daily min temperatures.

900 **Figure 4:** Three left columns: parameters of the GEV distribution fitted to observations (left panels) and
901 the model simulations (center panels for the distribution of annual maxima in daily temperature (TXx)).
902 μ refers to the location parameter which is related to the mean value, σ the scale parameter, related to
903 the range, and ζ the shape parameter, diagnosing if the distribution is heavy-tailed (large value of ζ).
904 The differences between the parameters of the observed and simulated GEV fits are shown in the right
905 column of panels. For μ and ξ the difference is expressed in absolute terms; σ the difference is
906 expressed as a ratio. Stippling indicates areas where the observed-simulated difference is larger than
907 the 95% confidence intervals. Three right columns: same as left columns for the GEV distributions of
908 the minimal temperatures Tnn.

909 **Figure 5:** As Figure 4 but for the distribution of annual maxima in daily precipitation. The fourth row of
910 panel shows the ratio of the scale parameter σ and location parameter μ , with the difference again
911 expressed as a ratio.

912 **Figure 6:** Centroids of the four weather regimes sea-level pressure anomalies as obtained from the
913 NCEP/NCAR re-analyses (a-d for winter, i-l for summer) and HadGEM3-A (Ensemble of 15 members, e-
914 h for winter, m-p for summer). First column: Atlantic Ridge (AR) regime, second column: Blocking (BLO)
915 regime, third column: Negative NAO (NAO-) regime and fourth column: Zonal (ZO) regime for winter
916 (d, h) and Atlantic Low (AL) regime for summer (l, p).

917 **Figure 7:** JJA Composites of the standardized near-surface temperature for hot summer events over
918 Central Europe in Had-GEM3-N216 historical forcing ensemble members 1-15 (lines 1-3), ensemble
919 mean (line 4, left) and and 20CR ensemble mean (line 4, right). The composites have been derived
920 from all cases where the area-averaged and 5-day averaged temperature over Central Europe is larger
921 than its 95th seasonal percentile in JJA.

922 **Figure 8:** as figure 7, but showing composites of the standardized near-surface geopotential height at
923 500mb during hot summer events from Figure 7 over Central Europe.

924 **Figure 9:** Relationship between July evapotranspiration and June soil moisture averaged across
925 European subregions (left panels), in observations (gray) and HadGEM3-A (black). The considered time
926 period is 1989-2005. Correlation between July evapotranspiration and June soil moisture (right panels)
927 in observations (top) and HadGEM3-A ensemble median (middle). Bottom plot indicates whether or
928 not HadGEM3-A ensemble captures observed coupling strength. Considered time period is 1989-2005.

929 **Figure 10:** Relationship between temperature and evapotranspiration in July averaged across European
930 subregions, in observations (gray) and HadGEM3-A ensemble median (black). The considered time
931 period is 1989-2005. The range of correlations across HadGEM3-A ensemble members is shown in red
932 if the observed correlation is not contained.

933 **Figure 11:** Correlation between temperature and evapotranspiration in July in observations (top) and
934 HadGEM3-A ensemble median (middle). Bottom plot indicates whether or not HadGEM3-A ensemble
935 captures observed coupling strength. Considered time period is 1989-2005.

936 **Figure 12:** Composite of average duration (HWD) of the *longest* summer heat wave following the (top)
937 20% driest, and (bottom) 20% wettest springs for (left) E-OBS (1950-2015) and for (right) fifteen
938 HadGEM3-A historical members (1960-2013), based on Standardised Precipitation Index (SPI) averaged
939 over each bounded region (i.e. each regions' HWD pattern is composited based on its own wet and dry
940 spring ranking). Stippling indicates points that show a statistically significant difference at the 95% level
941 between dry and wet spring composites, based on a two sample Wilcoxon signed-rank test (Hollander
942 and Wolfe 1999). Significant differences are only marked on the dry-spring composite maps. HWD
943 values for regions without heat waves are set to zero. Each composite consists of 14 and 165 springs
944 for E-OBS and HadGEM3-A (i.e. 11 springs \times 15 ensemble members), respectively.

945 **Figure 13:** Correlations of winter zonal mean zonal wind anomalies at 60N with that at 10 hPa as
946 function of pressure and time lag. Positive lags mean that the stratosphere leads. Light and dark
947 shading identify regions where the correlations are significantly different from zero at the 5% and 1%
948 levels as estimated with a Monte-Carlo method that takes serial correlations into account. Top left:
949 NCEP daily. Top right: NCEP monthly. Bottom left: A typical member from HadGEM3-A ensemble.
950 Bottom right: correlations between the stratospheric vortex (zonal mean wind at 60 N, 10 hPa) and the
951 NAO as function of lag (positive lags mean that the stratosphere leads). Annual cycle has been
952 removed. Winter (DJF). NCEP (green), a typical HadGEM-3A ensemble member (blue), NCEP
953 interpolated from monthly values (green, dashed). The NAO is calculated as the leading principal
954 component of sea-level pressure.

955 **Figure 14:** Correlations between the stratospheric vortex and surface temperatures. Annual cycle has
956 been removed. Winter months (DJF). Large dots indicate correlations that have been estimated to be
957 significantly different from zero (5 % level) as estimated with a Monte-Carlo method that takes serial
958 correlations into account. Upper left panel: Observations (E-Obs for surface temperature, NCEP for
959 stratospheric vortex). Other panels: Different members from HadGEM-3A ensemble.

960 **Figure 15.** November anomalies of maximum water level [m] for 1971-2010 based on reconstructed
961 model data (Coastdat-TRIM) and historical HadGEM3-A-TRIM (hist) ensemble members 1-7. Selected
962 grid boxes represent locations co-located with German cities of Travemünde (left) and Warnemünde
963 (right).

964 **Figure 16:** Quantile-quantile plots of the distributions of the three ERA-Interim vs. HadGEM3-A derived
965 indices of wind in the South-West Baltic sea (see main text for definitions of the indicators). All wind
966 values or wind speeds are expressed as ms^{-1} .

967

968

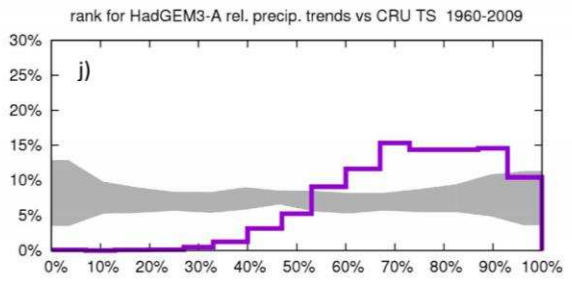
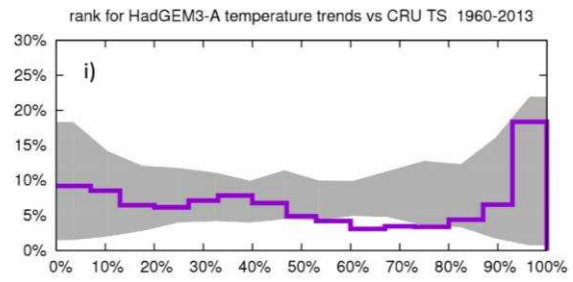
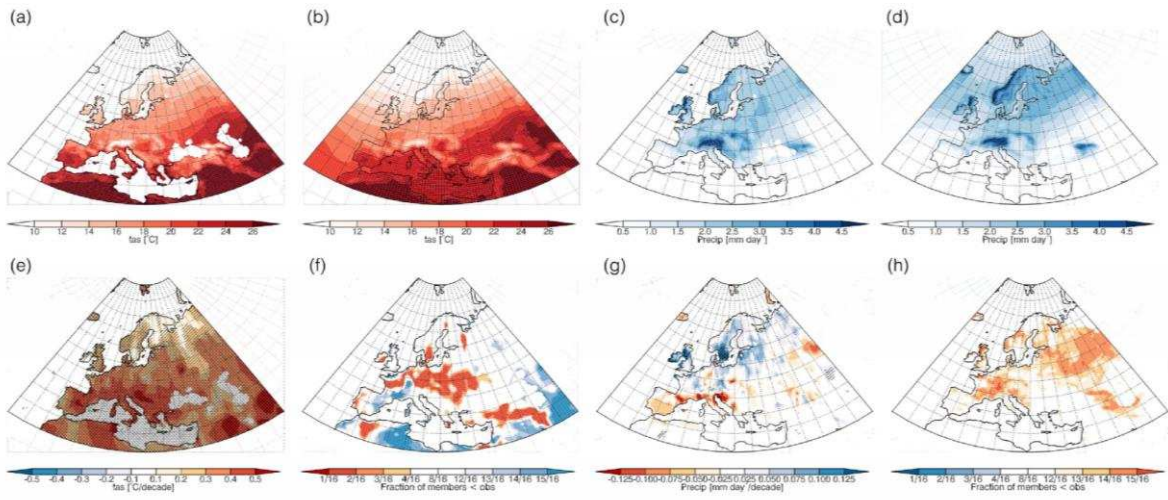
969

970

971

972

973

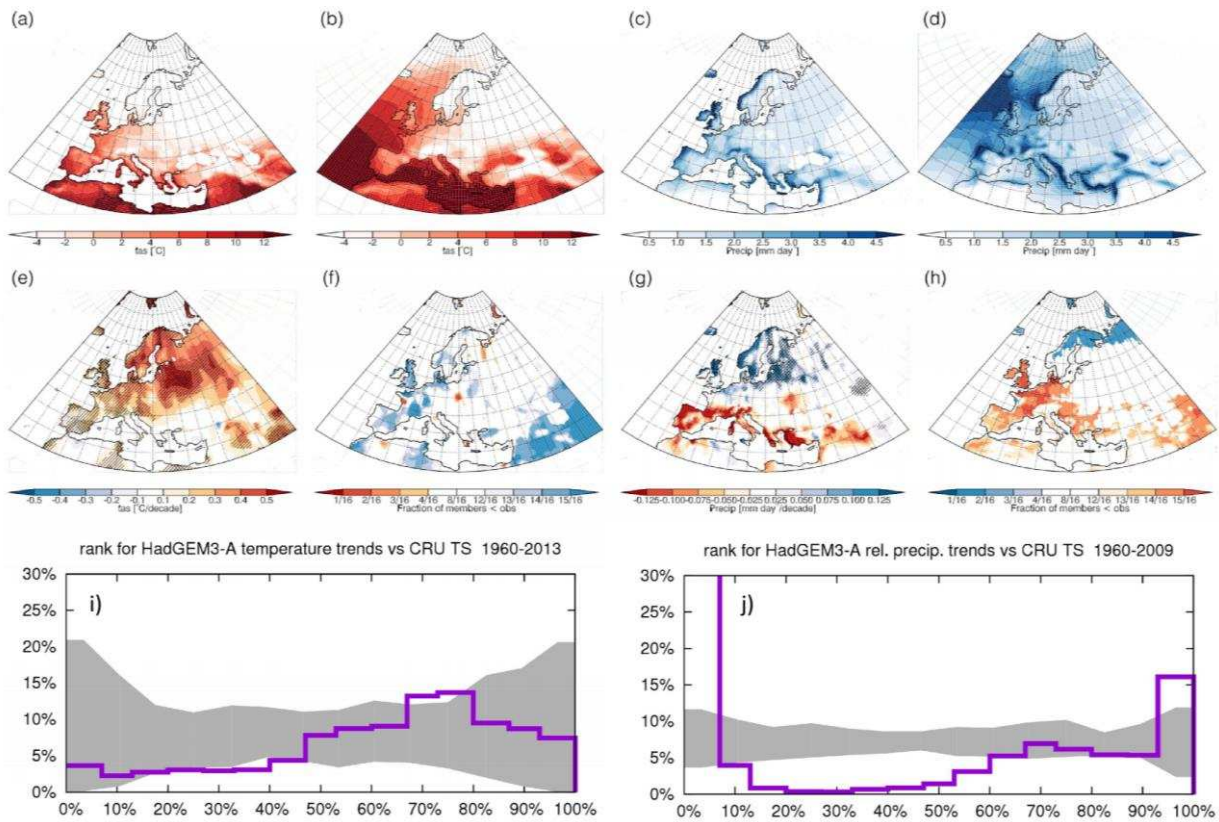


974
975
976
977
978

Figure 1

979

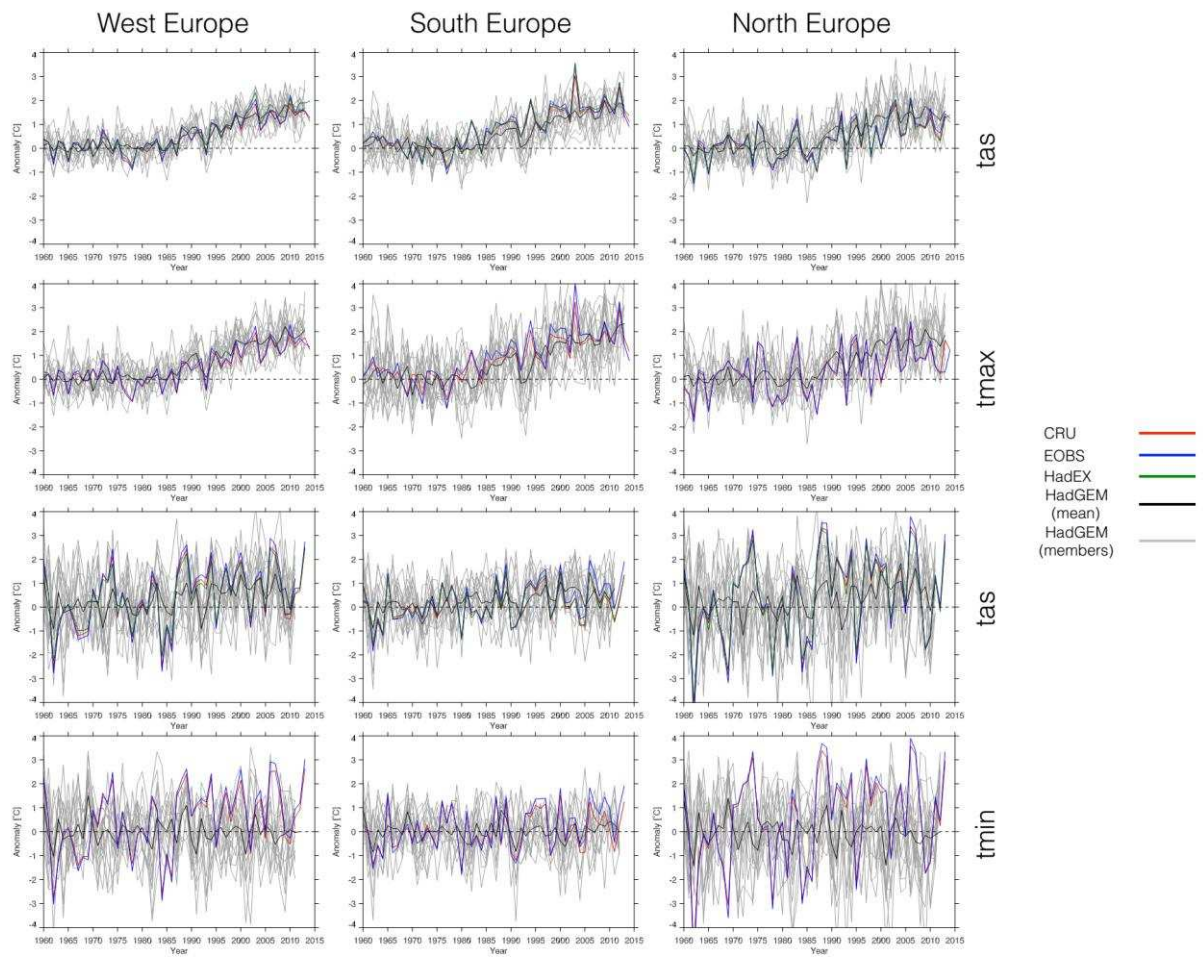
980



981

982

983 Figure 2

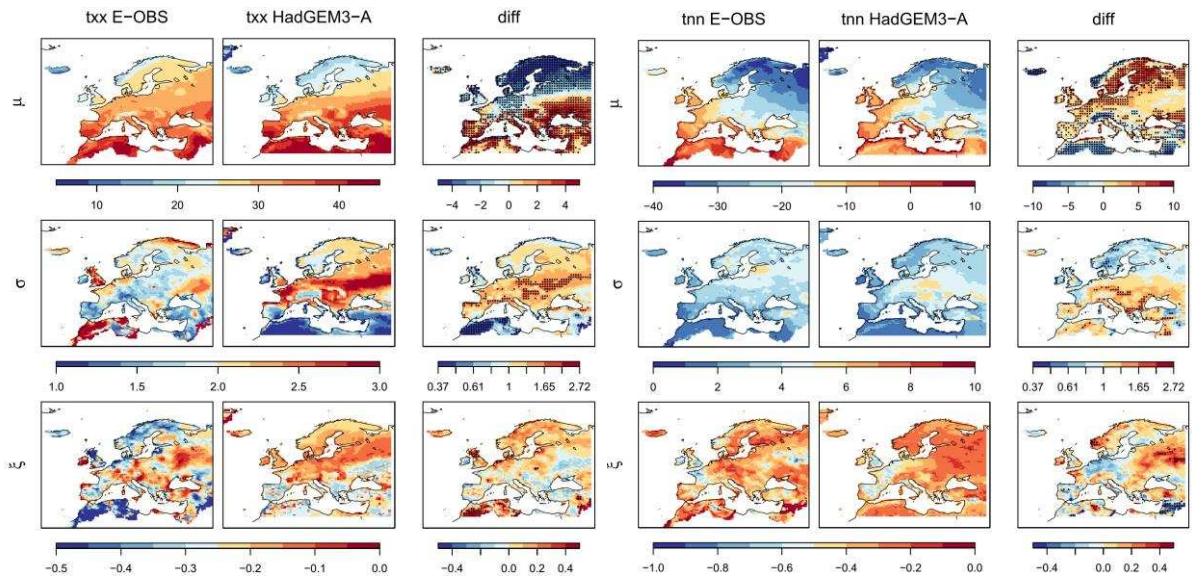


984

985

986 Figure3

987



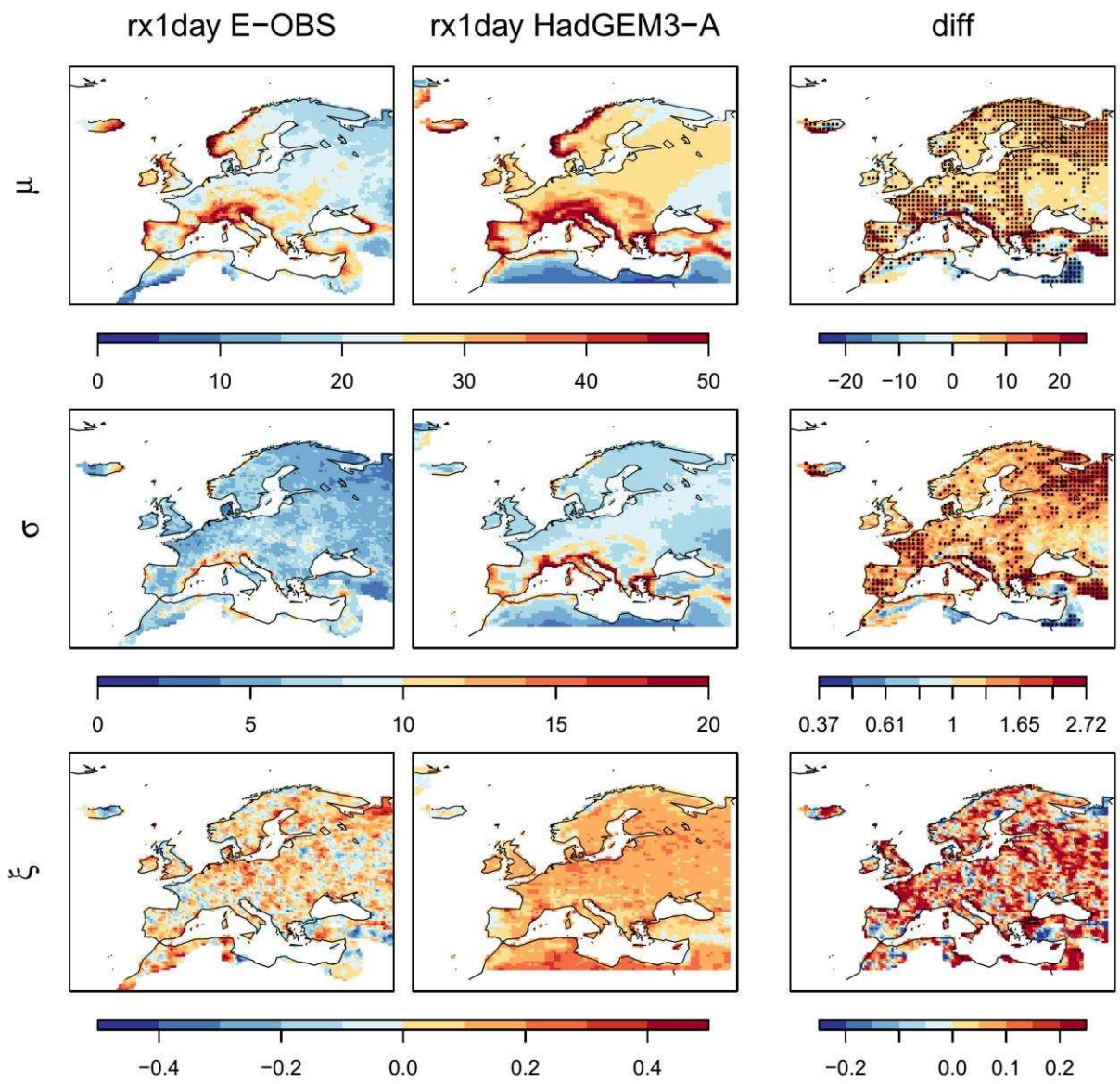
988

989

990

991 Figure 4

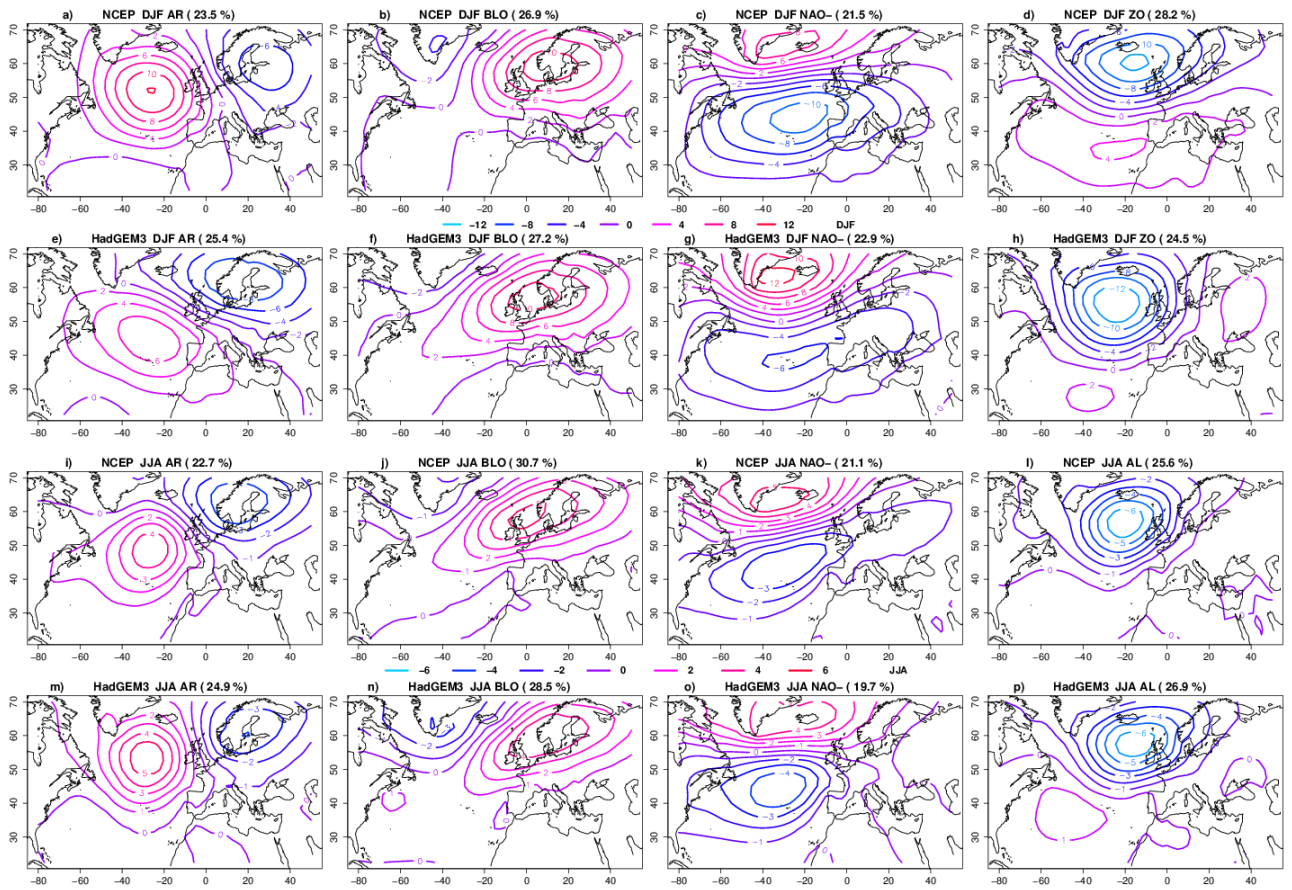
992



993

994 Figure 5

995

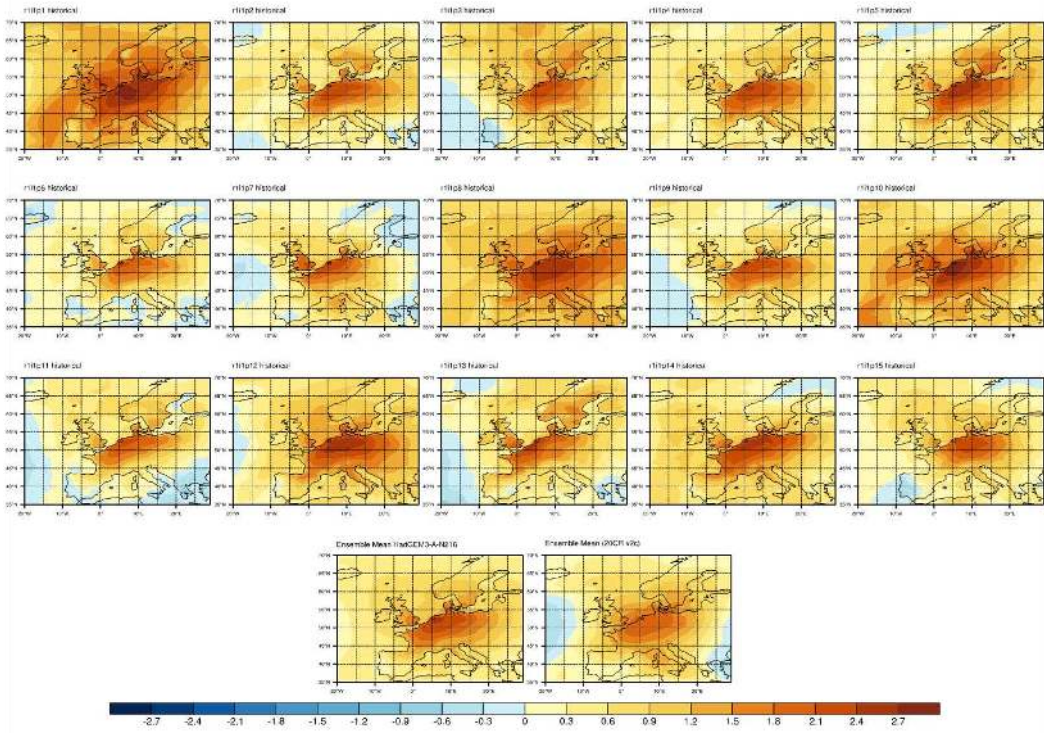


996

997

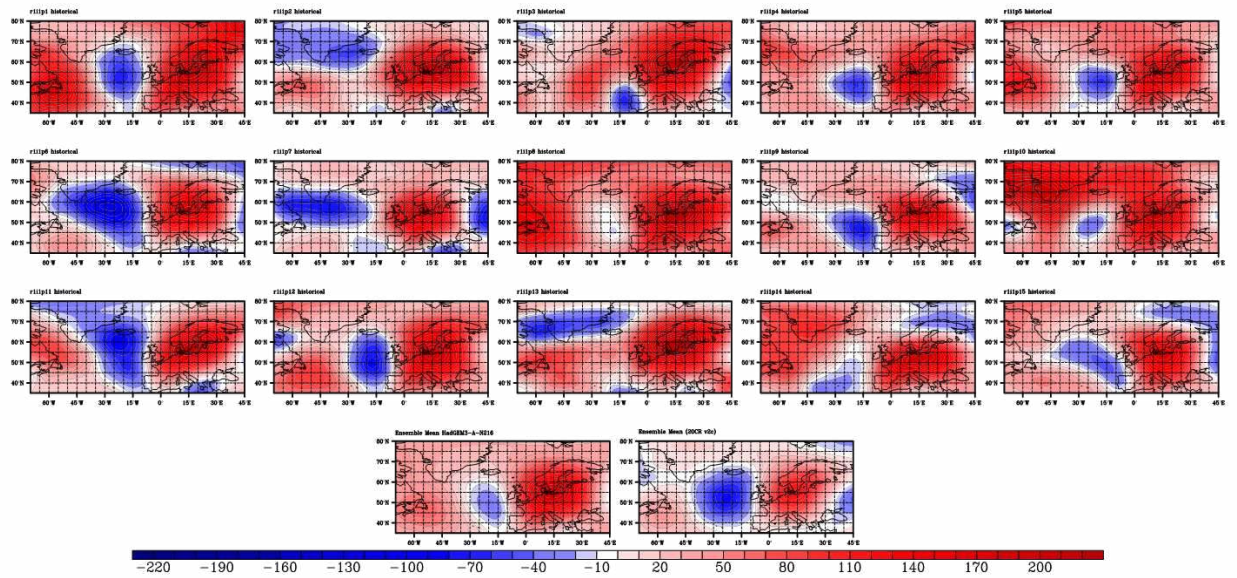
998

999 Figure 6



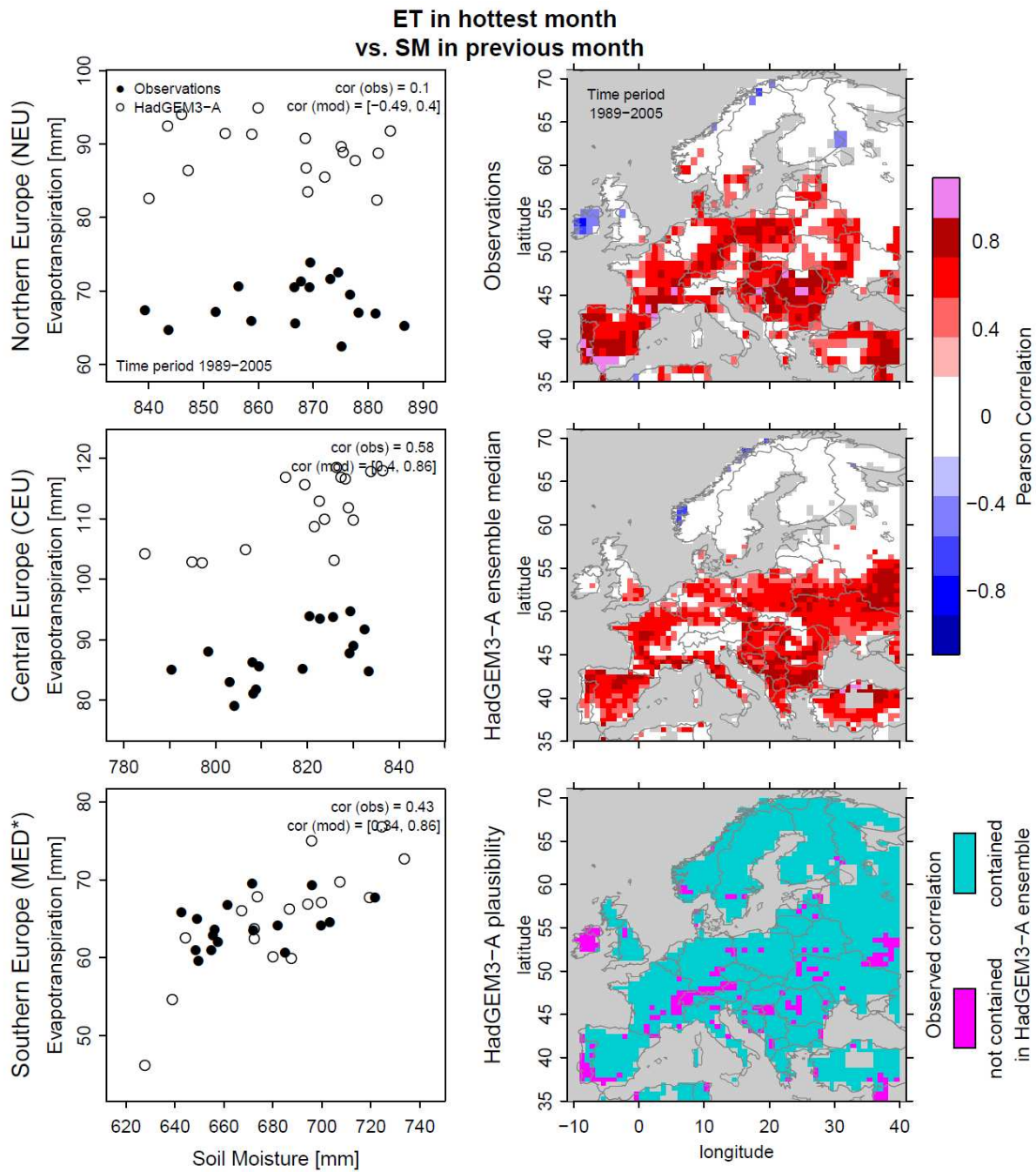
1000
 1001
 1002
 1003

Figure 7



1004
 1005

Figure 8



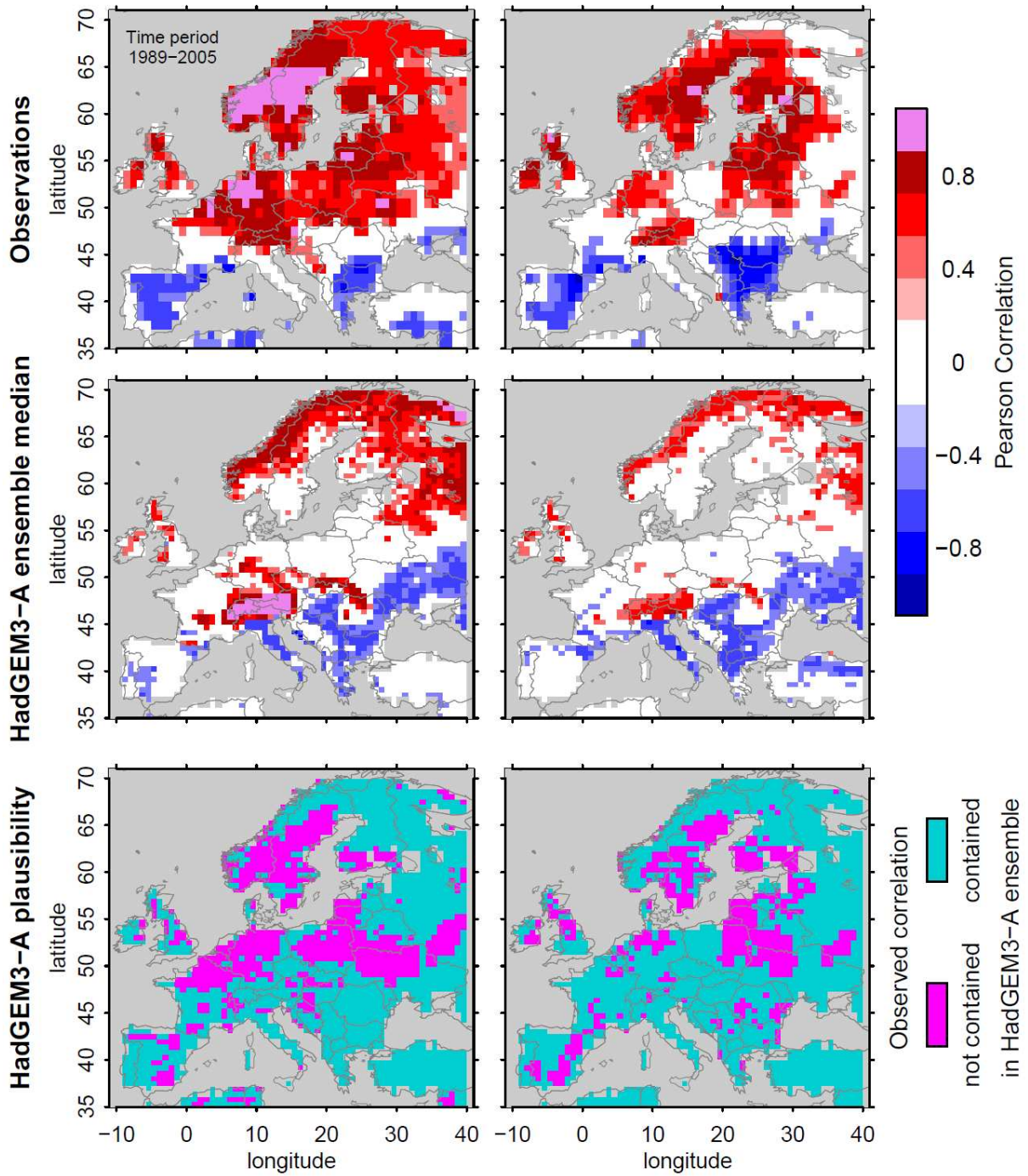
1007

1008

1009 Figure 9

1010

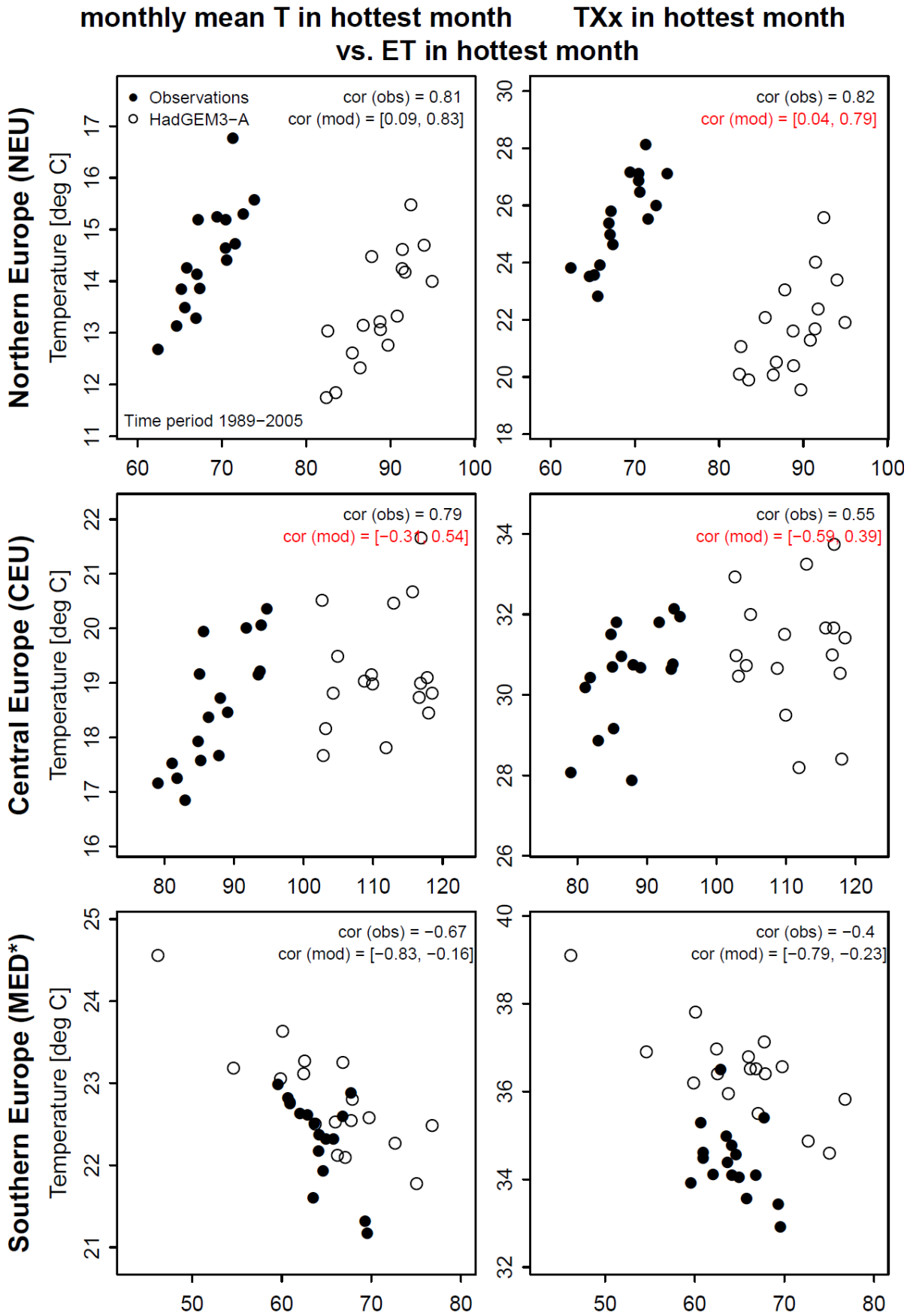
monthly mean T in hottest month TXx in hottest month
vs. ET in hottest month

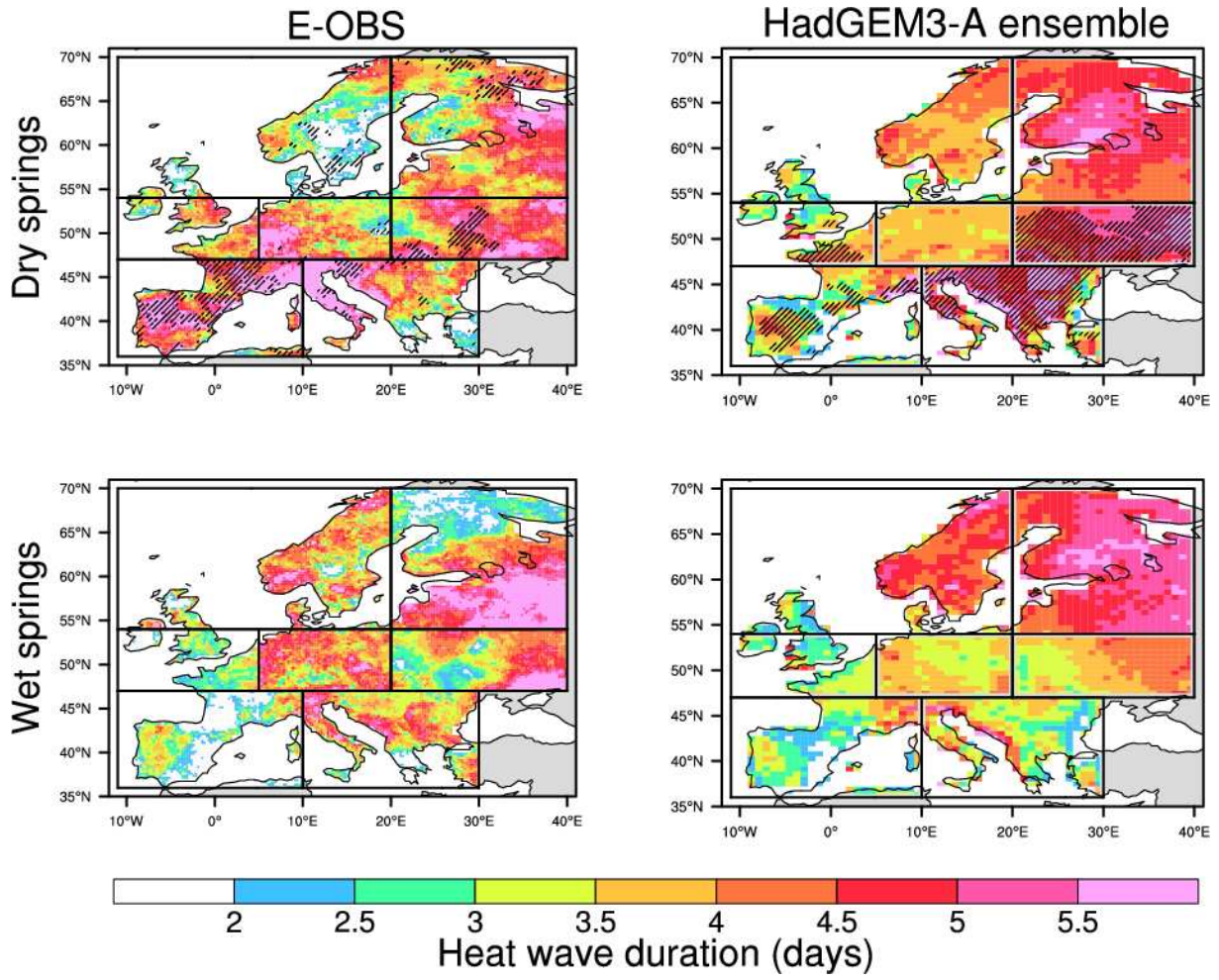


1011

1012 Figure 10

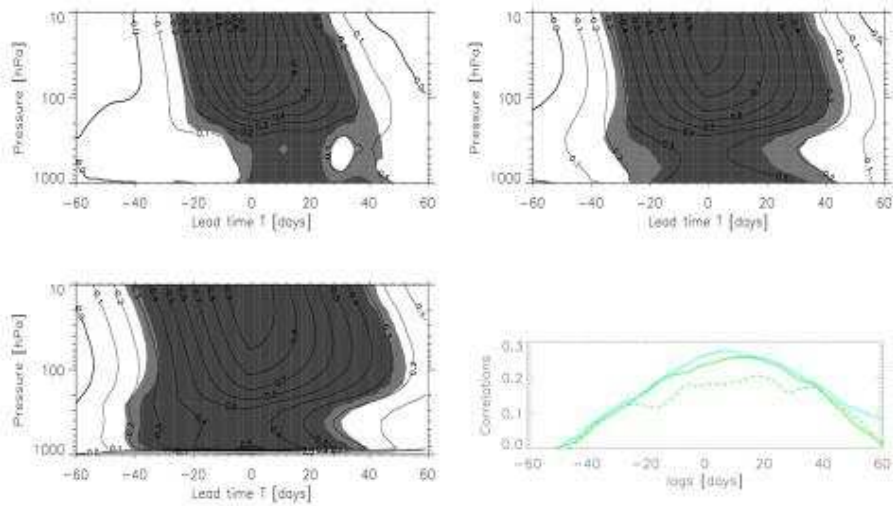
1013





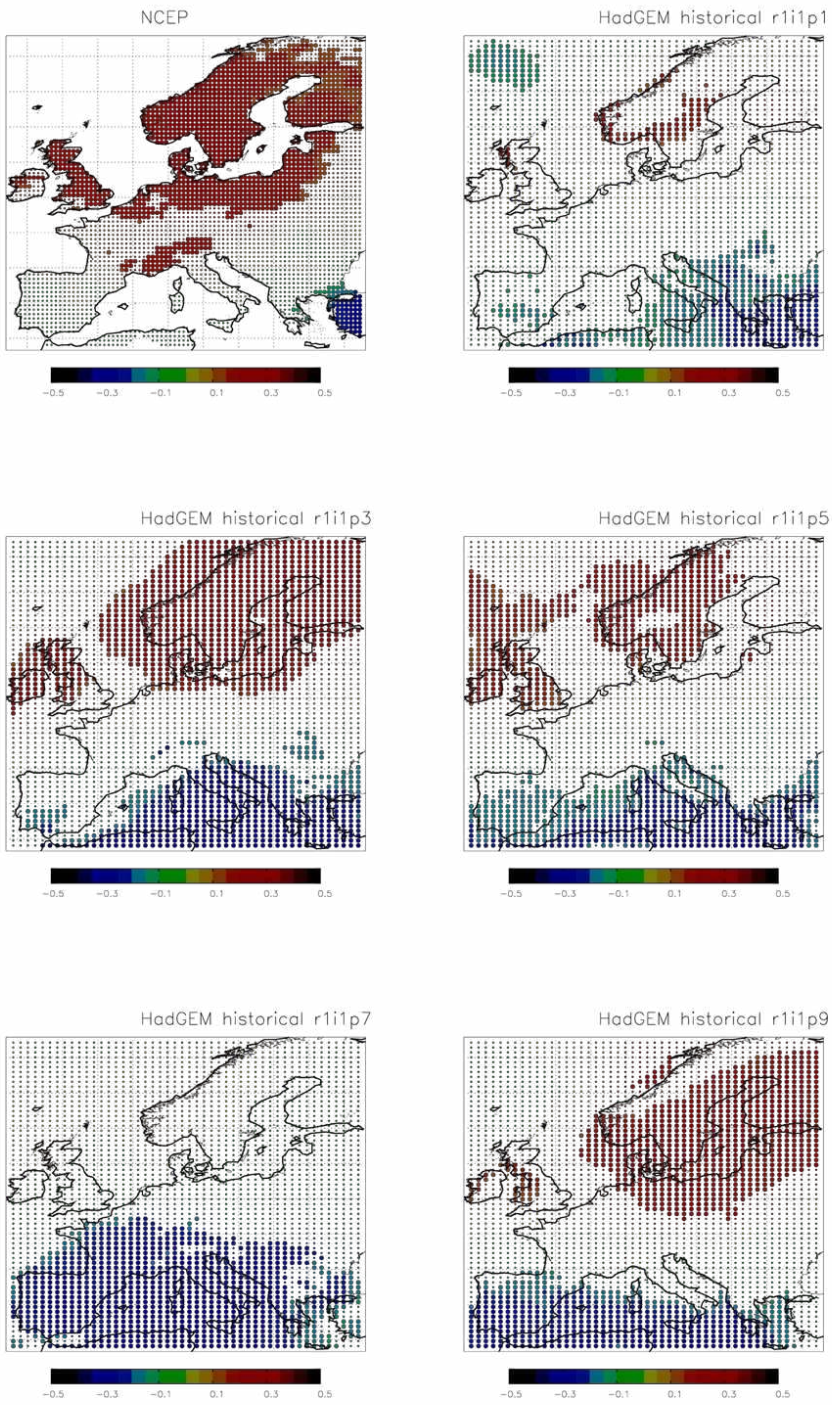
1017

1018 Figure 12



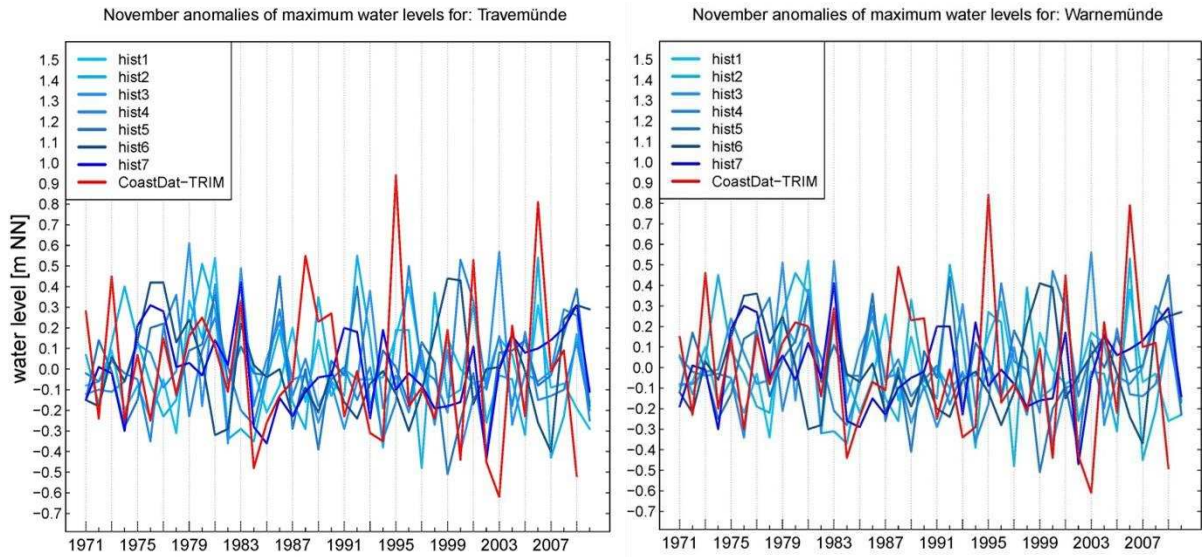
1019

1020 Figure 13



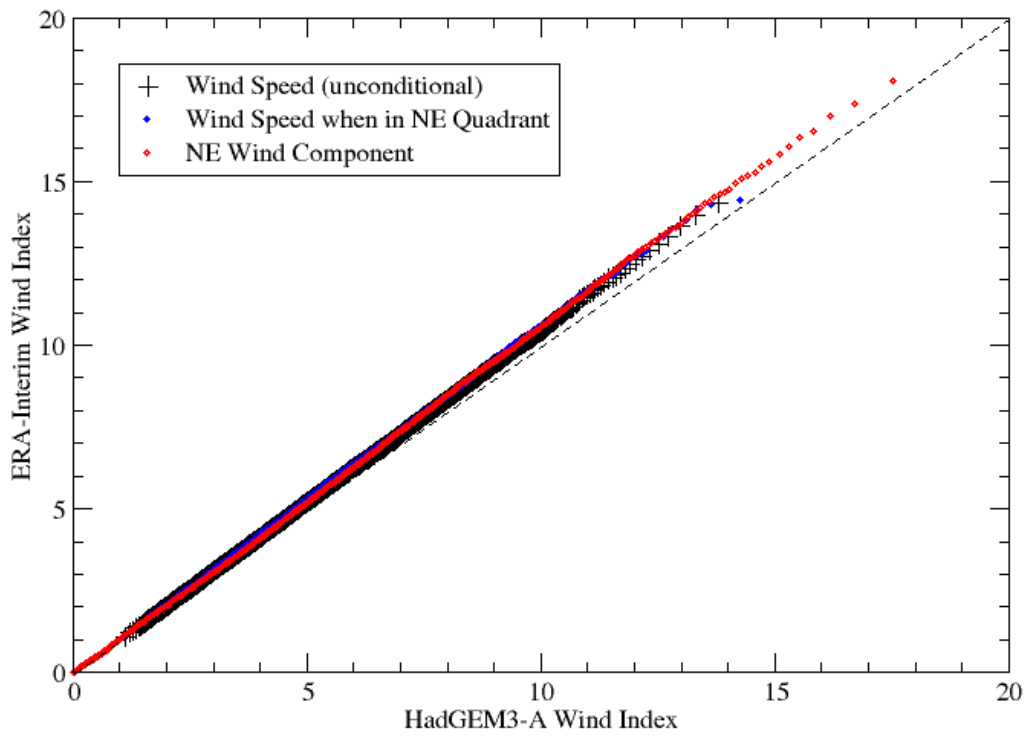
1021

1022 Figure 14



1023

1024 Figure 15



1025

1026 Figure 16

1027

1028 **Tables**

1029

	Dataset	Time period	Spatial Resolution
Temperature	E-OBS v14.0 (Haylock et al. 2008) CRUTS 3.23 (UEA 2015) 20CR reanalysis temperature data, averaged from 6hrly values	1960-2013	0.5°x0.5°
Precipitation	E-OBS v14.0 (Haylock et al. 2008)	1960-2013	0.5°x0.5°
Sea level pressure	NCAR/NCEP reanalyses NOAA 20CR reanalysis, version 2c	1948-2014	2.5°x2.5° 2°x2°
10-m winds	ERA-Interim reanalysis	1979-2013	0.7°x0.7°
Soil Moisture	SWBM Dataset (Orth and Seneviratne 2015)	1984-2013	0.5°x0.5°
Evapotranspiration	LandFlux-EVAL Dataset (Mueller et al. 2013)	1989-2005	1°x1°

1030

1031 **Table 1:** Overview of employed reference datasets

1032

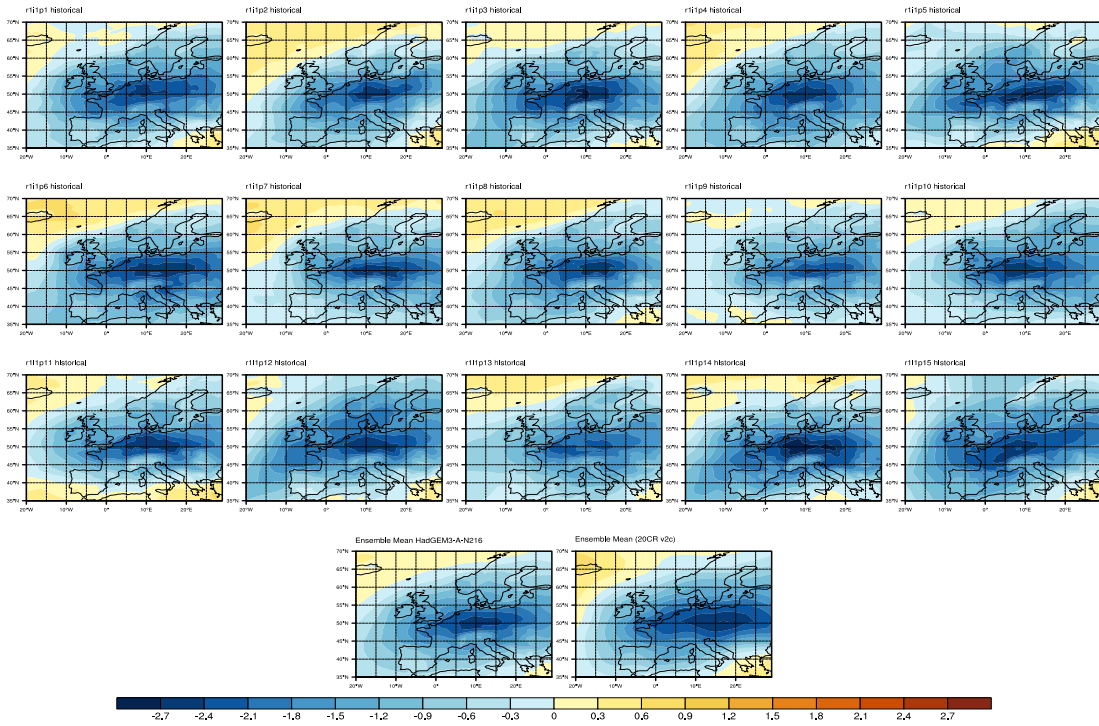
Regime	Winter: AR	Winter: BLO	Winter: NAO-	Winter: ZO	Summer: AL	Summer: BLO	Summer: NAO-	Summer: AR
NCEP/NCAR	24.4%	27.2%	21.0%	27.4%	22.6%	30.1%	21.2%	28.6%
HadGEM3-A (15 members)	23.8%	27.0%	22.5%	26.6%	18.5%	28.4%	24.6%	26.2%

1033

1034 **Table 2:** Weather regime occupancies (or frequencies) for each cluster, clusters being referenced from
1035 the NCEP/NCAR reanalyses, for each season.

1036

1037



1038

1039 **Supplementary Figure 1:** DJF Composites of the standardized near-surface temperature for cold winter

1040 events over Central Europe in Had-GEM3-N216 historical forcing ensemble members 1-15 (lines 1-3),

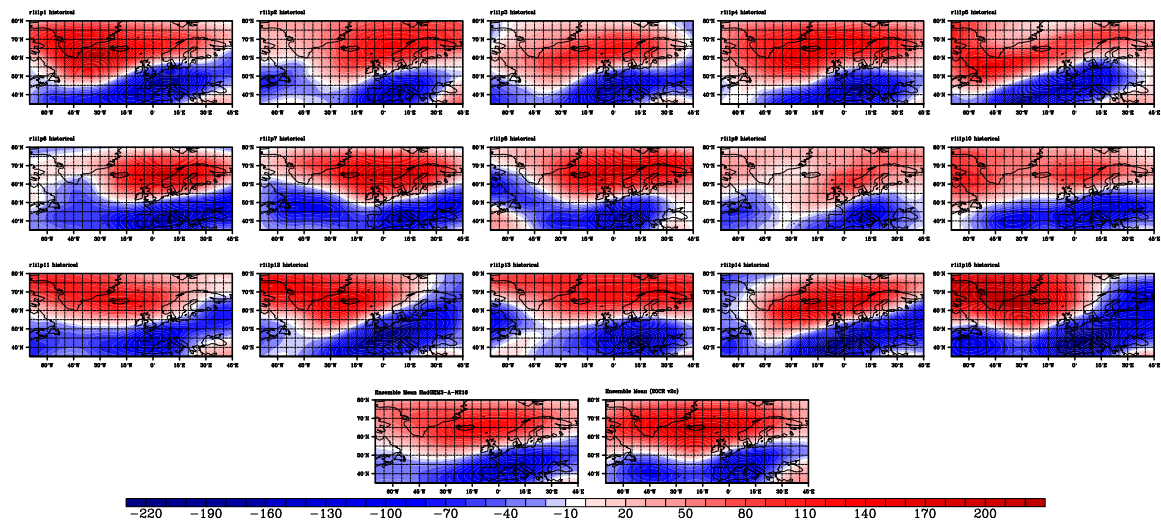
1041 ensemble mean (line 4, left) and and 20CR ensemble mean (line 4, right). The composites have been

1042 derived from all cases where the area-averaged temperature over Central Europe is smaller than its

1043 5th seasonal percentile in DJF.

1044

1045



1046

1047 **Supplementary Figure 2:** DJF Composites of the standardized geopotential height at 500mb for cold
1048 winter events over Central Europe in Had-GEM3-N216 ensemble members 1-15 (lines 1-3), ensemble
1049 mean (line 4, left) and and 20CR ensemble (line 4, right). The composites have been derived from all
1050 cases where the area-averaged temperature over Central Europe is lower than the 5th seasonal
1051 percentile in DJF of the associated temperature.

1052