Was the cold European winter 2009-2010 modified by anthropogenic climate 1 change? An attribution study.

Citation for published version:
Tett, S 2018, 'Was the cold European winter 2009-2010 modified by anthropogenic climate 1 change? An attribution study.', *Journal of Climate*. https://doi.org/10.1175/JCLI-D-17-0589.s1.

Digital Object Identifier (DOI):
10.1175/JCLI-D-17-0589.s1.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Climate

Was the cold European winter 2009-2010 modified by anthropogenic climate change? An attribution study.

Bo Christiansen*

Danish Meteorological Institute, Copenhagen, Denmark

Carmen Alvarez-Castro

Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Université Paris-Saclay, Gif sur Yvette, France

Nikolaos Christidis, Andrew Ciavarella

UK Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK

Ioana Colfescu

National Centre for Atmospheric Science, University of Leeds, UK

Tim Cowan

School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK

Jonathan Eden

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

Mathias Hauser

Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

Nils Hempelmann

Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Université Paris-Saclay, Gif sur Yvette, France

Katharina Klehmet

Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

Fraser Lott

UK Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK

Cathy Nangini

Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Université Paris-Saclay, Gif sur Yvette, France

Geert Jan van Oldenborgh

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

René Orth

Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

Peter Stott

UK Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK

Simon Tett

School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK

Robert Vautard

Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Université Paris-Saclay, Gif sur Yvette, France

Laura Wilcox

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

Pascal Yiou

Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Université Paris-Saclay, Gif sur Yvette, France

**Corresponding author address:* Bo Christiansen, Bo Christiansen, Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark.

E-mail: boc@dmi.dk

ABSTRACT

An attribution study has been performed to investigate the degree to which the unusually cold European winter 2009-2010 was modified by anthropogenic climate change. Two different methods have been included for the attribution: one based on a large HadGEM3-A ensemble and one based on a statistical surrogate method. Both methods are evaluated by comparing simulated winter temperature means, trends, standard deviations, skewness, return periods, and 5 % quantiles with observations. While the surrogate method performs well, HadGEM3-A in general underestimates the trend in winter by a factor of 2/3. It has a mean cold bias dominated by the mountainous regions and also underestimates the cold 5 % quantile in many regions of Europe. Both methods show that the probability of experiencing a winter as cold as 2009-2010 has been reduced by approximately a factor of two due to anthropogenic changes. The method based on HadGEM3-A ensembles gives somewhat larger changes than the surrogate method because of differences in the definition of the unperturbed climate. The results are based on two diagnostics: the coldest day in winter and the largest continuous area with temperatures colder than twice the local standard deviation. The results are not sensitive to the choice of bias correction except in the mountainous regions. Previous results regarding the behavior of the measures of the changed probability have been extended. The counter-intuitive behavior for heavy-tailed distributions is found to hold for a range of measures and for events that become more rare in a changed climate.

1. Introduction

An increased frequency of occurrence of extreme events such as flooding and heat waves has been reported (Frich et al. 2002; Alexander et al. 2006; Meehl et al. 2009; Coumou and Rahmstorf 2012; Peterson et al. 2012; Fischer and Knutti 2015) and, as the potentially most adverse consequences of climate change are related to extremes, there has been an increased interest in the attribution of such events (see, e.g., Field et al. 2012; National Academies of Sciences, Engineering, and Medicine 2016). A particular challenge is the attribution of single events. While there are a number of papers addressing event attribution of flooding and heat waves, there has not been much work done in this area addressing cold spells. Cold spells also increase morbidity and mortality, although the effect is weaker than for extreme warm events (Conlon et al. 2011). Furthermore, extreme winter conditions have serious detrimental effects on infrastructure such as damage to railways, closed airports, and frozen power lines (see, e.g., Doll et al. 2014, and references therein).

Part of the lesser interest in the attribution cold spells – at least in Europe – can be found in a weaker change in winter temperatures than in summer temperatures (see section 4). Together with the larger natural variability in winter, this makes changes in cold spells harder to detect. Cold spells in Europe are closely connected to the North Atlantic Oscillation (NAO) and blocking (Buehler et al. 2011), with a negative NAO index suggestive of cold European winters. Stratospheric sudden warmings propagate downwards on sub-seasonal time-scales and lead statistically to a negative phase of the NAO and associated colder temperatures in Europe (Baldwin and Dunkerton 1999; Christiansen 2001). In addition to the general warming expecting to reduce cold extremes (Van Oldenborgh et al. 2014), there have also been discussions about dynamical effects related to anthropogenic forcings that may change European winter temperatures and cold spells. One proposed connection is a positive correlation between autumn sea-ice extent and the atmospheric circulation, e.g., the NAO, the following winter, which has been studied both in observations (Francis et al. 2009; Overland and Wang 2010; Liu et al. 2012; Tang et al. 2013) and with modelling approaches (Petoukhov and Semenov 2010; Orsolini et al. 2011; Yang and Christensen 2012; Mori et al. 2014). In another model study Sévellec et al. (2017) found a link between sea-ice and the Atlantic meridional overturning circulation. With retreating sea-ice due to a general warming – and the Arctic amplification of that warming – such connections could help to explain the occurrence of recent cold winters in Europe. However, recent results (Li et al. 2015; Gerber et al. 2014; Screen 2017) suggest that the relationship between sea-ice, the NAO, and cold spells may be a chance occurrence or at least is very fragile. Recently, Francis (2017) related the unsettled science to a potential combination of a low signal-to-noise ratio and deficiencies in the models, the experimental designs, and the metrics of circulation changes. Other broad review of the Arctic influence on mid-latitudes are presented by Overland et al. (2015) and Cohen et al. (2014), while the reviews by Vihma (2014) and Gao et al. (2014) focus on the connection between sea-ice and mid-latitude weather and climate. Low-frequency changes in European cold spells may also be related to an intensified anticyclone that drives changes in the Siberian high (Zhang et al. 2012).

Here, we present an event attribution study of the cold European winter 2009-2010. The attribution is based on two different methods; the first is based on the ensembles produced by the HadGEM3-A (Hadley Centre Global Environment Model version 3) atmospheric model and the second on ensembles generated by a statistical surrogate method.

The paper is organized as follows. In section 2 we describe the data and the diagnostics used for the event attribution of cold spells. Therein, we also briefly describe the meteorological details of the winter 2009-2010 (see also WMO (2010)) focusing on these diagnostics. The two methods for generating ensembles – the HadGEM3-A model and the statistical surrogate method – are described in section 3. In section 4 we evaluate these two methods against observations. In section 5 we present the resulting risk ratios. In appendix A we expand the discussion of the framing issue of attribution of single events from Christiansen (2015) to be more relevant for the present study. The extension includes other measures of the risk than just the Fractional Attributable Risk and also the situation where the considered event becomes less frequent in the changed climate. The conclusions are presented in section 6.

2. The observations, the diagnostics, and the winter 2009-2010

For surface temperature observations we use the E-OBS (version 12) daily mean gridded data-set on a 0.5x0.5 longitude/latitude land-only grid (Haylock et al. 2008). Uncertainties in the E-OBS data and comparisons with re-analyses are presented in van der Schrier et al. (2013), who find good agreement between European mean trends in the different data-sets.. We also use daily zonal wind from the National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis (NCEP/NCAR) reanalysis on a 2.5$\times$2.5 longitude/latitude grid and 17 pressure levels from 1000 to 10 hPa (Kalnay et al. 1996). To calculate the NAO index we use NCEP daily sea-level pressure on a 2.5x2.5 longitude/latitude grid. For all three data sets we use the 54 year long period 1960-2013 which is also the period for which the experiments with HadGEM3-A have been performed (see section 3). We select E-OBS data for Europe, defined here as latitudes between 35 and 70° N and longitudes between 10° W and 30° E. For the E-OBS data we exclude grid-points where more than 5 % of the days are missing data. This affects only small regions on the African coast. Grid-points that are missing data between 0 and 5 % of the days are filled using nearest neighbour interpolation. This affects a few grid-points on the African coast and in Turkey.

The NAO is calculated by empirical orthogonal function (EOF) analysis of winter (DJF) monthly anomalies of sea-level pressure for latitudes between 20 and 80° N and longitudes between 90° W and 40 ° E. The anomalies are first weighted by the square-root of the cosine of the latitudes and linearly detrended. Daily values of the NAO index are then found by projecting the leading EOF onto daily sea-level pressure anomalies (see, e.g., Blessing et al. 2005).

There are many possible diagnostics of the severity of cold winters including different combinations of the duration, extent, and intensity of the cold periods. In the following we focus on two diagnostics. The first diagnostic is defined on grid-cell scales as the minimum temperature over the whole winter. The second diagnostic, herefrom denoted the blob index, is a spatially integrated measure defined as the largest continuous area with temperature anomalies less than -2σ, where σ is the local, seasonally varying standard deviation, i.e., the standard deviation calculated for each grid-point and for each day of the year. Thus, the blob index is a combined measure of both the spatial coherence and the intensity of the cold spell. The blob index is calculated for each day separately and for convenience expressed as a fraction of the total European land area. Both diagnostics are calculated from daily mean temperatures. The first diagnostic measures the intensity of the cold period while the second diagnostic also takes spatial extent into account, and is similar to the heat-wave diagnostic used in Christiansen (2015).

We now briefly describe the winter 2009-2010 with focus on the chosen diagnostics; the minimum temperature over whole winter and the blob index. The winter 2009-2010 was a relatively cold winter with a series of strong cold spells of which the strongest appeared in the middle of December. The blob index reached a value of 0.38 on 19th December (Fig. 1, top panel), which is large but exceeded in both earlier and later winters, e.g., in the winter of 2011-2012. On 19th December 2009 the temperature was below normal almost everywhere except for few regions in

Northern Scandinavia (Fig. 2). The coldest anomalies, below -4σ, are found in the middle of Germany[1].

The temperature of the coldest day of the winter 2009-2010 confirms that this year was unusually cold in many regions of Europe (Fig. 3). In Germany, Spain, Great Britain, and Scandinavia temperatures as cold as in 2009-2010 are rarely found in other years in the period 1960-2013.

The winter 2009-2010 was, as many other cold winters, dominated by a strong negative NAO (Wang et al. 2010; Ouzeau et al. 2011; Buchan et al. 2014) (demonstrated in the upper panel of Fig. S1 in the supplement). However, this winter might not have been as cold as previous winters with the same NAO levels, suggesting an impact of a general warming climate (Cattiaux et al. 2010). The negative NAO was connected to a weak stratospheric vortex (Cohen et al. 2010; Vargin 2015) – as demonstrated in the lower panel in Fig. S1 – although the main factor responsible for the strong negative NAO has been suggested to be related to internal tropospheric dynamical processes (Jung et al. 2011).

3. The two ensemble methods

To make statements about the attributable risk of the observed extreme event (the winter 2009-2010) we need information about the frequencies of similar events of different magnitudes in both the unperturbed climate and in the climate under anthropogenic forcings (Allen 2003; Stott et al. 2004, 2013). For each of the climates the probability for finding an event at least as extreme as the observed event is calculated. The risk ratio is then defined as the ratio between these two probabilities. See also appendix A for a more precise definition of the risk ratio and other measures of the attributable risk. To obtain these frequencies we here use ensembles both from

[1]The lead author got stuck in airports in Manchester and then Amsterdam on the way home from AGU. The meteorological conditions are described here `https://en.wikipedia.org/wiki/Winter_of_2009-2010_in_Europe`

the atmospheric general circulation model HadGEM3-A and ensembles obtained by a surrogate field method that produces fields with the same spatial and temporal structure as an observed target field. These methods complement each other as they make different assumptions about the effect of anthropogenic climate change. Note, that for the HadGEM3-A approach the unperturbed climate is represented by pre-industrial (1850) conditions while for the surrogate method it is represented by 1960 conditions.

a. The dynamical model

Two ensembles, each with 15 members, have been produced with HadGEM3-A covering the years 1960-2013. The horizontal resolution is N216 and the vertical resolution is L85 with 50 tropospheric and 35 stratospheric layers. The version used here is discussed in Ciavarella et al. (2017) and includes the Global Atmosphere 6.0 (GA6) atmospheric science package (Walters et al. 2016). Both ensembles were recently used for attribution analysis by Christidis et al. (2016), Eden et al. (2016), and Burke et al. (2016). A detailed analysis of the perturbed (historical) ensemble regarding the skill in extreme events is presented in Vautard et al. (2017). We further note that no significant correlations between the Arctic autumn sea-ice and the winter NAO are found in these ensembles. This holds both when total Arctic sea ice and regional sea-ice (e.g., the Kara-Barents Seas) is considered.

The two ensembles differ through the external climate forcings included, one is driven with both natural and anthropogenic forcings (historical) and the other with only natural forcings (histnat). Natural external forcings are variability in total solar irradiance at the top of the atmosphere, and volcanic activity implemented through a latitudinal variation of stratospheric aerosol optical depth. Anthropogenic forcings include well-mixed greenhouse gases, zonal-mean ozone concentrations, aerosol emissions, and land use changes. The external forcings are obtained from sources used by the Coupled Model Intercomparison Project Phase 5 (CMIP5) generation of models (Taylor et al. 2011). In the histnat experiments, anthropogenic forcings are held at pre-industrial levels taken to be those of 1850. Boundary conditions at the bottom of the atmosphere are given by sea-surface temperatures (SST) and sea-ice concentrations fields. In the historical experiments the SSTs and the sea-ice are prescribed from observed values (HadISST1.1, Rayner et al. 2003) while for the histnat experiments an estimate of the change due to anthropogenic influence is removed from the observations (Christidis et al. 2013). This estimate comes from ensembles of simulations with and without anthropogenic forcings generated with 19 coupled models for the C20C+ detection and attribution project (`http://portal.nersc.gov/c20c/experiment.html`).

Both ensembles share a common atmospheric initialization on 1st December 1959 from ERA-40 reanalysis fields (Uppala et al. 2005). The differences between ensemble members are produced by two stochastic physics schemes that generate small differences in the physics of each simulation (Christidis et al. 2013).

b. Ensemble surrogate field method

The method is based on a simple algorithm to produce ensembles of surrogate fields based on observations. This method produces surrogate fields with the same spatial and temporal structure – as measured with instantaneous and lagged cross-correlations – as the original observed field of surface temperatures. The method was used in Christiansen (2015) for attribution of heat waves and in a study of the significance of the increase in warm records (Christiansen 2013). The surrogate fields are generated with a phase-scrambling procedure described in Christiansen (2007, 2013) which is very similar to the multivariate method introduced by Prichard and Theiler (1994) based on the univariate amplitude adjusted Fourier transform method (AAFT) by Theiler et al. (1992).

The general outline of the procedure is familiar from bootstrap methods; first a transformation of the original field into stationary anomalies is performed, stationary surrogate anomalies are produced from the original stationary anomalies, and the final surrogate field is produced by applying the inverse transformation to the surrogate anomalies.

The stationary anomalies of the original observed surface temperature field are obtained by removing the average annual cycle and the secular variations – trends and variability on the lowest frequencies estimated by a 3rd order polynomial fit – at each geographical position. The resulting stationary anomalies are Fourier transformed, then the Fourier phases are randomized but with the same random phases for all grid-points, and finally inverse Fourier transforms are performed to get the stationary surrogate anomalies. Now the average annual cycles are restored at each geographical position to get a surrogate field of the unperturbed climate state, i.e., 'the world that could have been without climate change'. Also adding the secular trends to this field gives us a surrogate of the perturbed climate.

Repeating this process with different randomizations allows us to calculate ensembles of fields for both the unperturbed climate and the perturbed climate. From these ensembles the relevant distributions of the diagnostic can be calculated and the risk ratio for an observed event can be estimated.

The surrogate method is fast and flexible and can therefore also be used for sensitivity studies and to test the robustness of the risk ratio to methodological choices. The method does not depend on physical parameterizations but only on statistical assumptions. A fundamental assumption is that it is possible in the observations to empirically separate internal variability from climate change. Here this separation is performed by assuming different temporal scales for the two types of variability. The method was tested in details in Christiansen (2015) and found to be adequate for temperature fields while problems may arise for fields which are strongly non-Gaussian. In agreement with the analysis in Christiansen (2015) we find here similar results for cold spells when climate change is defined by 5th or 7th order polynomials.

4. Evaluation

In this section we investigate the extent to which HadGEM3-A and the surrogate methods reproduce the relevant features of the observations. Our confidence in the calculated risk ratios depends on the methods ability to reproduce long-term temperature trends as well as cold extremes.

The statistical significance of trends and differences is estimated by Monte-Carlo methods that take the possible serial correlations of the data into account. The statistical significance of trends are calculated by a phase-scrambling method (Theiler et al. 1992; Christiansen 2001) for which the 'bootstrap' members retain the full auto-correlation spectrum of the original detrended time-series. The significance of differences are calculated by a block-bootstrap method assuming that data separated by 15 days are independent. This separation corresponds to roughly twice the temporal decorrelation length of surface temperatures (see, e.g., Christiansen 2015).

We will use 'historical' and 'histnat' to denote the two ensembles from HadGEM3-A. For the surrogate method we use 'perturbed' and 'unperturbed' ensembles. So 'histnat' and 'unperturbed' ensembles refer to the counter-factual world that could have been.

Some general evaluations related to cold spells were presented in Vautard et al. (2017) based on the historical HadGEM3-A ensemble. It was concluded that there were no major processes hindering the representation of cold spells. Here we will focus on quantities directly related to the two diagnostics and compare the evaluations of the dynamical model and the surrogate method.

a. The European mean perspective

The observed spatially averaged European winter (DJF) mean temperature has a linear trend of 0.30 °C/decade (95 % confidence interval is $[0.12, 0.51]$ °C/decade) in the period 1960-2013 (Fig. 4). This is somewhat larger than the ensemble mean of the HadGEM3-A historical ensemble which shows a trend of 0.20 °C/decade (95 % interval $[0.12, 0.28]$). Both these trends are significant to the 5 % level while only approximately half of the individual HadGEM3-A historical ensemble members show significant trends. However, 3 out of the 15 ensemble members show a trend that is comparable to that of the observations. The trends are probably due to a combination of increasing greenhouse gases and decreasing European aerosol emissions. However, there is no significant difference in the trends calculated for the whole period, the period before 1985, and the period after 1985, neither for observations nor models. It is also worth noting that the HadGEM3-A model has a negative bias which is dominated by mountainous regions as seen in the next sub-section.

The ensemble mean of the perturbed ensemble of surrogates has a linear trend of 0.34 °C/decade (significant to the 5 % level, 95 % interval $[0.26, 0.42]$) close to that of the observations as should be expected by construction. The ensemble of surrogates shows less variation among ensemble members than does the HadGEM3-A ensemble, and all of them show significant trends. The unperturbed ensemble mean and the histnat ensemble mean show weak and insignificant trends. The NAO index has a weak non-significant trend in the observations while it is almost zero in the two HadGEM3-A ensembles (not shown).

The correlation of the European mean winter temperature between observations and the ensemble mean of the HadGEM3-A historical ensemble is 0.47 (95 % confidence interval is $[0.15, 0.71]$). For the HadGEM3-A histnat ensemble the correlation is 0.29 ($[0.01, 0.53]$). As expected the correlations for the surrogate ensembles are smaller, 0.28 ($[-0.14, 0.60]$) and 0.02 ($[-0.28, 0.32]$), reflecting that for this method only the trend will contribute. For the observations the correlation between the European mean winter temperature and the NAO index is 0.67 ($[0.40, 0.82]$), and similar values (0.61 and 0.63) are found for the two HadGEM3-A ensembles. Correlations of winter mean NAO index between observations and the two HadGEM3-A ensemble means are 0.19 ($[-0.03, 0.41]$) and 0.22 ($[-0.03, 0.46]$), while the correlation between the NAO index in the two ensemble means is 0.52 ($[0.29, 0.70]$). Thus, for both observations and the HadGEM3-A ensembles the SSTs determine a considerable part of the average European land temperature and the NAO index and the land temperature are well correlated. However, the NAO itself is only to a limited extent determined by SSTs (see, e.g., Greatbatch 2000, and references therein).

To get an overall impression of the changes in winter extremes we normalize the local temperatures for each grid-point with the local, seasonally varying standard deviation (calculated for each grid-point and for each day of the year) and pool them all together (Fig. 5). The challenge of detection and attribution of cold extremes becomes clear: although there is a general change in the distributions the changes are particularly small for the cold tail. This is quantitatively different from summer temperatures (Fig. S2) which show a general shift of the whole distribution toward warmer values. Both the HadGEM3-A historical ensemble and the perturbed surrogate show changes comparable to observations. Note also that the distributions in winter are heavily negatively skewed so that the values in the negative tail are numerically larger than those in the positive tail. This is in agreement with the observation (Twardosz and Kossowska-Cezak 2016) that more extreme cold than extreme warm winters are observed.

The blob diagnostic combines intensity and spatial coherence of the cold spell and requires a specific validation. In Fig. 1 the diagnostic is shown as function of time for a random historical HadGEM3-A ensemble member and for a random perturbed surrogate ensemble member. The two ensemble members compare well with observations. Figure 6 shows the return periods including only winter days of the historical HadGEM3-A and the perturbed surrogate ensembles, as well as for observations. We see that both the surrogate method and HadGEM3-A reproduce the observed return periods of the largest continuous area very well. However, there is a tendency for the HadGEM3-A to overestimate the return periods for events smaller than 0.35.

b. The local perspective

In sub-section 1 we present an evaluation based on all winter days while we in sub-section 2 briefly add to the evaluation of the temperatures of the coldest winter days presented in Vautard et al. (2017).

1) Evaluation based on all winter days

The mean of the local temperatures over the winters 1960-2013 is relatively well modelled in the historical HadGEM3-A ensemble (Fig. 7), with a bias that is small (although statistically significant) except for the alpine region and regions in Northern Scandinavia. In these mountainous regions the model is up to 5°C colder than the observations. The long term mean difference between the historical and histnat model is statistically significant and positive everywhere with the strongest warming in the north eastern part of Europe – reaching 4°C in Finland – and the weakest warming in the south western part. For the surrogate method (not shown) the long term mean is by construction almost indistinguishable from that of the observations.

The linear trend of the local temperatures over the winters 1960-2013 (Fig. 8) is positive nearly everywhere in the observations with the largest trends in the north eastern regions. The trends are statistically significant in large areas. The same pattern but of weaker strength and lower significance is found in the historical HadGEM3-A experiments (see also Vautard et al. (2017)).

The trends for the perturbed surrogate have the same magnitude as in observations. For the histnat and unperturbed ensembles the trends are close to zero everywhere. The pattern of the differences in the mean between HadGEM3-A historical and histnat ensembles (bottom right panel in Fig. 7) and the trends in observations and the HadGEM3-A historical ensembles (left panels in Fig. 8) are in general agreement with the expected Arctic amplification.

The standard deviation, the skewness, and the 5 % quantile of the local temperatures are shown in Figs. 9, 10, and 11. These quantities are calculated from winter anomalies over the period 1960-2013 after removing the seasonal cycle and the secular trend in form of a 3rd order polynomial fit. The figures include the observations (upper panels), the historical HadGEM3-A and perturbed surrogate (middle panels), the difference between the historical HadGEM3-A and observations and the difference between the historical and histnat HadGEM3-A (lower panels).

Compared to the observations, the standard deviation in the historical HadGEM3-A model is overestimated in the mountainous regions (Fig. 9). The modelled skewness is strongly overestimated compared to observations in Scandinavia, while it is underestimated in north-eastern parts of Europe. Only small differences are found in southern Europe (Fig. 10). The 5 % quantile is overestimated in the model compared to observations in parts of Northern Europe while it is underestimated in the mountainous regions (Fig. 11). This is a combination of the differences in standard deviation and skewness.

Comparing the HadGEM3-A historical and histnat experiments we find smaller differences. The standard deviation in the historical version is larger everywhere compared to the histnat version but the differences are small. The 5 % quantile has increased everywhere except for Spain, although the differences are statistically significant only in few regions. The pattern of the changes in the 5 % quantile is largely in agreement with the patterns of the changes in the long term means and the trends in the historical HadGEM3-A model.

The comparison above was done with a single ensemble member. But the described results are robust across the ensemble members and similar results are found for the ensemble mean. For the perturbed surrogate the long term values of standard deviations, skewness, and 5 % quantile are very well represented as expected.

For a good representation of the extremes it is not only necessary that the long term values of the variance and skewness are well represented; also the year-to-year variations of these quantities should be correctly represented. The spatial averages of the winter means of temperature, the variance, and the skewness are shown as a function of the year in Fig. 12 for observations, for a historical HadGEM3-A ensemble member, and for a perturbed surrogate. It is obvious that the observed temporal variability of these quantities are well represented by both the HadGEM3-A and the surrogate. The main deviation is the cold bias in the HadGEM3-A mentioned earlier. The anti-correlation between winter means and variances was also observed in (Yiou et al. 2009).

2) Evaluation of the coldest winter days

Fitting a generalized extreme value (GEV) distribution to the coldest winter days Vautard et al. (2017) found that the historical HadGEM3-A experiments underestimate the location parameter in the mountainous regions. This is in agreement with the results for the 5 % quantile presented in the previous sub-section. The scale parameter is reasonably well represented but in Eastern Europe the model overestimates the shape parameter (too long cold tail). Again, this is in agreement with the results for the skewness shown in the previous sub-section.

Here we use a Kolmogorov-Smirnov test to see if observed and modelled distributions of the temperatures of the coldest winter days are equal. We also show how different forms of bias correction change the results of the test. This is important when choosing the form of correction used when calculating the risk ratios (section 5). The test is applied to each grid-point and for each grid-point the observed sample consists of 53 numbers (one value for each winter) and the modelled sample of 53*15 numbers (as we have 15 ensemble members). As a measure of the overall similarity of the observed and modelled coldest days we use the fraction of grid-points for which we can reject the null-hypothesis of identical distributions at the 5 % level.

For the raw data from the HadGEM3-A historical experiments we can reject the null-hypothesis at the 5 % level in 71 % of the grid-points. The p-values from the test are shown in Fig. S3. For the perturbed surrogate ensembles the corresponding fraction is only 7.5 %, indicating that the cold extremes are well represented by the surrogate approach.

If we perform a bias correction with the difference between the means over all winter days (not just the coldest) a small improvement is seen; now the null-hypothesis is rejected for a smaller fraction, 61 %, of the grid-points. If we also scale with the standard deviations of all winter days (so the observations and model both have same mean and same variance in each grid-point) we get a drastic improvement to 26 %. However, bias correction with the mean of only the coldest winter days brings the fraction of grid-points where we can reject the null-hypothesis down to 5.4 %. Thus some differences in the distributions are particular to the extremes; the differences can not just be described as differences in the mean and standard deviations of winter days.

Fortunately, although the different corrections have different – and in some cases substantial – influence on the distributions themselves we find that for the risk ratios the influence of the corrections are rather small (section 5).

5. The risk ratios

The distributions of the temperatures of the coldest winter days and of the blob index have been calculated for both the HadGEM3-A ensembles (historical and histnat) and the surrogate ensembles (perturbed and unperturbed).

The significance and error bars have been calculated by bootstrapping the values contributing to each distribution. For temperature of the coldest day this amounts to 15*53 values: one value for each winter in each of the 15 ensembles. For the blob index it is 15*53*90 values as we have 90 values each winter. Note that the resulting significance and error bars only include the effects of finite ensemble size.

For the temperatures of the coldest winter days the distributions are calculated for each grid-point. Two examples are shown in Fig. 13; a grid-point near Oslo and a grid-point near Utrecht. These grid-points are typical for mountainous and non-mountainous regions, respectively. Considering first HadGEM3-A, we see that for both locations the distributions for the historical ensemble have moved towards warmer values compared with the histnat ensemble. For the grid-point near Utrecht the modeled distribution and the observations (grey vertical lines) agree well. For this location the risk ratio of the winter 2009-2010 is 0.44 but it should be noted this winter was not extreme at this location. Recall that a risk ratio less than one indicates a reduced probability for an event as extreme as the observed. For the grid-point near Oslo the modeled distribution and the observations do not agree (see discussion of model bias in section 4). The observed winter 2009-2010 (vertical green line) is a cold winter at this location but falls in the middle of the modelled distributions. Correcting the observed temperature for the mean winter bias (orange vertical line) improves the situation significantly. Without the bias correction the risk ratio is 0.44 and with the bias correction it is 0.05. Norway is the region where the bias correction has the largest impact followed by the Alpine region. Outside these areas the effect of the bias correction on the risk ratio is typically less than 0.15. Considering the surrogate method we find as expected that the changes in the modelled distributions are smaller and that the distributions compare well with the observations. Now the risk ratios are 0.71 for both locations.

The geographical distribution of the risk ratios for the coldest winter day is shown in Fig. 14. We see that the probability for a 2009-2010 event has been reduced over almost all of Europe. This holds for both the HadGEM3-A based analysis and the surrogate method although most values are moderate. The HadGEM3-A based analysis in general gives larger changes (and more significant grid-points) than the surrogate method which can be understood from the fact that the histnat ensemble with HadGEM3-A represents pre-industrial conditions while the corresponding unperturbed ensemble with the surrogate method represents the 1960s. The mean risk ratio over Europe is 0.69 for HadGEM3-A. Although, as we saw in section 2, bias correction will influence the distributions themselves it has a smaller effect on the risk ratios outside the mountainous regions. Correcting with the mean of all winter days gives a mean risk ratio of 0.65, while correcting with the mean of the coldest days gives a mean risk ratio of 0.69.

Using only data since 1985 (bottom panel of Fig. 14) we find lower risk ratios for both the HadGEM3-A and the surrogate methods. This should be expected as this period is warmer than the period 1960-1985 in the histnat and perturbed ensembles. However, the lower risk ratios may also partly be due to the smaller number of degrees of freedom in the shorter period (see Appendix A).

The risk ratio of the 2009-2010 event measured with the blob index – which combines the spatial coherence and the intensity of the cold spell – is shown Fig. 15. When the whole period is considered the risk ratio of the 2009-2010 event is not significantly different for either HadGEM3-A or the surrogate method. However, when only data from 1985 are considered the risk ratio is 0.47 (95 % confidence interval is $[0.36, 0.58]$) for HadGEM3-A and 0.65 ($[0.50, 0.82]$) for the surrogate method, and is significantly different from 1 in both cases. Again HadGEM3-A gives larger and more significant changes than the surrogate method. Note that for the largest values of the blob index the 95 % confidence intervals are based on few events and are therefore not robust.

Although the result that risk ratios differ more from 1 when calculated from the period after 1985 than when calculated from the whole period is in agreement with a stronger warming there might also be an effect of the selection problem. In the longer period there is more events to choose from (i.e., it includes more independent degrees of freedom) and the longer period will therefore favor risk ratios closer to 1 (see section 6 and the analytic explanation in appendix A).

6. Conclusions

We have investigated the possibility of attributing the cold European winter 2009-2010 to anthropogenic changes. Two different methods for event attribution have been included: one based on the HadGEM3-A ensembles and one based on the statistical surrogate method described in Christiansen (2015). The surrogate method is based on a simple algorithm to produce ensembles of surrogate fields for both the unperturbed climate and the perturbed climate. These ensembles differ locally by the observed secular low-frequency variability. The method is based on observations and the surrogate fields by construction have the same spatial and temporal structure as the original observed field. The HadGEM3-A ensembles differ in applied forcings, with the histnat ensemble including only natural forcings and the historical ensemble also including the effects of anthropogenic changes. While the histnat HadGEM3-A ensemble represents pre-industrial (1850) conditions the unperturbed surrogate ensemble represents 1960 conditions.

Focusing the evaluation on HadGEM3-A, we found that the trend in winter means over 1960-2013 is in general under-estimated by a factor of 2/3 although there is a considerable spread among the ensemble members. HadGEM3-A also has a mean cold bias dominated by the mountainous regions. The modelled winter standard deviation compares well to observations except for the Norwegian coast and the Alpine region where it is somewhat overestimated. In observations the skewness is negative almost everywhere. The model underestimates the strength of the negative skewness in Scandinavia and many of the western parts of Europe while it overestimates the strength of the negative skewness in central Europe. Together this results in the cold 5 % quantile being overestimated in many regions of Europe except in the mountainous areas. For the extremes – such as the coldest day in winter – we do find some differences between the HadGEM3-A ensemble and the observations. Fortunately, the risk ratios are not sensitive to these deficiencies.

For the attribution we considered two diagnostics; the coldest day in winter for each grid-point and the largest continuous area with temperatures more than two local standard deviations below the mean. The results for the risk ratio were presented using both the whole period 1960-2013 and the later period 1985-2013 to build the distributions. For the largest continuous area no significant change in the risk ratio was found for either the HadGEM3-A model or the surrogate method when the whole period was included. When only the briefer period was included both methods gave statistically significant (different from 1 at the 5 % level) risk ratios for the 2009-2010 event of around 0.5. For the temperature of the coldest day in winter, values less than 1 were found over most of Europe. Lower values were found for HadGEM3-A compared to the surrogate method. Smaller and more significant values were found when only the later period was considered. For this period the HadGEM3-A model and the surrogate method agree on the general pattern with the lowest values in the Western Europe (except the Norwegian coast).

In the perturbed surrogates any low-frequency effect of retreating sea-ice would automatically be included while, as mentioned in section 3a, there are no significant correlations between the Arctic autumn sea-ice and the winter NAO in the HadGEM3-A historical ensemble. The latter observation does not completely rule out an influence of sea-ice on the temperatures in the HadGEM3-A ensemble. However, the fact that we get comparable results about the risk ratios in both the surrogate method and the HadGEM3-A approach suggests that the effect of retreating sea-ice is not very important for the risk ratios.

In appendix A we addressed some issues of attribution of single events. We saw that the counter-intuitive behavior found for the Fractional Attributable Risk (FAR) in Christiansen (2015) also holds for the risk ratio and the simple ratio of probabilities; these measures do not increase monotonically with the strength of the event for heavy tailed distributions. As shown in Vautard et al. (2017) cold extremes might actually have distributions that are difficult to distinguish from heavy tailed distributions (shape parameters of GEV distributions close to 0). Note also that the risk ratios found with the surrogate approach (Fig. 15) do not show a clear decrease with the strength of the event. We also saw that all three measures are sensitive to the 'selection problem'; they depend on the number of degrees of freedom and therefore on the choice of region and period used when counting the events that are similar to the observed extreme event. In agreement with the analytical results we found in section 5 that the risk ratios for the whole period were larger than the risk ratios for the period after 1985. Although some of the explanation can be found in the increased warming in the later period, it further demonstrates that the attribution of single events contains some amount of subjectivity. This point is emphasized by the very low risk ratios found when only the period 2007-2012 is considered (bottom row in Fig. 15). In fact, even lower risk ratios are found when only the winter 2009-2010 is considered (not shown). Finally we saw that the issues described in Christiansen (2015) also exist when the event under consideration becomes less frequent in the changed climate as for the cold events of the present study.

However, we take some comfort in the fact that the two very different methods in general agree on the risk ratio. As mentioned above, the somewhat larger changes found for HadGEM3-A compared to the surrogate approach are because the histnat and the unperturbed ensembles represent different periods. As mentioned in Christiansen (2015) the surrogate method has both advantages and disadvantages, the main advantages being that it is fast and does not require extensive computer resources. The results in the present paper confirm that the surrogate method can be used as an alternative for dynamical methods when considering event attribution. It is also reassuring that the two very different diagnostics in general agree on a reduced risk of cold spells.

Acknowledgments. This work was supported by the EUCLEIA project funded by the European Union's Seventh Framework Programme [FP7/2007-2013] under Grant Agreement No. 607085. The NCEP Reanalysis data were provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). P. Yiou and C. Alvarez-Castro were supported by ERC Grant No. 338965-A2C2.

APPENDIX

Framing issues in attribution of single events

There is an ongoing debate about the interpretation and usefulness of the attribution of single events to climate change (Bindoff et al. 2013; Hansen et al. 2014; Hannart et al. 2015; Otto et al. 2015; Christiansen 2015; National Academies of Sciences, Engineering, and Medicine 2016). In particular, Christiansen (2015) studied the influence of heavy tails and the 'selection problem', i.e., the consequence of the fact that the event under consideration is not independent but selected precisely because it is an extreme. While Christiansen (2015) focused on the Fractional Attributable Risk we here expand the study to include other measures. We will also include the situation where the event under consideration becomes more rare in the changed climate (as expected for cold spells).

The situation and notation are briefly described as follows. For an observation x we denote the probability density in the unperturbed climate as $p^{uc}(x)$ and the cumulative density as $P^{uc}(x)$. In the perturbed climate the corresponding quantities are $p^{pc}(x)$ and $P^{pc}(x)$. Here, the perturbed climate refers to the climate under anthropogenic changes and the unperturbed climate to 'the world that might have been', i.e., the climate without anthropogenic changes. An often used measure of the increased risk for x is the Fractional Attributable Risk (FAR) defined as $(\tilde{P}^{pc}(x) - \tilde{P}^{uc}(x))/\tilde{P}^{pc}(x)$, where $\tilde{P} = 1 - P$ (Allen 2003; Stott et al. 2004, 2013). Here, we assume an event on the right tail of the distribution. Other possible measures are the risk ratio $\tilde{P}^{pc}(x)/\tilde{P}^{uc}(x)$ and the simple ratio of probabilities $p^{pc}(x)/p^{uc}(x)$.

We first assume that climate change amounts to a simple shift $p^{pc}(x) = p^{uc}(x-c)$, $c = 0.3$. This is a reasonable first order approximation as discussed in Christiansen (2015). Also note that in a study of climate-model simulations with future levels of greenhouse gases, de Vries et al. (2012) finds that changes in the frequency of cold spells in Western Europe can be explained by changes in the mean and variance. Under this assumption, Christiansen (2015) showed that while the FAR increases monotonically with x when $p^{uc}(x)$ is Gaussian, this is not the case when $p^{uc}(x)$ has a heavy tail. In this case the FAR has a maximum for a finite value of x. Christiansen (2015) also studied the effect of the 'selection problem' defined above. In this case the relevant probability is not $p^{uc}(x)$ but rather $p_n^{uc}(x_{max})$: the probability density of the largest value, x_{max}, of n variables. Note, that when the n variables are independent and identically distributed we have the identity $P_n = P^n$ for the cumulative densities.

While Christiansen (2015) only considered the FAR, we here show results also for the risk ratio and the simple ratio $p^{pc}(x)/p^{uc}(x)$ (Fig. A1). We see that all three measures behave similarly. Under Gaussianity (left panels) they all increase with x and approach infinity for large x. However, for the distribution with the heavy tail (right panels), they all have a maximum whereafter they decrease. Also note, that for a given x all measures decrease as the number of degrees of freedom increases.

The analysis above assumes that the event under consideration becomes more frequent in the changed climate. For the cold spells analysed in the present paper – and a few previous attribution studies (Christidis et al. 2013, 2014) – the situation is the opposite. The relevant assumption is now $p^{pc}(x) = p^{uc}(x+c)$. Results for this case is shown in Fig. A2. Now the FAR and the two other measures decrease monotonically under Gaussianity while for distributions with heavy tails they reach a minimum for a finite value of x. We also see that all measures increase as the number of degrees of freedom increases.

Thus, the conclusions of Christiansen (2015) based on the FAR also hold for the other measures and when the considered event becomes more infrequent. The 'selection problem' cannot be avoided; all three measures change drastically when the number of degrees of freedom increases. All three measures are sensitive to deviations from Gaussianity; for heavy-tailed distributions the measures do not change monotonically so for the most extreme events the measures reports less changes in the risk than for more intermediate values.

References

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05 109, doi:10.1029/2005JD006290.

Allen, M., 2003: Liability for climate change. *Nature*, **421 (6926)**, 891–892.

Baldwin, M. P., and T. J. Dunkerton, 1999: Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *J. Geophys. Res.*, **104**, 30 937–30 946.

Bindoff, N. L., and Coauthors, 2013: *Climate Change 2013: The Physical Science Basis. [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and*

P.M. Midgley (eds.)], chap. 10 Detection and Attribution of climate change: From global to regional, 867–952. Cambridge University Press, doi:10.1017/CBO9781107415324.022.

Blessing, S., K. Fraedrich, M. Junge, T. Kunz, and F. Lunkeit, 2005: Daily North-Atlantic Oscillation (NAO) index: Statistics and its stratospheric polar vortex dependence. *Meteor. Z.*, **14**, 763–769, doi:10.1127/0941-2948/2005/0085.

Buchan, J., J. J.-M. Hirschi, A. T. Blaker, and B. Sinha, 2014: North Atlantic SST anomalies and the cold North European weather events of winter 2009/10 and December 2010. *Mon. Weather Rev.*, **142**, 922–932, doi:10.1175/MWR-D-13-00104.1.

Buehler, T., C. C. Raible, and T. F. Stocker, 2011: The relationship of winter season North Atlantic blocking frequencies to extreme cold or dry spells in the ERA-40. *Tellus A*, **63**, 212–222, doi:10.1111/j.1600-0870.2010.00492.x.

Burke, C., P. Stott, A. Ciavarella, and Y. Sun, 2016: Attribution of extreme rainfall in southeast China during May 2015. *Bull. Amer. Meteor. Soc.*, **97 (12)**, S92–S96, doi:10.1175/BAMS-D-16-0144.1.

Cattiaux, J., R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron, 2010: Winter 2010 in Europe: A cold extreme in a warming climate. *Geophys. Res. Lett.*, **37**, L20 704, doi:10.1029/2010GL044613.

Christiansen, B., 2001: Downward propagation of zonal mean zonal wind anomalies from the stratosphere to the troposphere: Model and reanalysis. *J. Geophys. Res.*, **106**, 27 307–27 322.

Christiansen, B., 2007: Atmospheric circulation regimes: Can cluster analysis provide the number? *J. Climate*, **20**, 2229–2250.

Christiansen, B., 2013: Changes in temperature records and extremes: Are they statistically significant? *J. Climate*, **26**, 7863–7875, doi:10.1175/JCLI-D-12-00814.1.

Christiansen, B., 2015: The role of the selection problem and non-gaussianity in attribution of single events to climate change. *J. Climate*, **28**, 9873–9891, doi:10.1175/JCLI-D-15-0318.1.

Christidis, N., M. McCarthy, A. Ciavarella, and P. A. Stott, 2016: Human contribution to the record sunshine of winter 2014/15 in the United Kingdom. *Bull. Amer. Meteor. Soc.*, **97 (12)**, S47–S50, doi:10.1175/BAMS-D-16-0143.1.

Christidis, N., P. A. Stott, and A. Ciavarella, 2014: The effect of anthropogenic climate change on the cold spring of 2013 in the United Kingdom. *Bull. Amer. Meteor. Soc.*, **95 (9)**, S79–S82.

Christidis, N., P. A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copsey, J. R. Knight, and W. J. Tennant, 2013: A new HadGEM3-A-Based system for attribution of weather- and climate-related extreme events. *J. Climate*, **26 (9)**, 2756–2783.

Ciavarella, A., and Coauthors, 2017: The HadGEM3-A N216 system for probabilistic attribution of extreme weather and climate events. In preparation.

Cohen, J., J. Foster, M. Barlow, K. Saito, and J. Jones, 2010: Winter 2009–2010: A case study of an extreme Arctic Oscillation event. *Geophys. Res. Lett.*, **37**, L17 707, doi:10.1029/2010GL044256.

Cohen, J., and Coauthors, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627, doi:10.1038/ngeo2234.

Conlon, K. C., N. B. Rajkovich, J. L. White-Newsome, L. Larsen, and M. S. O'Neill, 2011: Preventing cold-related morbidity and mortality in a changing climate. *Maturitas*, **69**, 197–202, doi:10.1016/j.maturitas.2011.04.004.

Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nat. Clim. Change*, **2**, 491–496, doi:10.1038/nclimate1452.

de Vries, H., R. J. Haarsma, and W. Hazeleger, 2012: Western European cold spells in current and future climate. *Geophys. Res. Lett.*, **39**, L04 706, doi:10.1029/2011GL050665.

Doll, C., C. Trinks, N. Sedlacek, V. Pelikan, T. Comes, and F. Schultmann, 2014: Adapting rail and road networks to weather extremes: case studies for southern Germany and Austria. *Natural Hazards*, **72**, 63–85, doi:10.1007/s11069-013-0969-3.

Eden, J. M., K. Wolter, F. E. L. Otto, and G. J. v. Oldenborgh, 2016: Multi-method attribution analysis of extreme precipitation in Boulder, Colorado. *Environ. Res. Lett.*, **11**, 124 009, doi:10.1088/1748-9326/11/12/124009.

Field, C. B., and Coauthors, 2012: Managing the risks of extreme events and disasters to advance climate change adaptation. *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, UK, and New York, NY, USA*.

Fischer, E. M., and R. Knutti, 2015: Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Clim. Change*, **5 (6)**, 560–564, doi:10.1038/nclimate2617.

Francis, J. A., 2017: Why are Arctic linkages to extreme weather still up in the air? *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-17-0006.1.

Francis, J. A., W. Chan, D. J. Leathers, J. R. Miller, and D. E. Veron, 2009: Winter Northern Hemisphere weather patterns remember summer Arctic sea-ice extent. *Geophys. Res. Lett.*, **36**, L07 503, doi:10.1029/2009GL037274.

Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Clim. Res.*, **19 (3)**, 193–212.

Gao, Y., and Coauthors, 2014: Arctic sea ice and Eurasian climate: A review. *Adv. Atmos. Sci.*, **32**, 92–114, doi:10.1007/s00376-014-0009-6.

Gerber, F., J. Sedláček, and R. Knutti, 2014: Influence of the western North Atlantic and the Barents Sea on European winter climate. *Geophys. Res. Lett.*, **41**, 561–567, doi:10.1002/2013GL058778.

Greatbatch, R. J., 2000: The North Atlantic Oscillation. *Stoch. Environ. Res. Risk Assess.*, **14**, 213–242, doi:10.1007/s004770000047.

Hannart, A., J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil, 2015: Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-14-00034.1.

Hansen, G., M. Auffhammer, and A. R. Solow, 2014: On the attribution of a single event to climate change. *J. Climate*, **27 (22)**, 8297–8301, doi:10.1175/JCLI-D-14-00399.1.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20 119, doi:10.1029/2008JD010201.

Jung, T., F. Vitart, L. Ferranti, and J.-J. Morcrette, 2011: Origin and predictability of the extreme negative NAO winter of 2009/10. *Geophys. Res. Lett.*, **38**, L07 701, doi:10.1029/2011GL046786.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.*, **77**, 437–471.

Li, C., B. Stevens, and J. Marotzke, 2015: Eurasian winter cooling in the warming hiatus of 1998-2012. *Geophys. Res. Lett.*, **42**, 2015GL065 327, doi:10.1002/2015GL065327.

Liu, J., J. A. Curry, H. Wang, M. Song, and R. M. Horton, 2012: Impact of declining Arctic sea ice on winter snowfall. *Proc. Natl. Acad. Sci. USA*, **109**, 4074–4079, doi:10.1073/pnas.1114910109.

Meehl, G. A., C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, 2009: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the US. *Geophys. Res. Lett.*, **36**, L23 701.

Mori, M., M. Watanabe, H. Shiogama, J. Inoue, and M. Kimoto, 2014: Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nat. Geosci.*, **7**, 869, doi:10.1038/ngeo2277.

National Academies of Sciences, Engineering, and Medicine, 2016: *Attribution of Extreme Weather Events in the Context of Climate Change*. The National Academies Press, Washington, DC, doi:10.17226/21852.

Orsolini, Y. J., R. Senan, R. E. Benestad, and A. Melsom, 2011: Autumn atmospheric response to the 2007 low Arctic sea ice extent in coupled ocean-atmosphere hindcasts. *Clim. Dyn.*, **38**, 2437–2448, doi:10.1007/s00382-011-1169-z.

Otto, F. E. L., E. Boyd, R. G. Jones, R. J. Cornforth, R. James, H. R. Parker, and M. R. Allen, 2015: Attribution of extreme weather events in Africa: a preliminary exploration of the science and policy implications. *Clim. Change.*, 1–13, doi:10.1007/s10584-015-1432-0.

Ouzeau, G., J. Cattiaux, H. Douville, A. Ribes, and D. Saint-Martin, 2011: European cold winter 2009-2010: How unusual in the instrumental record and how reproducible in the ARPEGE-Climat model? *Geophys. Res. Lett.*, **38**, L11 706, doi:10.1029/2011GL047667.

Overland, J., J. A. Francis, R. Hall, E. Hanna, S.-J. Kim, and T. Vihma, 2015: The melting Arctic and midlatitude weather patterns: Are they connected? *J. Climate*, **28**, 7917–7932, doi:10.1175/JCLI-D-14-00822.1.

Overland, J. E., and M. Wang, 2010: Large-scale atmospheric circulation changes are associated with the recent loss of Arctic sea ice. *Tellus A*, **62**, 1–9, doi:10.1111/j.1600-0870.2009.00421.x.

Peterson, T. C., P. A. Stott, and S. Herring, 2012: Explaining extreme events of 2011 from a climate perspective. *Bull. Amer. Meteor. Soc.*, **93**, 1041–1067, doi:10.1175/BAMS-D-12-00021.1.

Petoukhov, V., and V. A. Semenov, 2010: A link between reduced Barents-Kara sea ice and cold winter extremes over northern continents. *J. Geophys. Res.*, **115**, D21 111, doi:10.1029/2009JD013568.

Prichard, D., and J. Theiler, 1994: Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.*, **73 (7)**, 951–954, doi:10.1103/PhysRevLett.73.951.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108 (D14)**, 4407, doi:10.1029/2002JD002670.

Screen, J. A., 2017: The missing Northern European winter cooling response to Arctic sea ice loss. *Nat. Commun.*, **8**, 14 603, doi:10.1038/ncomms14603.

Sévellec, F., A. V. Fedorov, and W. Liu, 2017: Arctic sea-ice decline weakens the Atlantic Meridional Overturning Circulation. *Nat. Clim. Change*, **7**, 604, doi:10.1038/nclimate3353.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614, doi:10.1038/nature03089.

Stott, P. A., and Coauthors, 2013: Attribution of weather and climate-related events. *Climate Science for Serving Society*, G. R. Asrar, and J. W. Hurrell, Eds., Springer Netherlands, 307–337, URL http://link.springer.com/chapter/10.1007/978-94-007-6692-1_12.

Tang, Q., X. Zhang, X. Yang, and J. A. Francis, 2013: Cold winter extremes in northern continents linked to Arctic sea ice loss. *Environ. Res. Lett.*, **8**, 014 036, doi:10.1088/1748-9326/8/1/014036.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2011: An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, 1992: Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, **58**, 77–94, doi:10.1016/0167-2789(92)90102-S.

Twardosz, R., and U. Kossowska-Cezak, 2016: Exceptionally cold and mild winters in Europe (1951–2010). *Theor. Appl. Climatol.*, **125**, 399–411, doi:10.1007/s00704-015-1524-9.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.

van der Schrier, G., E. J. M. van den Besselaar, A. M. G. Klein Tank, and G. Verver, 2013: Monitoring European average temperature based on the E-OBS gridded data set. *J. Geophys. Res.*, **118**, 5120–5135, doi:10.1002/jgrd.50444.

Van Oldenborgh, G. J., R. Haarsma, H. De Vries, and M. R. Allen, 2014: Cold extremes in North America vs. mild weather in Europe: The winter of 2013–14 in the context of a warming world. *Bull. Amer. Meteor. Soc.*, **96**, 707–714, doi:10.1175/BAMS-D-14-00036.1.

Vargin, P., 2015: Stratospheric polar vortex splitting in December 2009. *Atmos.-Ocean*, **53**, 29–41, doi:10.1080/07055900.2013.851066.

Vautard, R., and Coauthors, 2017: Evaluation of the HadGEM3-A simulations in view of detection and attribution of human influence on extreme events in Europe. *Clim. Dyn., under review*.

Vihma, T., 2014: Effects of Arctic sea ice decline on weather and climate: A review. *Surveys in Geophysics*, **35**, 1175–1214, doi:10.1007/s10712-014-9284-0.

Walters, D., and Coauthors, 2016: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev. Discuss.*, **2016**, 1–52, doi:10.5194/gmd-2016-194.

Wang, C., H. Liu, and S.-K. Lee, 2010: The record-breaking cold temperatures during the winter of 2009/2010 in the Northern Hemisphere. *Atmos. Sci. Lett.*, **11**, 161–168, doi:10.1002/asl.278.

WMO, 2010: Assessment of the observed extreme conditions during the 2009/2010 boreal winter. Tech. Rep. TD-1550, World Meteorological Organization.

Yang, S., and J. H. Christensen, 2012: Arctic sea ice reduction and European cold winters in CMIP5 climate change experiments. *Geophys. Res. Lett.*, **39**, L20 707, doi:10.1029/2012GL053338.

Yiou, P., D. Dacunha-Castelle, S. Parey, and T. T. Huong Hoang, 2009: Statistical representation of temperature mean and variability in Europe. *Geophys. Res. Lett.*, **36**, L04 710, doi:10.1029/2008GL036836.

Zhang, X., C. Lu, and Z. Guan, 2012: Weakened cyclones, intensified anticyclones and recent extreme cold winter weather events in Eurasia. *Environ. Res. Lett.*, **7**, 044 044, doi:10.1088/1748-9326/7/4/044044.

LIST OF FIGURES

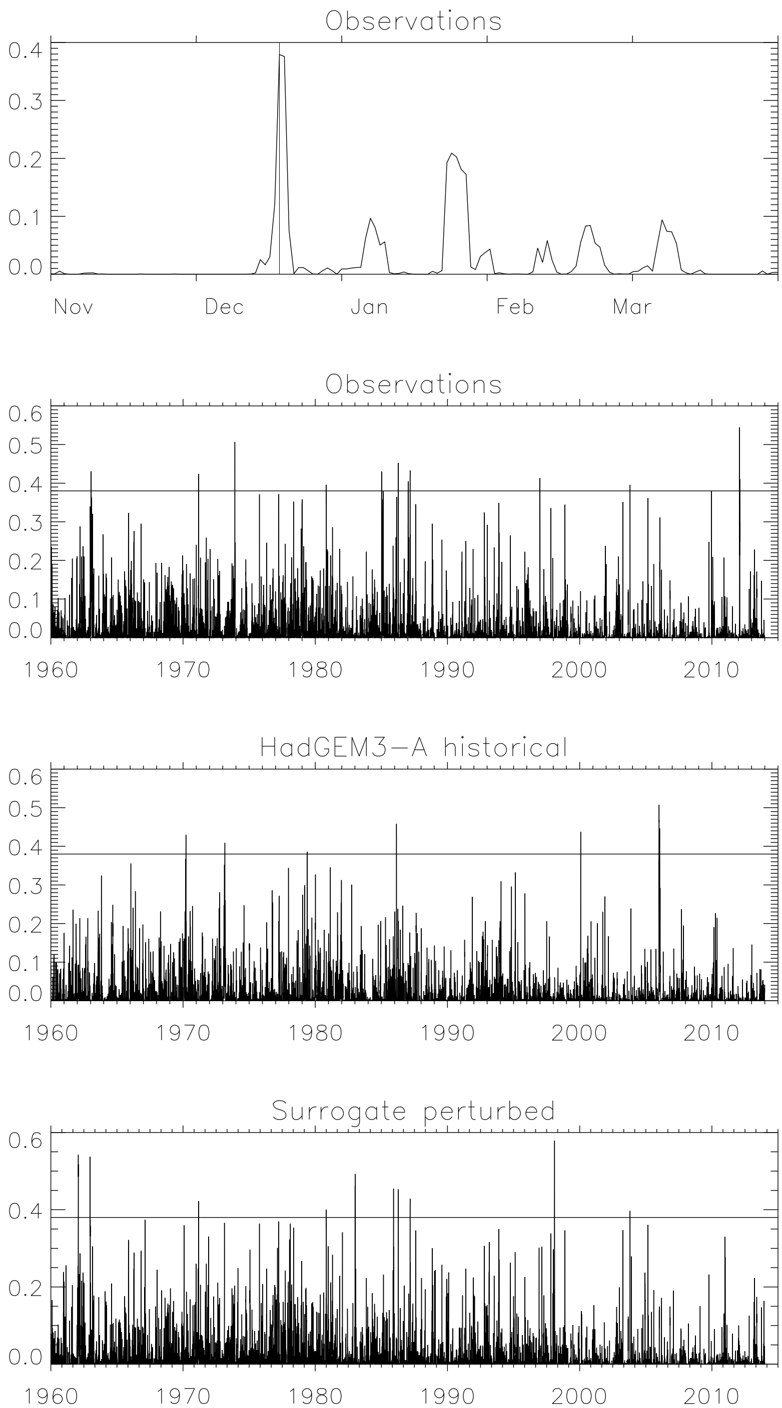


FIG. 1. Blob index as function of time. The blob index is the area of the largest continuous region with temperature anomalies below -2σ, where σ is the local, seasonally varying standard deviation. It is normalized with the total area of the considered region. First panel: Winter 2009-2010 from E-OBS. Second panel: 1960-2013 from E-OBS. Third panel: 1960-2013 from a historical HadGEM3-A ensemble member. Fourth panel: 1960-2013 from a perturbed surrogate ensemble member. Three lowest panels include all year.

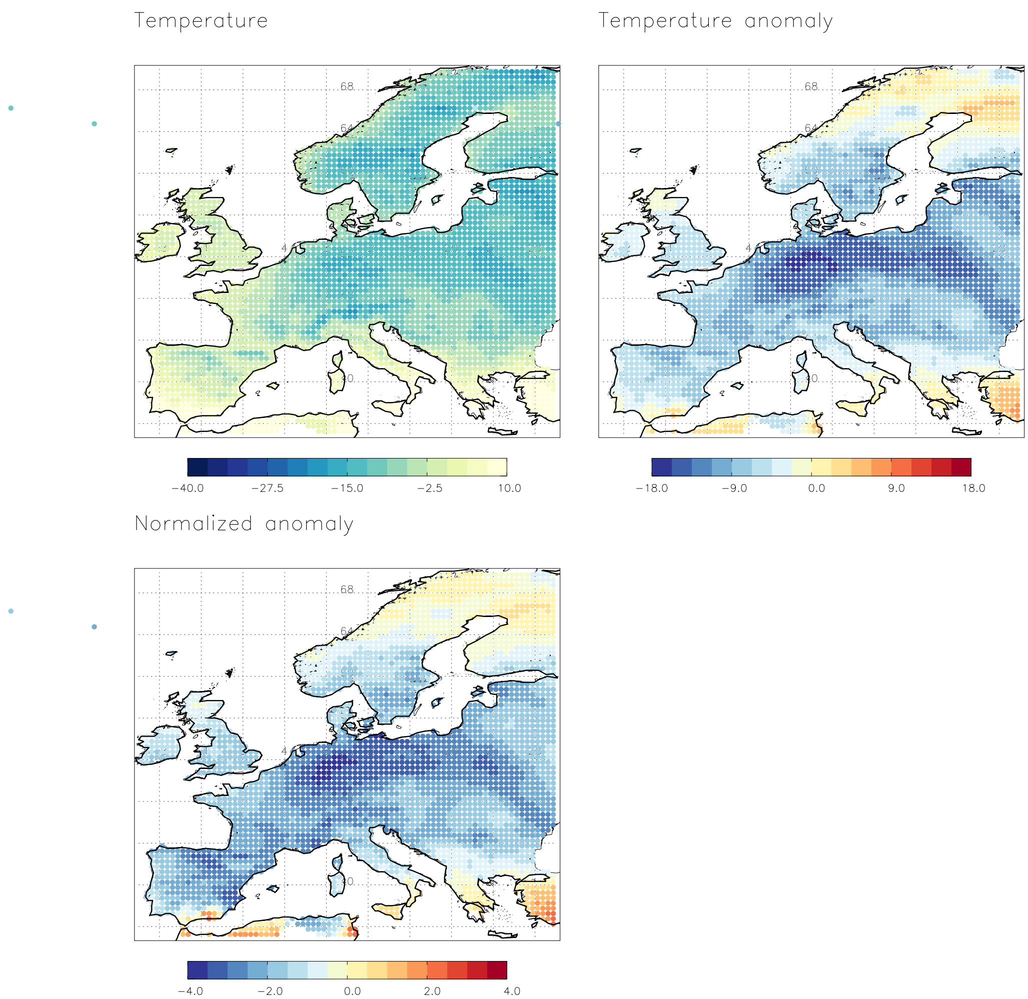


FIG. 2. The temperature of 19th Dec. 2009 which is the winter day of 2009-2010 with the largest blob index. Top left: Temperature [°C]. Top right: Anomaly after removing annual cycle [°C]. Bottom: Anomaly normalized with seasonal standard deviation.

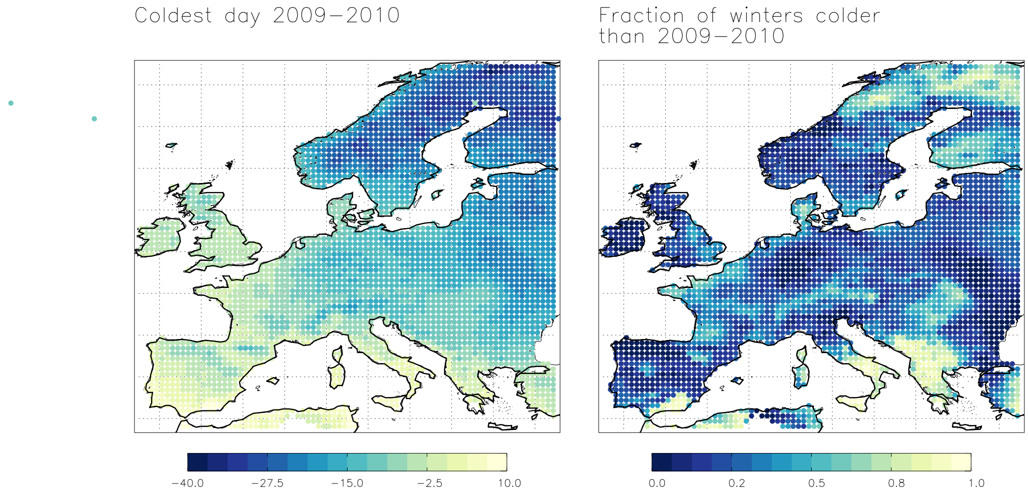


FIG. 3. Left: Temperature [°C] of the coldest day in winter 2009-2010 found individually for each grid-point. Right: Fraction of winters in 1960-2013 with days colder than the coldest day in winter 2009-2010.

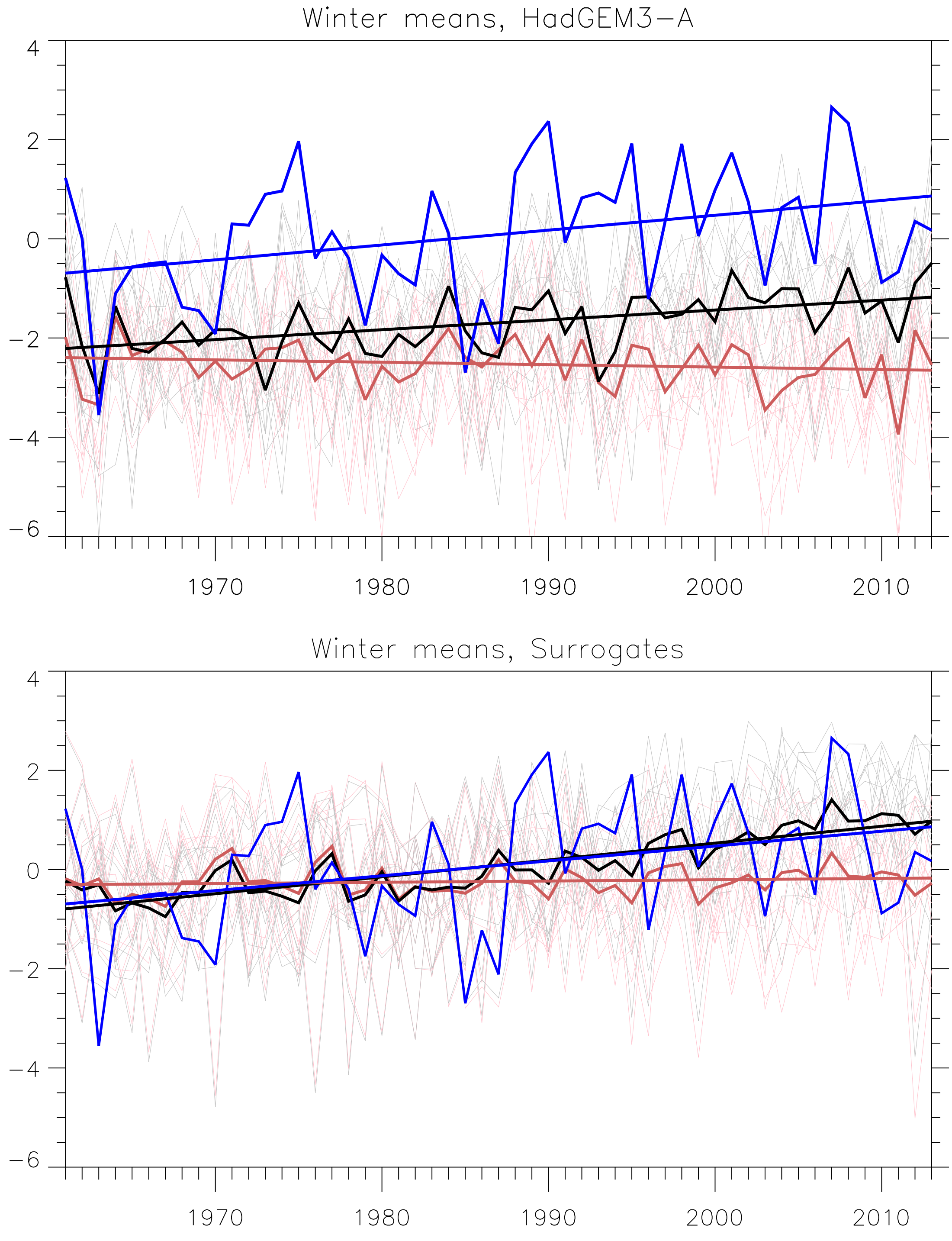


FIG. 4. European mean winter temperatures [°C] as function of time. Top: HadGEM3-A. Bottom: Surrogate. Observations: blue curve. Historical and perturbed ensembles: black curves, Histnat and unperturbed ensembles: red curves. Ensemble means shown with thick curves.

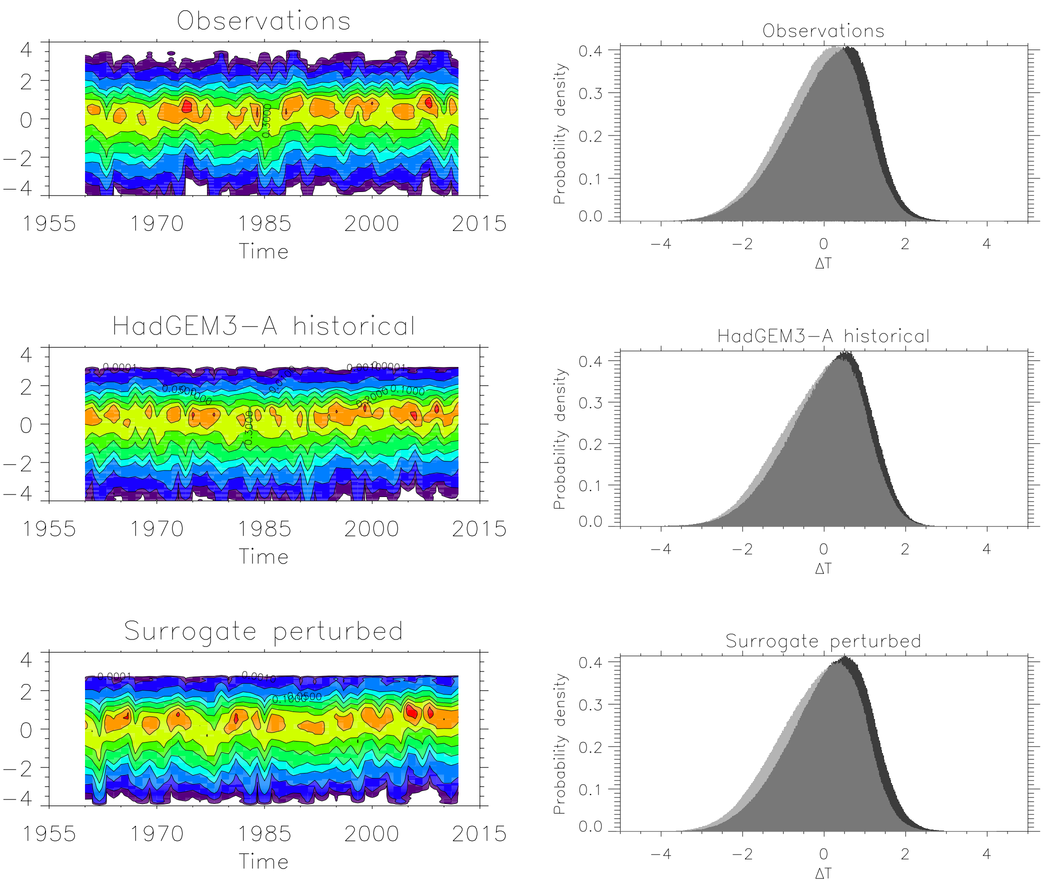


FIG. 5. Local daily winter temperatures have been normalized with their seasonally varying standard deviation and pooled. Left panel shows the distribution as function of time. Contour levels are 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5. Right panel shows the distributions before (light shading) and after (dark shading) 1985. From top: Observations (E-OBS), HadGEM3-A historical, and perturbed surrogate.

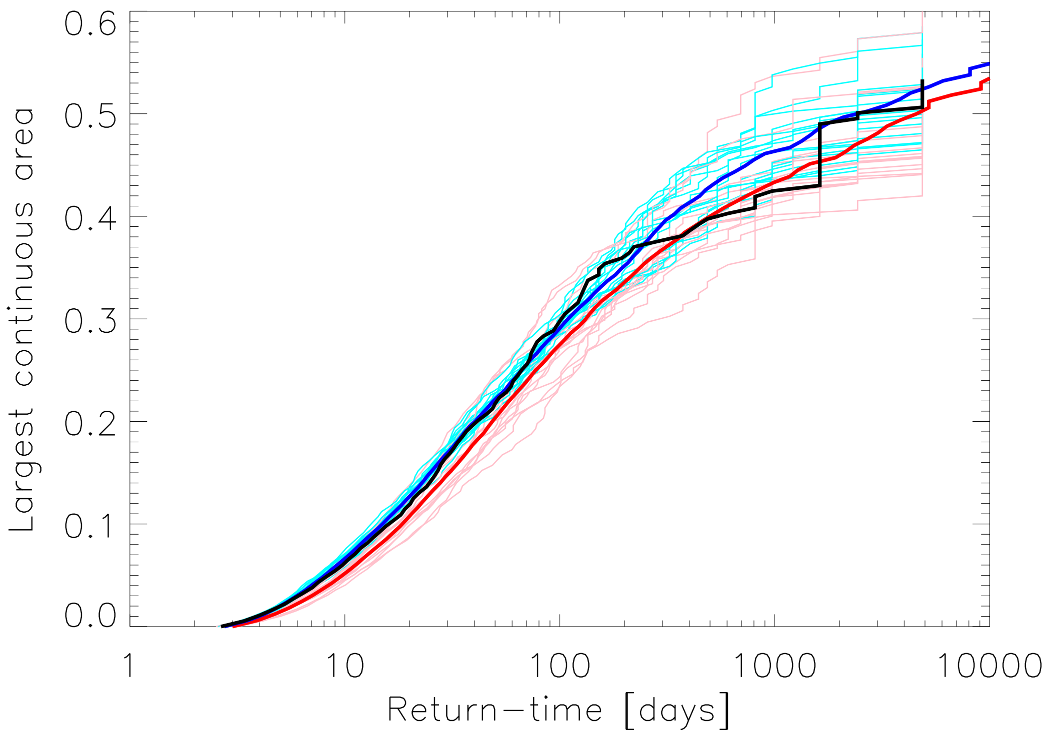


FIG. 6. Return periods of the blob index (largest continuous area) for winter. Black: observations. Blue: surrogates. Red: HadGEM3-A. Thin curves are individual ensemble members, thick curves are pooled ensembles. Only historical and perturbed ensembles shown.

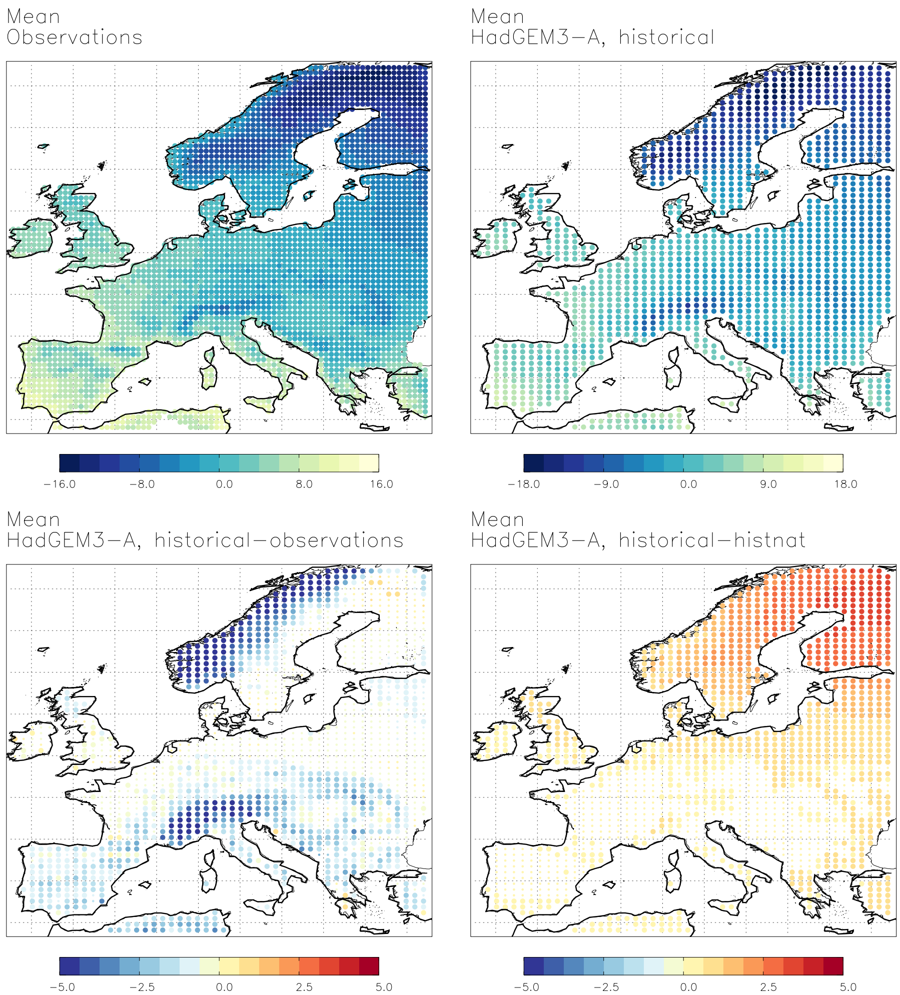
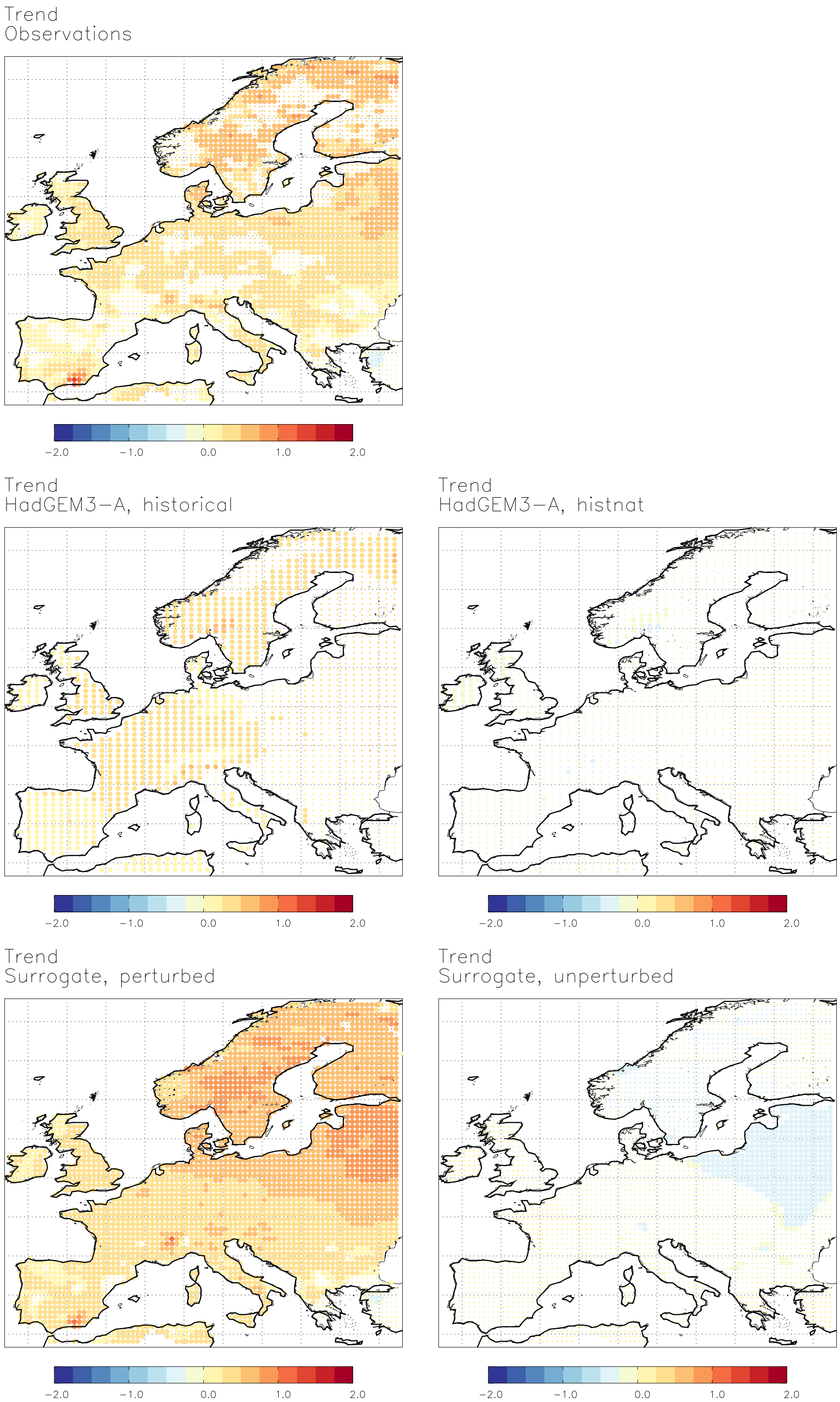
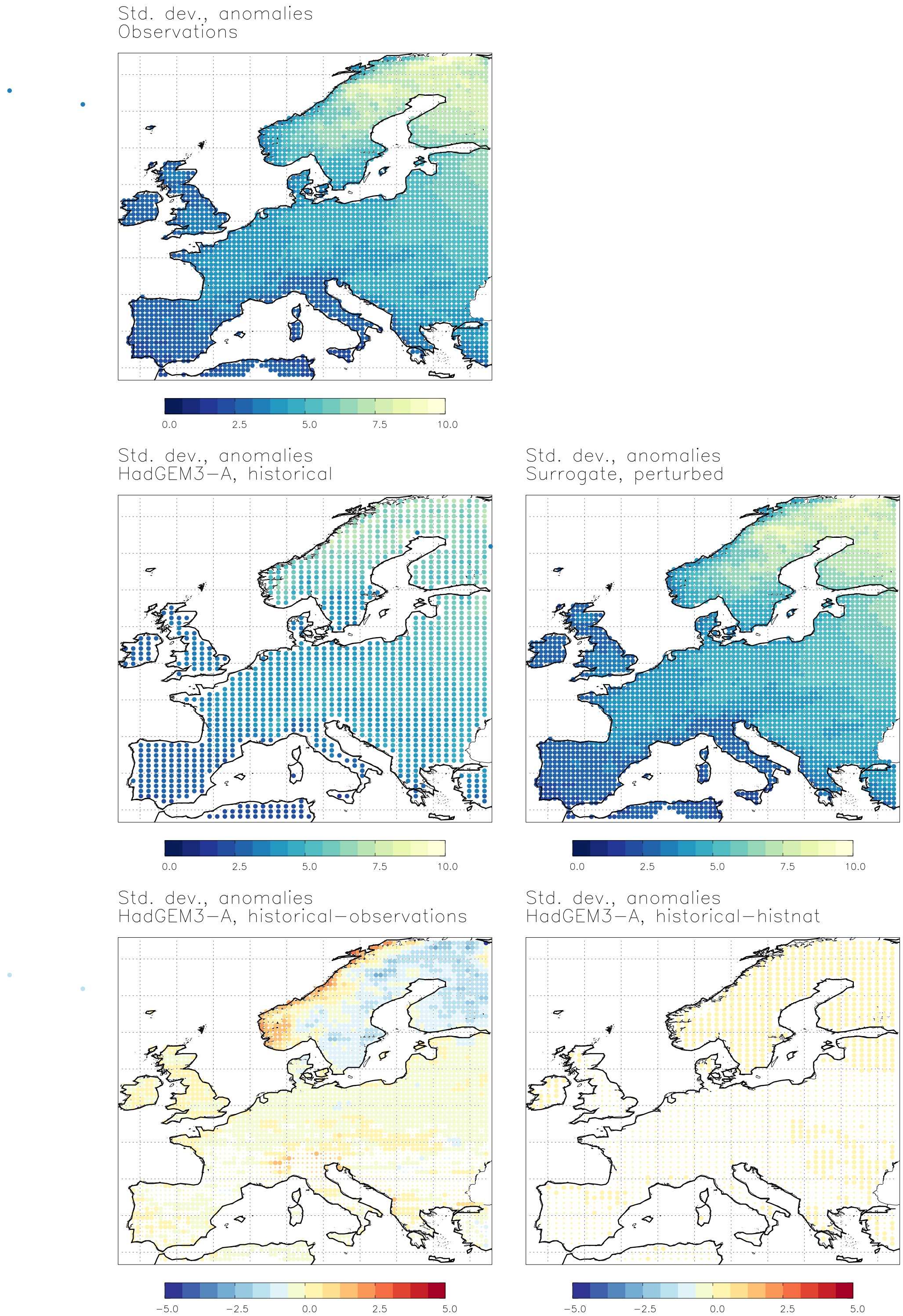


FIG. 7. Top: Long term winter means of grid-point temperatures in observations and historical HadGEM3-A. Bottom left: Model bias. Differences in long term mean between HadGEM3-A historical and observations. Bottom right: Differences in long term mean between HadGEM3-A historical and histnat. Unit: [°C]. Large dots where differences are estimated to be statistically significant at the 5 % level.



862 FIG. 8. Linear trends of grid-point temperatures in winter [°C/decade]. Top: Observations. Middle: Historical
863 and histnat HadGEM3-A. Bottom: Perturbed and unperturbed surrogate. Large dots where trends are estimated
864 to be statistically significant at the 5 % level.



865 FIG. 9. Standard deviation of winter anomalies of grid-point temperatures [°C]. Top: Observations. Middle: Historical HadGEM3-A and perturbed surrogate. Bottom: Difference between historical HadGEM3-A and observations and difference between historical and histnat HadGEM3-A. Large dots where differences are estimated to be statistically significant at the 5 % level.

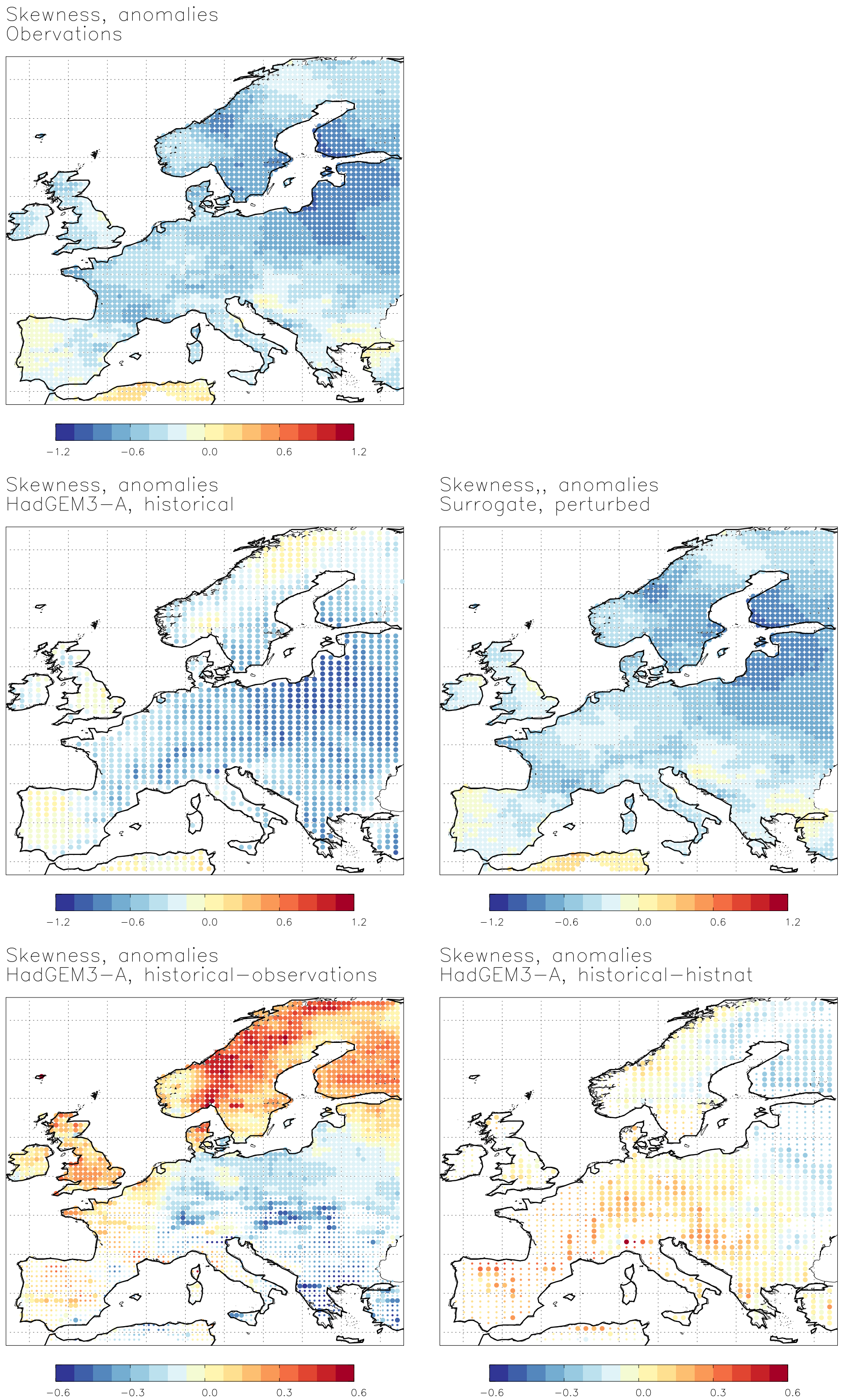


FIG. 10. As Fig. 9 but for skewness of winter anomalies of grid-point temperatures.

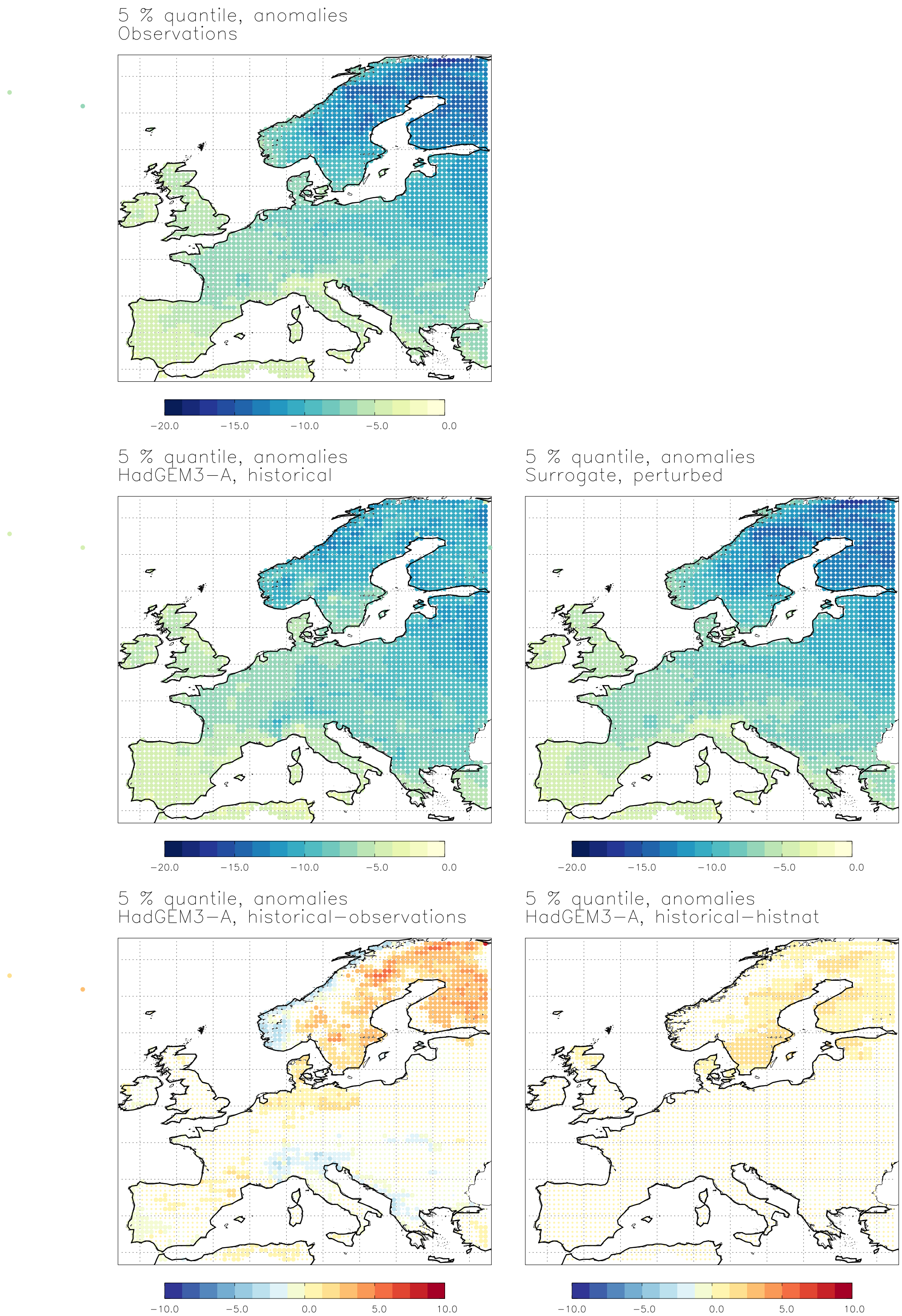


FIG. 11. As Fig. 9 but for the 5 % quantile of winter anomalies of grid-point temperatures [°C].

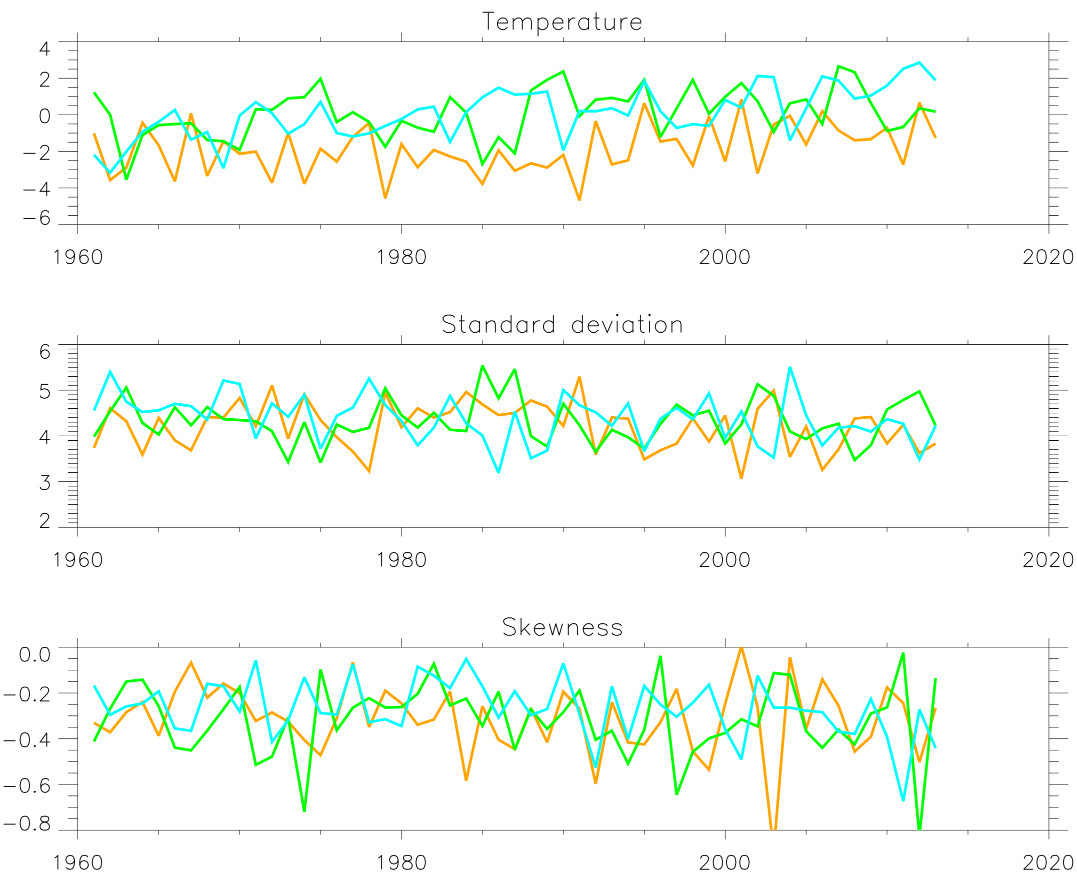


FIG. 12. Time development for spatial mean of temperature [°C], standard deviation [°C], and skewness. Standard deviation and skewness calculated from anomalies. Observations (green), HadGEM3-A historical (orange), surrogate perturbed (cyan).

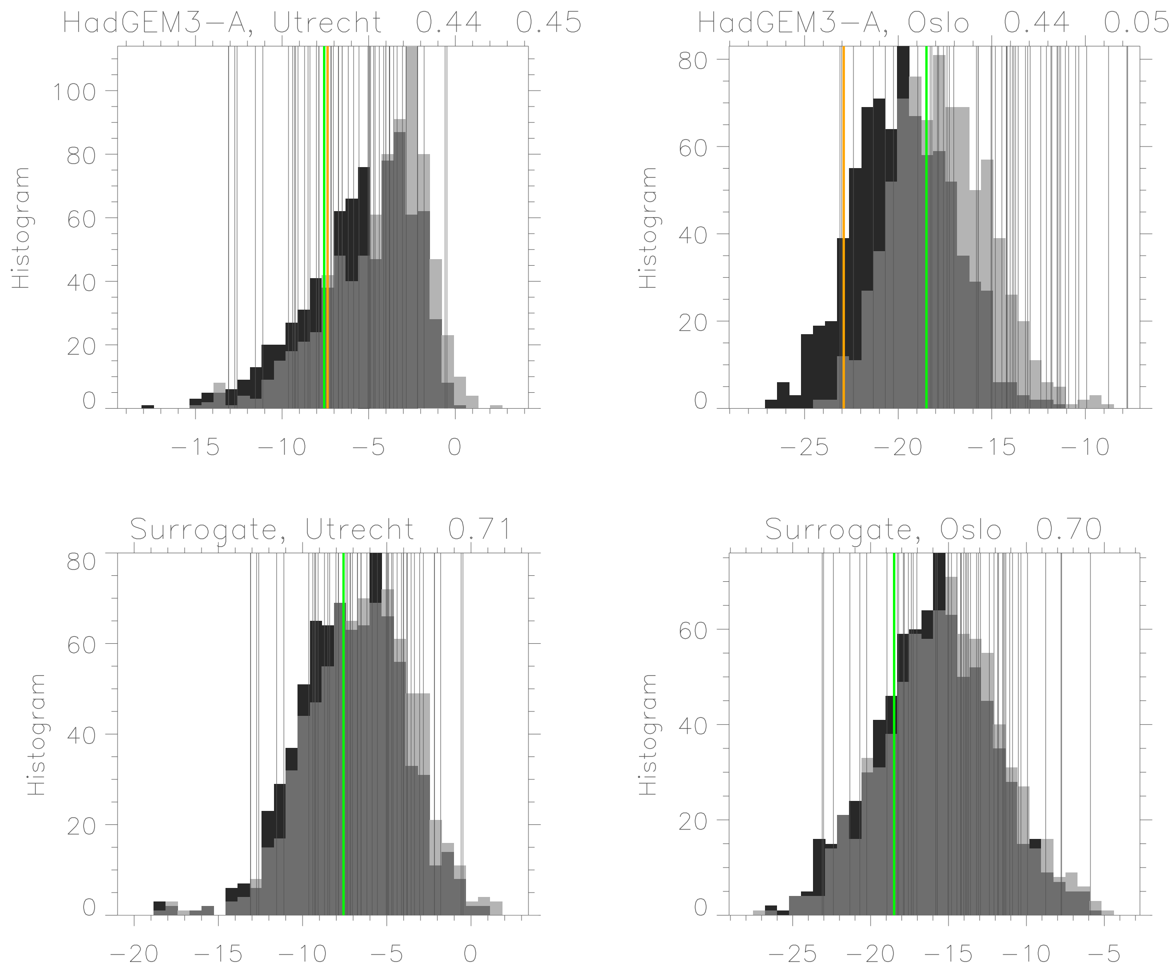


FIG. 13. The distributions of the temperatures [°C] of the coldest day in winter for grid-points near Utrecht and Oslo. Based on 15*53 winters. Historical or perturbed climate: light shading. Histnat or unperturbed climate: dark shading. Thin vertical gray lines are the observed winters. Green vertical line is the observed winter 2009-2010. Orange vertical line is this winter corrected with mean bias. Risk ratios are provided at the top of the panels. For the HadGEM3-A the second number includes bias correction.

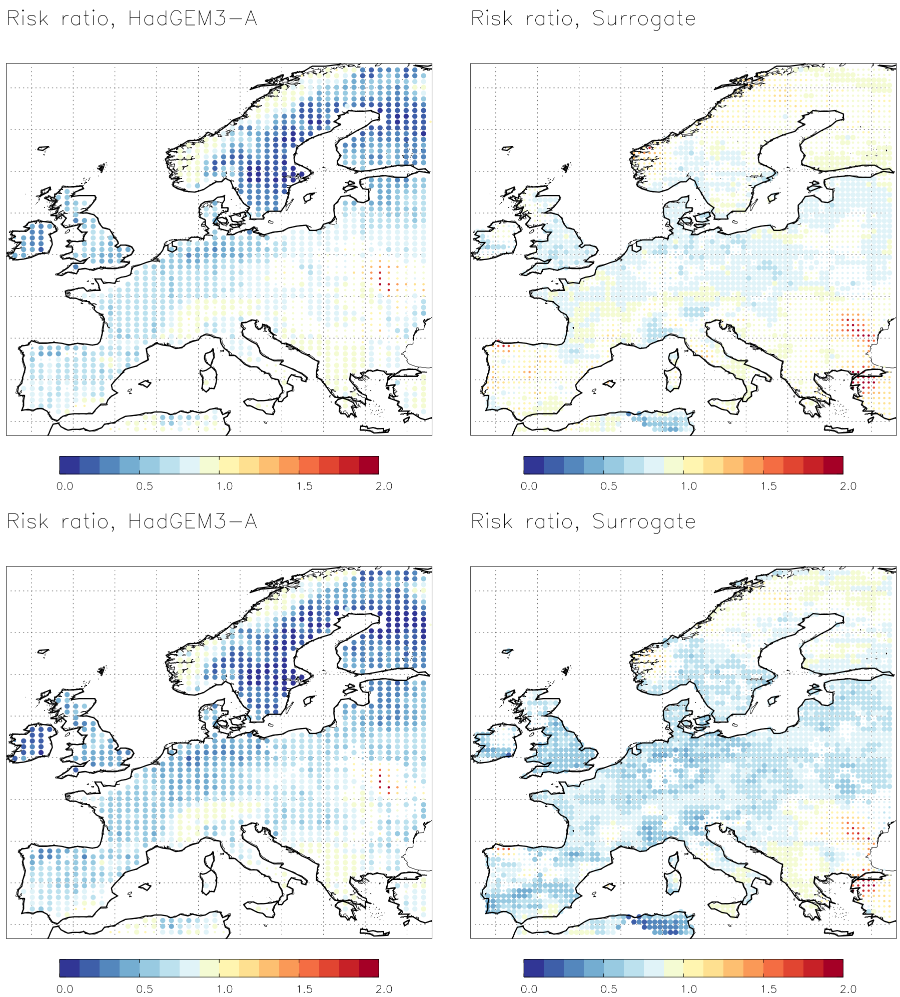


FIG. 14. Maps of the risk ratios of the temperature of the coldest day in the winter 2009-2010. Densities calculated over all winter days. Left: HadGEM3-A. Right: Surrogate method. Top: based on the full period 1960-2013. Bottom: Based on 1985-2013. Large dots where the ratio is estimated to be significantly different from 1 (5 % level).

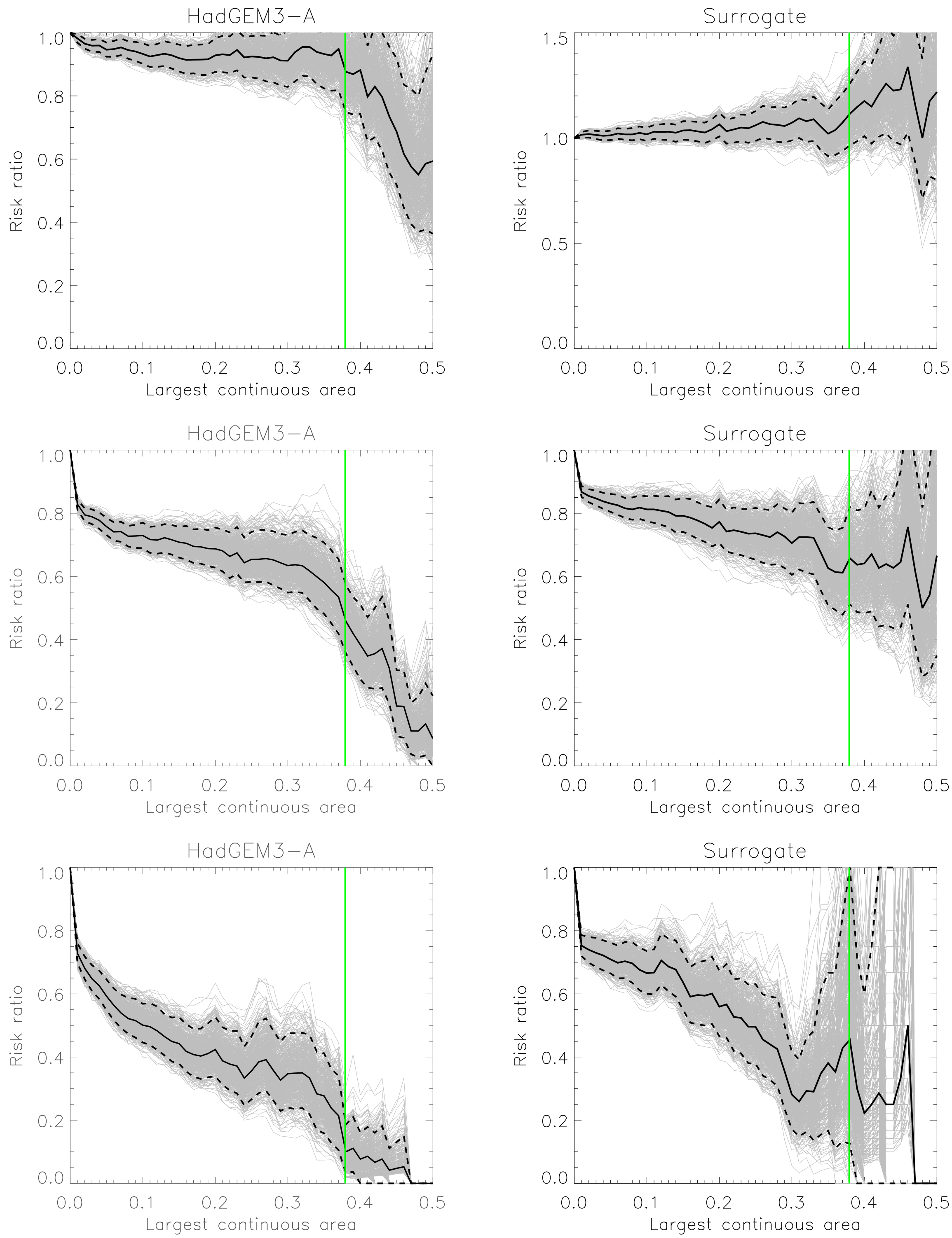


FIG. 15. The risk ratio (thick black curve) for the blob index, i.e., the largest continuous area with temperature anomalies less than -2σ. Vertical green line: observed value for winter 2009-2001. Thin black curves: bootstraps. Black dashed curves: 95 % confidence interval. Left: HadGEM3-A. Right: Surrogate method. Top: based on the full period 1960-2013. Middle: Based on 1985-2013. Bottom: Based on 2007-2012.

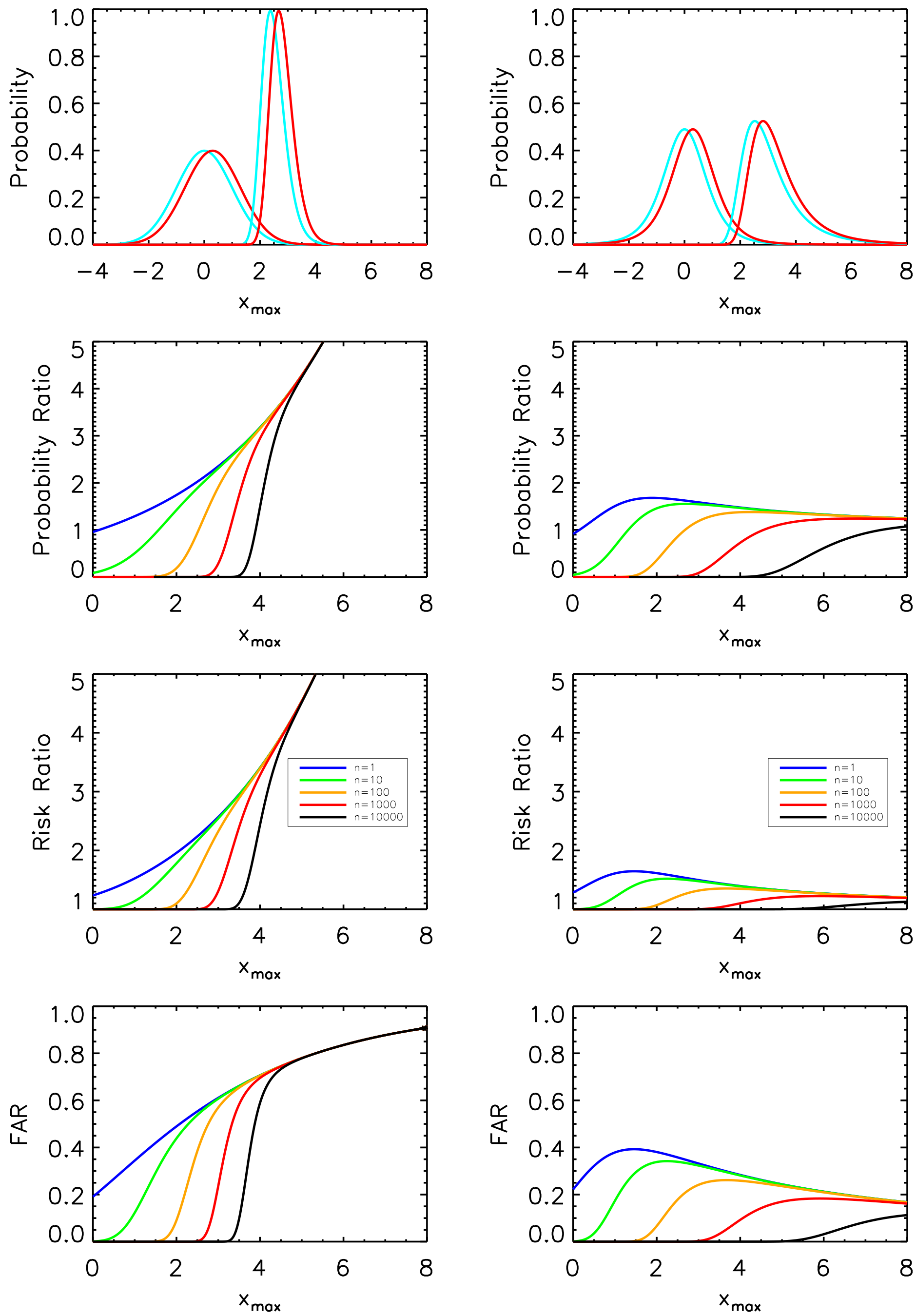


Fig. A1. First row: Probability densities of the largest value, x_{max}, of n independent and identically distributed variables for $n = 1$ and $n = 100$. Cyan: the unperturbed case, p_1^{uc} and p_{100}^{uc}. Red: under climate change, p_1^{pc} and p_{100}^{pc}. The perturbed and unperturbed cases related by $p_1^{pc}(x) = p_1^{uc}(x-c)$, $c = 0.3$. These curves are shown in logarithmic scale in Fig. S4 in the supplement. Second, third and fourth rows: The ratio of probabilities p_n^{pc}/p_n^{uc}, the risk ratios $RR = (1-P_n^{pc})/(1-P_n^{uc})$, and the FARs $\frac{(1-P_n^{pc})-(1-P_n^{uc})}{(1-P_n^{pc})} = 1 - 1/RR$ as function of x_{max}. In left panel p_1^{uc} is Gaussian, in right panel it is t-distributed with 5 degrees of freedom. In each panel are shown results for n=1 (blue), 10 (green), 100 (orange), 1000 (red), 10000 (black).

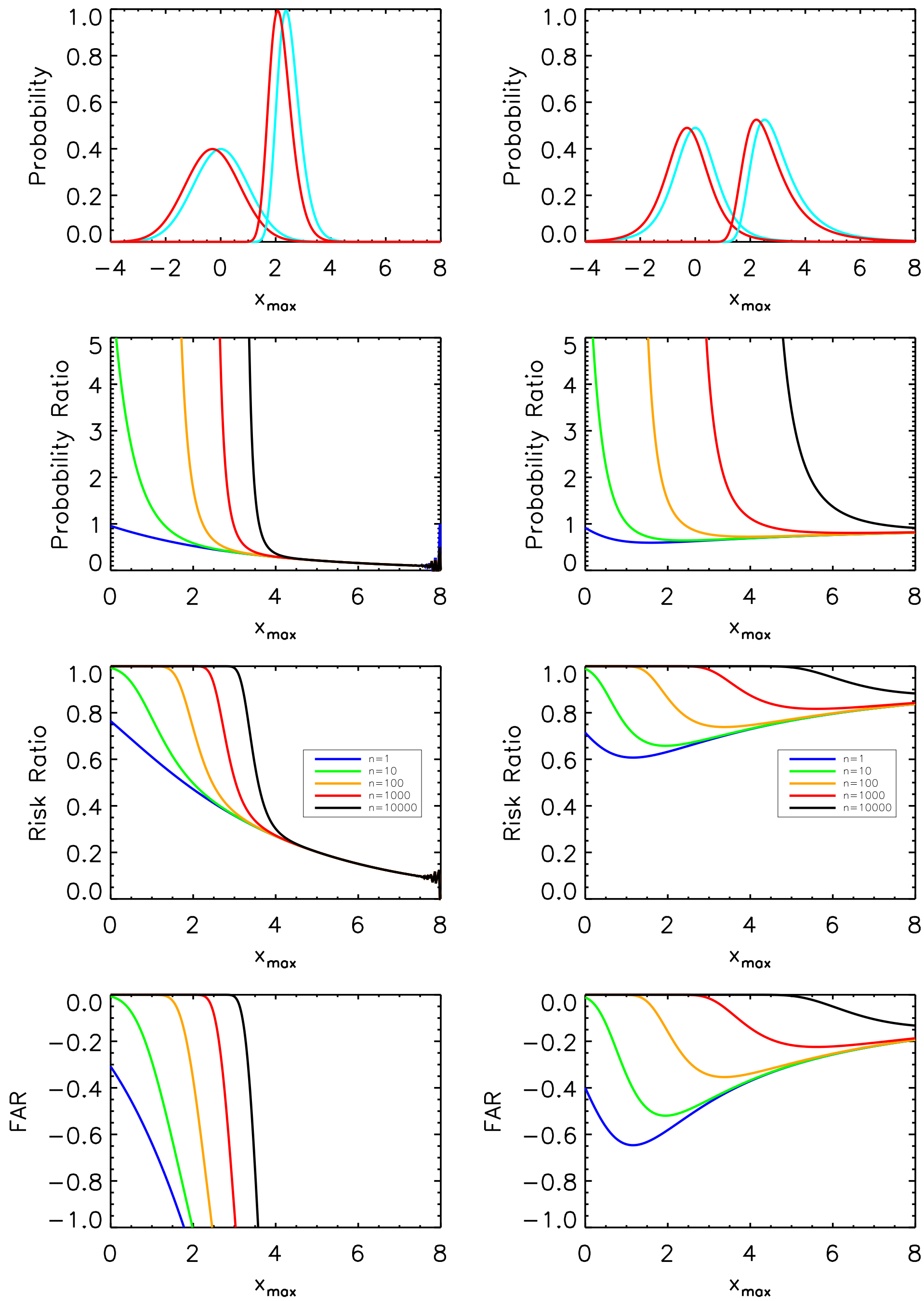


Fig. A2. As for Fig. A1 but with the perturbed climate given by $p_1^{pc}(x) = p_1^{uc}(x+c)$, $c = 0.3$, indicating fewer positive extremes in the perturbed climate.