

# *Can an ensemble climate simulation be used to separate climate change signals from natural variability?*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Bengtsson, L. and Hodges, K. I. ORCID:  
<https://orcid.org/0000-0003-0894-229X> (2019) Can an ensemble climate simulation be used to separate climate change signals from natural variability? *Climate Dynamics*, 52 (5-6). pp. 3553-3573. ISSN 0930-7575 doi: 10.1007/s00382-018-4343-8 Available at <https://centaur.reading.ac.uk/76414/>

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/s00382-018-4343-8>

Publisher: Springer

All outputs in CentAUR are protected by Intellectual Property Rights law, 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](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability?

L. Bengtsson<sup>1</sup> · K. I. Hodges<sup>1</sup> 

Received: 10 July 2017 / Accepted: 30 June 2018  
© The Author(s) 2018

## Abstract

The contribution of internal unforced variability to climate change is explored using a 100-member ensemble climate simulation for the period 1850–2005. The ensemble simulation is based on the Max Planck Institute for Meteorology climate model, ECHAM6, where all members have been exposed to the identically same radiative forcing. The range of global mean surface temperature warming over the 1850–2005 period, based on all members, is 0.65–1.10 °C. The distribution of the global mean surface temperature about the ensemble mean has a standard deviation (StD) of ~0.14 °C which slowly decreases in time. Regionally, the largest decrease in the ensemble StD occurs in the Northern Hemisphere winter. Comparing the temporal StD with that from the observed HadCRUT4 surface temperature data indicates that the majority of the ensemble members have a larger temporal StD than the observations suggesting that the model simulations might overestimate the variance. There are clear random 20-year linear trends in global mean surface temperature anomalies as well as significant regional 50-year linear trends. Even with an ensemble mean warming trend, typical of the early twenty-first century, a global hiatus in temperature of 20 years duration is possible to occur by chance. The results support the view that observed decadal and multi-decadal anomalies in the twentieth century were significantly influenced by internal processes of the climate system. This is particularly the case for the observed global warming trend of 1910–1940 and the global cooling trend of 1940–1970. Global mean precipitation hardly increases with time in the ensemble simulations, but in agreement with theory regional changes occur, with increasing precipitation in polar regions and in some tropical areas. In the subtropics there are reductions in precipitation. Long-lasting regional anomalies of significant amplitudes occur by chance in the ensemble integration.

## 1 Introduction

In many parts of the world the climate varies considerably from year to year and from decade to decade. An important question is whether such fluctuations are due to internal processes in the climate system, including prominent weather events, or whether the causes are external due to changes in radiative forcing, such as those caused by greenhouse gases, aerosols (e.g. from volcanoes) or solar variability. This question is difficult to answer using observational data as historical records only show what has occurred and not what might

have occurred. However, new insights can be gained of the role of stochastic weather processes on climate fluctuations, where the external forcing happens to be exactly known, by investigating large ensembles of coupled climate model integrations for the last 150 years.

Examination of long, multi-centennial climate simulations (Hunt and Elliot 2006; Bengtsson et al. 2006a) suggest that there are considerable variations in regional trends of temperature and precipitation over multi-decadal time periods even if there are no changes in the external forcing, except through the daily and annual cycles. A major challenge is how to identify changes that occur by chance and to what extent the forced modes can be separated from internal modes and unforced processes and whether unforced processes are predictable, at least in a statistical sense.

Improved understanding of the space and temporal variability of the climate system is one of the key objectives of the world climate research program, in particular the CLIVAR programme (ICPO 1998). While global surface temperatures broadly follow the external forcing on long

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00382-018-4343-8>) contains supplementary material, which is available to authorized users.

---

✉ K. I. Hodges  
[k.i.hodges@reading.ac.uk](mailto:k.i.hodges@reading.ac.uk)

<sup>1</sup> Department of Meteorology, University of Reading,  
Reading, UK

time scales (Bengtsson and Schwartz 2013), considerable deviations also occur, such as the warming maximum in the 1930s and 1940s as well as the cooler episode in the 1960s and 1970s (e.g. Bengtsson and Schwartz 2013). In recent years the hiatus in global warming since 1998 has been in the forefront of the climate change debate (Trenberth 2015; Fyfe et al. 2016; von Känél et al. 2017), though this is thought to have ended around 2014, and has been associated with weaker ENSO activity (Hu and Federov 2017) or volcanic activity (Santer et al. 2014). A central scientific question concerns the influence of weather driven processes on long-term climate change. Specifically it is important to determine how long a hiatus or a temporary cooling can last while being simultaneously exposed to an externally forced climate warming.

The Earth's climate system, like all complex systems, is exposed to internal dynamical processes that *a priori* limit its predictability (Poincaré 1892; Lyapunov 1892, 1992; Lorenz 1964). Because of this inherent property, climate like weather can never be accurately predicted. However, in a similar way that ensemble weather prediction can indicate the level of predictive skill at weather time scales, an ensemble climate simulation can help to separate long term climate effects, caused by changes in external forcing (increasing greenhouse gases, solar irradiation or aerosols), from changes that are inherent to the climate system, including chaotic weather events. Whilst climate ensemble prediction has been in use over several decades, more recently the availability of large ensembles with several tens of simulations have recently become an additional tool in climate simulation studies for the study of climate change (Deser et al. 2012a, 2014; Kay et al. 2015; Wallace et al. 2012). These use a coupled climate model run at relatively low resolution in the atmosphere, typically T42 or T63, with model initial conditions varying by either randomly choosing different start times from a control simulation (Deser et al. 2012a) or by perturbing the temperature field by a very small value (Kay et al. 2015). Ensemble sizes are typically not more than 50 in number, even at these low resolutions due to computational limitations. External forcing from greenhouse gases, volcanic aerosols, solar forcing and forcing from anthropogenic aerosols are typically those observed for historical periods or estimated according to IPCC CMIP experimental design. Whilst internal variability and the impact of external forcing can be examined using a single model ensemble it is also important to assess internal variability from different model ensembles to determine the robustness of the results. The study of Deser et al. (2014) compared two ensembles from two different models over North America, the CCSM3 with 40 members and the ECHAM5 with 17 members for the period 2000–2060 and found that “precipitation trends are particularly subject to uncertainty as a result of internal variability” and that “intrinsic atmospheric circulation

variability is mainly responsible for the spread in future climate trends”. However, as with single model or multi-model studies it is important to compare results from ensemble studies with observations, where possible, to evaluate the realism of the estimates of internal variability and trends, at the same time being aware of observational uncertainty. For example, McKinnon et al. (2017) found the ensemble used by Kay et al. (2015) overestimated the uncertainty in trends due to internal variability for surface temperature over North America using resampling methods applied to observations to quantify uncertainty. Thompson et al. (2015) point out the difficulties in trying to estimate the internal variability of the climate from models and observations.

Hence in this paper we provide a new perspective of the role of internal climate variability and trends, by examining an experiment carried out by the Max Planck Institute for Meteorology (MPIMET) where 100 integrations have been undertaken with the MPIMET coupled climate model (Giorgetta et al. 2013) which includes observed and estimated forcing, analogous to the IPCC CMIP5 experimental design, such as greenhouse gases, volcanic aerosols, solar forcing and forcing from anthropogenic aerosols. This is a larger ensemble than used in previous similar studies and uses a different model. Also, we examine both the global and regional variability and trends of both temperature and precipitation over the historical period, as opposed to several other similar studies that have focussed primarily on North America with some using projections to attempt to understand the role of climate change and internal unforced variability in climate trends over the past and future 50 years (Deser et al. 2012b, 2014, 2016; Sigmond and Fyfe 2016). Some other studies have taken a more global view (Deser et al. 2012a; Kay et al. 2015; Hawkins et al. 2016; Dai and Bloecker 2018). In terms of regions we have focussed on the European region and the Arctic. We have also considered temperature trends in the mid-troposphere and contrasted these with those at the surface. Integrations have been performed between 1850 and 2005 with the aim of assessing the value of an ensemble experiment in determining to what extent a climate change signal can be separated from internal chaotic variability, but we do not address climate change directly by using future projections as in some studies. The model used in this study represents the observed internal variability sufficiently well (Giorgetta et al. 2013) and it can therefore be assumed that an ensemble of 100 members will be sufficient to separate the forced mode from the internal variability (Machete and Smith 2016). The distribution of trends in temperature and precipitation are compared with the observed trends and those from previous studies.

The Science questions that are addressed in this paper are:

- (i) How well can an ensemble integration be used to separate a climate change signal in temperature and

precipitation from the internal variability of these variables and what are the implications for climate predictability?

- (ii) What is the typical variance of global and regional multi-annual linear temperature trends and what is the likelihood of a long-term temperature hiatus? Were the multi-decadal temperature anomalies in the twentieth century caused by internal processes or are they related to external effects?
- (iii) Can any trend in the standard deviation and extremes of temperature and precipitation be identified during the last 150 years or are basic climate statistical parameters stationary?

The key findings of this study are:

- (i) Internal processes produce significant multi-decadal temperature anomalies and trends that are unrelated to external forcing.
- (ii) The warming trend of 1910–1940 and the cooling trend of 1940–1970 are markedly enhanced by internal processes.
- (iii) The standard deviation relative to the ensemble mean is unchanged in most areas except in the Arctic where a minor decrease is found and in the tropical belt with a corresponding slight increase.

The paper continues in Sect. 2 where we describe the data and the methodology used. In Sect. 3 the results are presented and in the final section the relevance of the study is discussed in terms of understanding climate change.

## 2 Data description and methodology

The ensemble model data used in this study has been produced with the Max-Planck-Institute Earth System Model (MPI-ESM) (Giorgetta et al. 2013). This consists of the ECHAM6 coupled atmosphere, ocean General Circulation Model (GCM) as well as sub-models for land surface processes and marine biogeochemistry. The ECHAM6 model is an updated version of the ECHAM5 model with new physical parameterisations. The ECHAM6 model resolution is T63L47 in the atmosphere and 1.5° horizontal resolution and 40 z-levels in the ocean. The simulation period for the ensemble generation is 1850–2005.

For each of the ensemble simulations the forcing from the well-mixed greenhouse gases increases monotonically from a value of  $\sim 0.3 \text{ W/m}^2$  in 1850 to  $2.5 \text{ W/m}^2$  at the end of the simulation in 2005. Solar variability varies by a regular 11-year cycle of  $\pm 0.3 \text{ W/m}^2$ . The effect from episodic volcanic eruptions is generally limited to at most a few years (Robock 2000; Driscoll et al. 2012). The handling of

anthropogenic aerosols in the model follows that of Kinne et al. (2013). In particular the global anthropogenic aerosol forcing is around zero in 1850 and reaches  $\sim -0.5 \text{ Wm}^{-2}$  at around 2000 (Fig. 24 in Kinne et al. 2013). Stratospheric aerosols from volcanic eruptions are included, represented by their optical properties, as zonal mean distributions dependent on latitude, pressure and time, and spectrally resolved as needed by the model radiation scheme. Only the direct effect of aerosols is included. Full details of the forcing data can be found in (Giorgetta et al. 2013). The ensemble is initialised by randomly selecting initial data from a control integration exposed to the same external forcing. Giorgetta et al. (2013) found that the model had a transient climate response to a doubling of  $\text{CO}_2$  of 2.0 K.

Data used in this study are primarily the monthly mean values for 2 m temperature, 500 hPa temperature and total precipitation (6 h accumulations). The main emphasis is on annual mean values but specific calculations have also been performed for summer (JJA) and winter (DJF). Both global and regional mean values for particular regions are used for the analysis, with area means calculated as area weighted means. For comparison with observations two data sets have been used, HadCRUT4 (Morice et al. 2012), also available for the period 1850–2005 and the Japanese 55 year reanalysis (JRA55) (Kobayashi et al. 2015) available from 1958 to 2016 though only data from the period 1958–2005 are used in this study.

The HadCRUT4 data is a gridded dataset of global historical surface temperatures produced by blending data from land-surface air temperature datasets and sea-surface temperature datasets and standardized to the period 1961–1990. The JRA55 data set is the latest reanalysis from the Japanese Meteorological Agency (JMA) based on their operational model current in December 2009. This combines a diverse range of bias corrected observations with a short range forecast (6 hourly cycle) using 4-dimensional data assimilation based on a spectral semi-Lagrangian atmospheric model at resolution TL369L60.

It should be noted that the observations of HadCRUT4 have problems with missing data, in particular at high latitudes, Africa and the southern hemisphere (Cowtan and Way 2014), which change with time, this will be considered in the analysis performed here.

In order to highlight typical changes in temperature and precipitation the ensemble model data is divided into three periods, each of 50 years duration. A first period, 1856–1905, with external forcing dominated by volcanic eruptions, a second period, 1906–1955, with minor increases in greenhouse gases as well as fewer volcanic eruptions and a third period, 1956–2005 with larger greenhouse gas forcing. We also specifically highlight the period 1910–1940, showing a marked warming trend, and 1940–1970 with a cooling trend.

It should be noted that the empirical comparison with observed data is limited by the fact that the observations only provide one realization of the past climate while the model provides 100 different and equally possible alternatives.

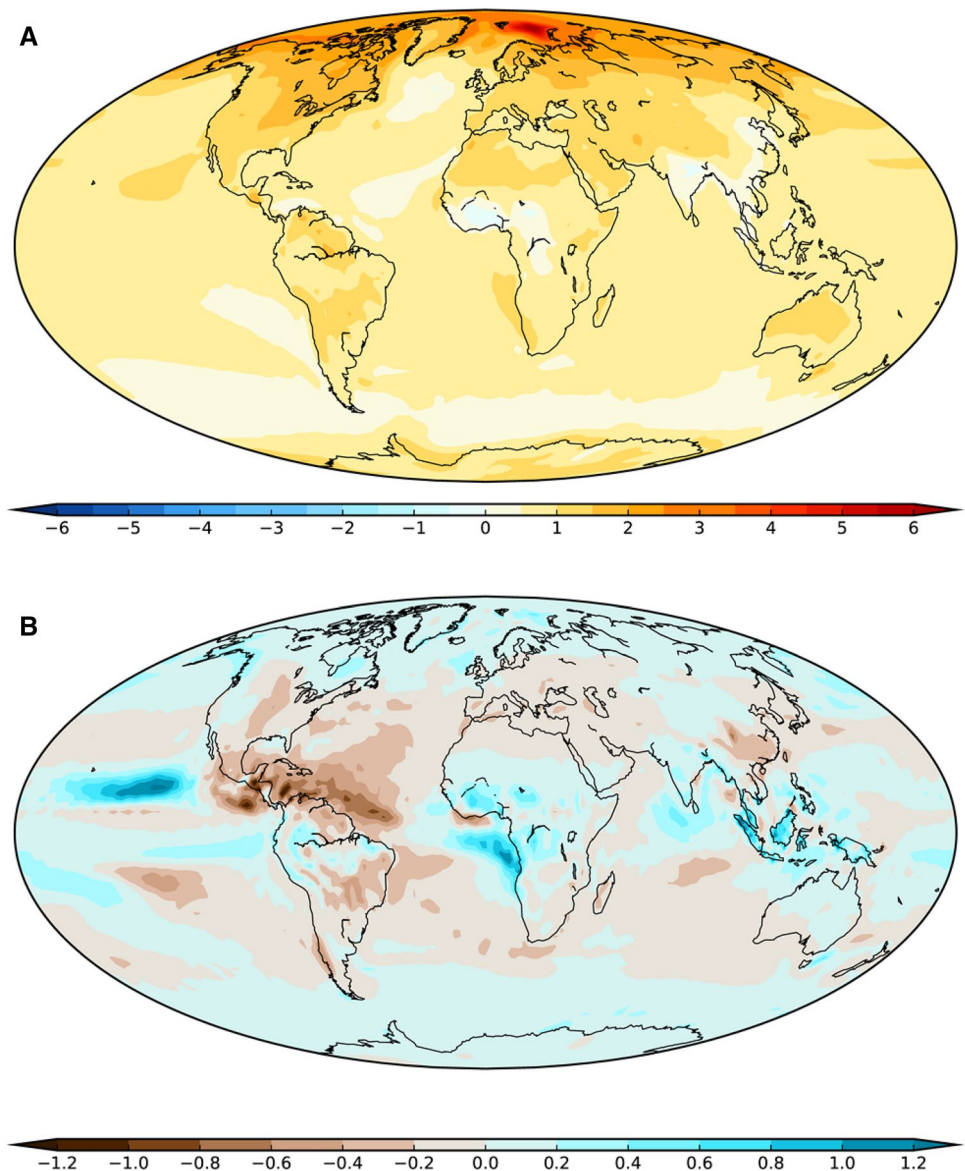
Combinations of standard statistical measures are used to evaluate the model ensemble, e.g. ensemble mean and standard deviation (StD). Linear trends are computed for 20 and 50-year periods for both the raw data and the anomalies. Anomalies are produced by detrending the time series for each ensemble member by fitting and subtracting a B-spline curve, an example of this is shown in Figure S1a. This method is also used for the observations. We have compared this method of detrending with simply subtracting the ensemble mean and they both result in practically identical results (see Figure S1b). Of course the spline method is the only approach that can be used for the single realisation of

the observations. All statistical calculations are performed in the R statistical package (R Core Team 2013), including significance tests where appropriate.

### 3 Results

The emphasis of the study is on the temperature field close to the surface and the temperature in the mid-troposphere at 500 hPa. Additionally, the total precipitation (convective plus large scale) and its relation to temperature is also explored. Figure 1 shows the change in temperature and precipitation for the full period. This has been calculated as the difference between the annual, ensemble mean, average over the first and last 10 years. The global mean change in temperature is  $\sim 0.9^\circ\text{C}$  and close to that observed (Morice

**Fig. 1** **a** Ensemble mean temperature change for the period 1850–2005, unit  $^\circ\text{K}$ . **b** The same for precipitation, unit mm/day. Change computed as average for last 10 years minus average for first 10 years



et al. 2012), with maximum warming in the eastern Barents Sea of  $\sim 6^\circ\text{C}$  which is a consequence of the reduced Arctic sea-ice in the latter period compared with earlier. Using the 100 ensemble members it is found that this warming falls in a range between 0.65 and  $1.10^\circ\text{C}$ . See Figure S2 in the supplementary material.

The difference in the global precipitation between the same two 10 year periods (Fig. 1b) is virtually unchanged ( $+0.01\text{ mm/day}$  or by 0.3%) but with considerable regional changes.

### 3.1 Global temperature

A summary of the near surface (2 m), global, annual mean temperature for the ensemble is shown in Fig. 2a, this has been normalised by the 1961–1990 period similar to HadCRUT4 (Morice et al. 2012). The heavy black line shows the ensemble mean 2 m temperature of the 100 members, the dashed black lines show the  $\pm 1$  StD and the blue lines the individual members. The ensemble mean curve indicates an average warming of around  $0.9^\circ\text{C}$  between 1850 and 2005, similar to the observations from HadCRUT4 (Morice et al. 2012) and JRA55 (Kobayashi et al. 2015) that are also overlaid in Fig. 2a. Figure 2b shows the distribution of linear trends for the ensemble together with those for the ensemble mean, HadCRUT4 and JRA55 for the common period of 1958–2005 covered by all the data sets. All trends are found to have p-values close to zero.

The ensemble StD, shown in Fig. 2c, is  $\sim 0.14^\circ\text{C}$  with an indication of a minor reduction towards the end of the period (see discussion below). Also shown in Fig. 2c is the effect of missing data on the StD (red line), this has been produced by interpolating the ensemble members to the HadCRUT4 grid and masking by the HadCRUT4 data and then re-computing the ensemble mean and StD. This shows that the effect of the missing data is to reduce the StD in the earlier period and that this increases with time eventually converging on the unmasked data in the latter part of the time series as less missing data is present in the later period. The linear trends of the StD's are  $-4 \times 10^{-4}\text{ K/decade}$  for the unmasked StD (p value of 0.03) and for the masked data it is  $1.4 \times 10^{-3}\text{ K/decade}$  (p value of 0.000).

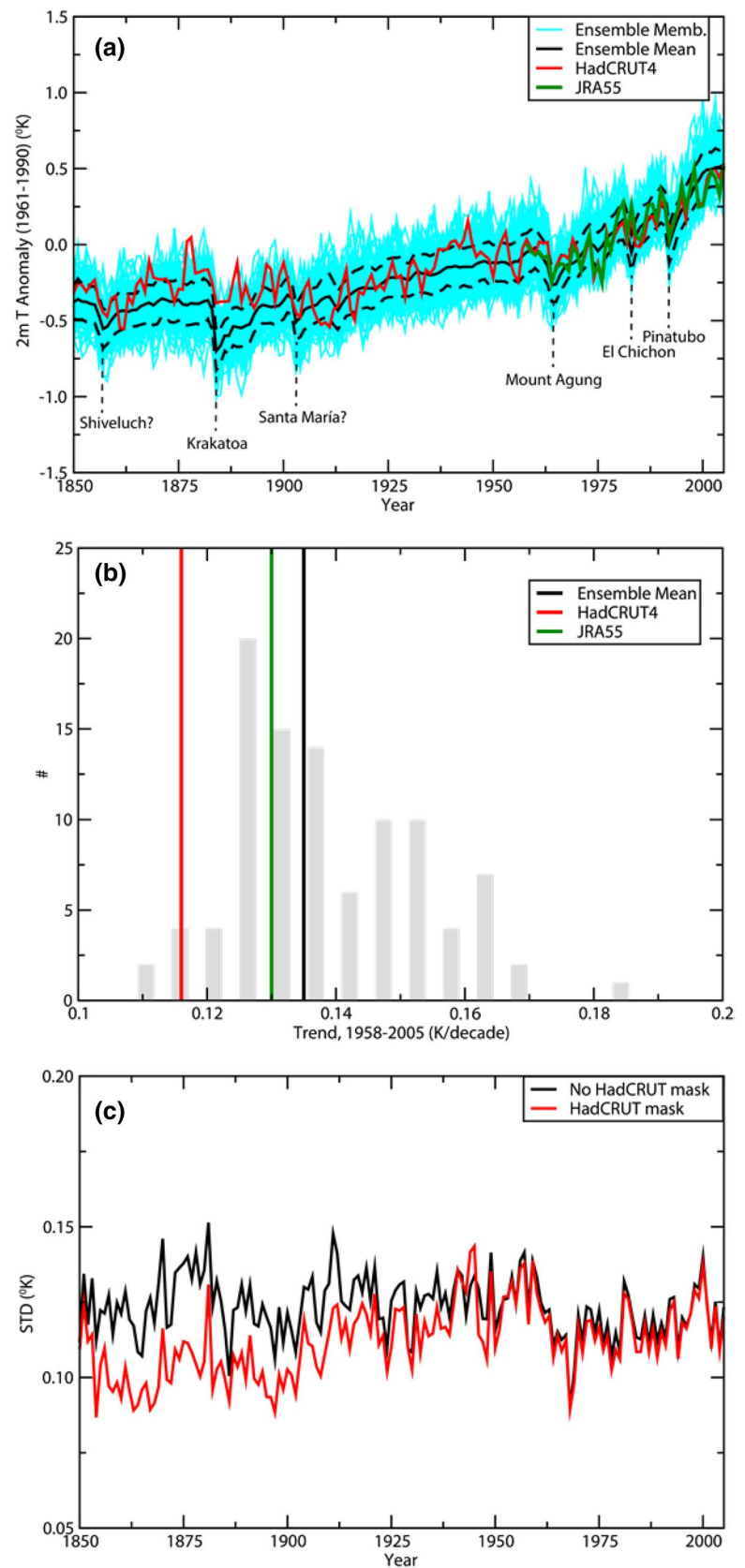
The standard deviation between members of the ensemble shows a decrease with time at middle and high latitudes with the largest decrease at  $70\text{--}90^\circ\text{N}$  in DJF. In other regions the decrease is minor or nil. In the tropical belt of  $10^\circ\text{S}\text{--}10^\circ\text{N}$  there is a slight increase, see Fig S3. The reduced Arctic variance at  $70\text{--}90^\circ\text{N}$  is presumably due to the less colder Arctic and reduced sea ice contributing to a reduced generation of available potential energy. The tropical increase on the other hand may be related to more active tropical systems such as ENSO-events but further research is needed to clarify this.

The difference between the coldest year (in the early part of the ensemble time series) and the warmest year (at the end of the ensemble time series) over all members is about  $2^\circ\text{C}$ . About half this difference is related to the long term warming trend, as represented by the ensemble mean. It is difficult to interpret the ensemble experiment relative to the observational records as the observations only represent one realization while the experiment has 100 possible realisations. An apparent difference is the response to volcanic eruptions, such as Krakatoa in 1883 and Mount Pinatubo in 1991 (see Fig. 2a), which are more pronounced and slightly stronger in the ensemble than in the observations. This is even apparent in the ensemble mean where the 2 m temperature response is larger than for the observations during volcanic events.

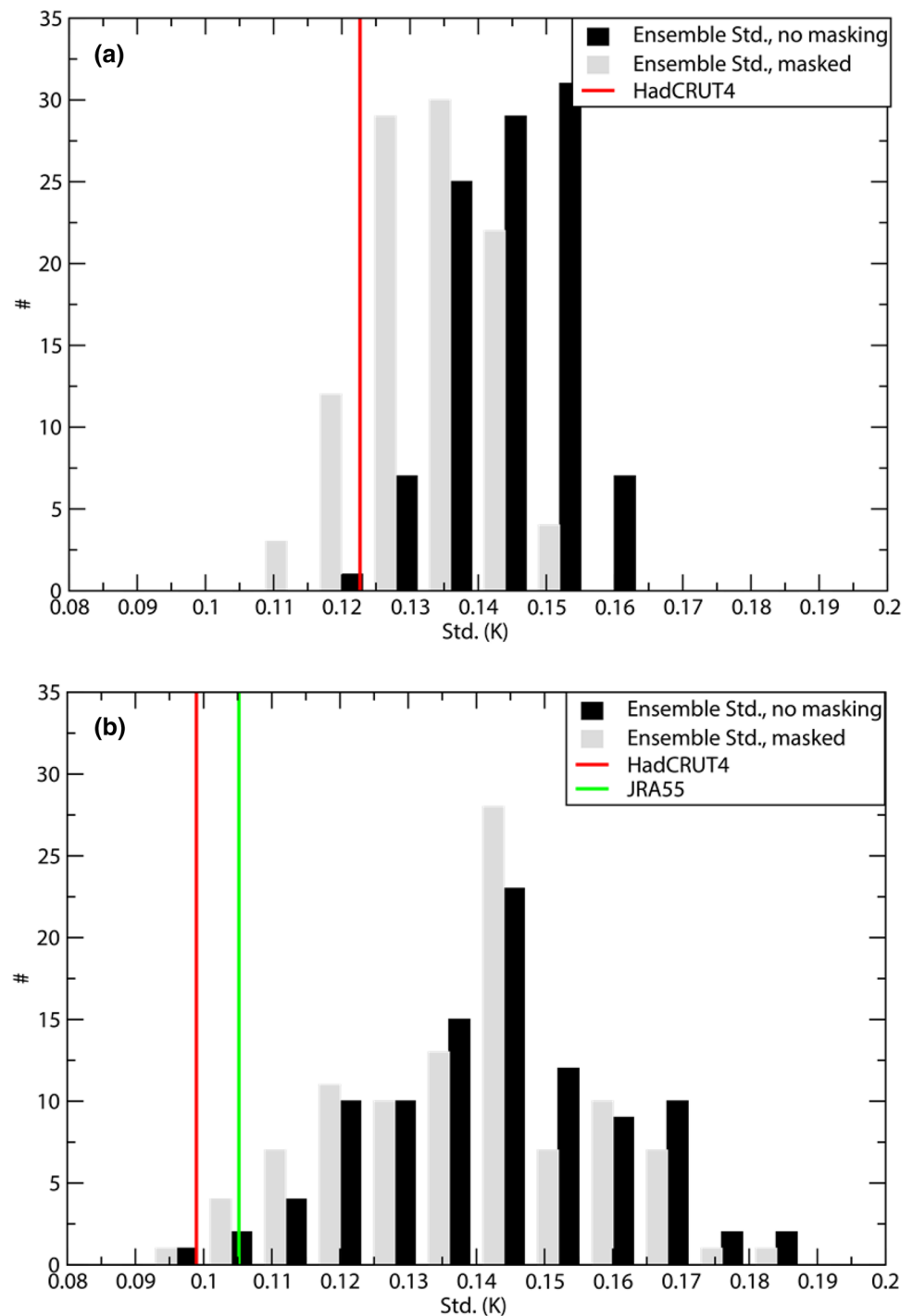
The HadCRUT4 observed 2 m temperature is well within the range of the 100 ensemble simulations but close to the lower end of the ensemble from the 1970s onwards. The same result is seen for the 2 m temperature from JRA55. The two observational data sets, obtained in very different ways, are seen to be more or less identical. Available upper air observations also show that they fall within the ensemble. This is exemplified in the supplementary material (Fig S4) using the 500 hPa temperature data for JRA55 for the last 50 years. Note that the 500 hPa temperature trend is slightly larger than the 2 m temperature trend both for the ensemble and the observation.

The time variance (presented as StD) for each of the 100 members is computed using the detrended time series, using the spline detrending methodology discussed in Sect. 2, in order to compare with the observations. This has been computed as two separate calculations for both the original ensemble members and the ensemble members masked by the HadCRUT4 data. This is shown in Fig. 3a, contrasted with the observed time variance of the observations from HadCRUT4 (similarly detrended). The time StD for the individual 100 members fall between 0.12 and  $0.16^\circ\text{C}$  for the unmasked data and 0.10 and  $0.15^\circ\text{C}$  for the masked data. The observed StD for HadCRUT4 is  $0.12^\circ\text{C}$ . Clearly incomplete data coverage, as implied from the masked ensemble data, has an impact on the StD distribution. However the HadCRUT4 StD sits well inside the distribution for the masked ensemble data but is on the lower edge of the distribution for the unmasked data. To explore this further the calculations have been repeated for the period 1956–2005, a time period when there are more reliable observations including from re-analyses such as JRA55, results are shown in Fig. 3b. Even during this period, with more reliable observations, the StD of practically all the members of the ensemble have a higher StD than both HadCRUT4 and JRA55.

**Fig. 2** **a** Annual global mean 2 m temperature as a function of time for all ensemble members (light blue), ensemble mean and  $\pm 1$  standard deviation (black). Observational results from HadCRUT4 (red) and JRA55 (green) are superimposed. All results are relative to the respective 1961–1990 mean as for HadCRUT4. Major volcanic eruptions are indicated, a ‘?’ indicates uncertainty to attribution. **b** Distribution of ensemble linear trends (gray) and linear trends for the ensemble mean (black), HadCRUT4 (red) and JRA55 (green) for the common 1958–2005 period. **c** Ensemble standard deviation as a function of time for the ensemble (black) and standard deviation when masked by the HadCRUT4 observations (red)



**Fig. 3** **a** Distribution of the annual, global mean, temporal standard deviation for detrended 2 m temperature for each ensemble member for the full period. The black values use all grid points to compute the global means and the grey values use the ensemble data masked by the HadCRUT4 observations for the full period. The standard deviation for the HadCRUT4 is shown as the red line. **b** the same but for the 1956–2005 period, including for JRA55 2 m temperature

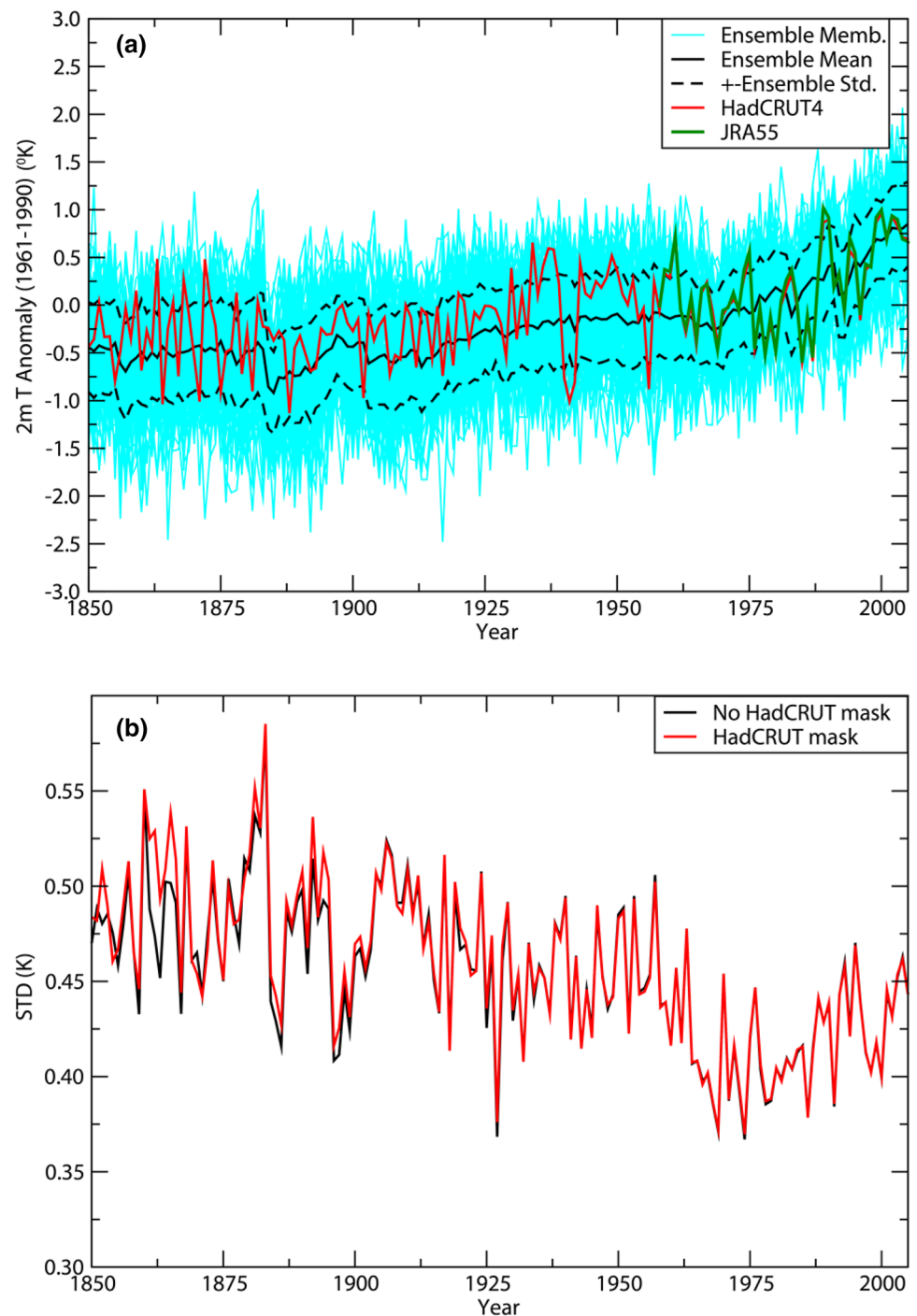


### 3.2 Regional surface temperature

A regional analysis has also been performed, with a particular focus on the area averaged 2 m temperatures for Europe (35°N–70°N, 10W–40E), with the results shown in Fig. 4a for the ensemble time series contrasted with similar results for HadCRUT4 and JRA55. The range of all members encompasses ~4 °C, in terms of the difference between

the coldest and warmest year for the whole period. As will be discussed below the dominant part of the variability is caused by internal differences between members (~3 °C). The ensemble mean warming is ~1 °C. The ensemble StD is shown in Fig. 4b and indicates a range of 0.33–0.55 °K with a distinct downward trend with time. Using the masked ensemble data now shows little difference from using the unmasked data indicating a more consistent data coverage

**Fig. 4** Same as Fig. 2 but for the European region

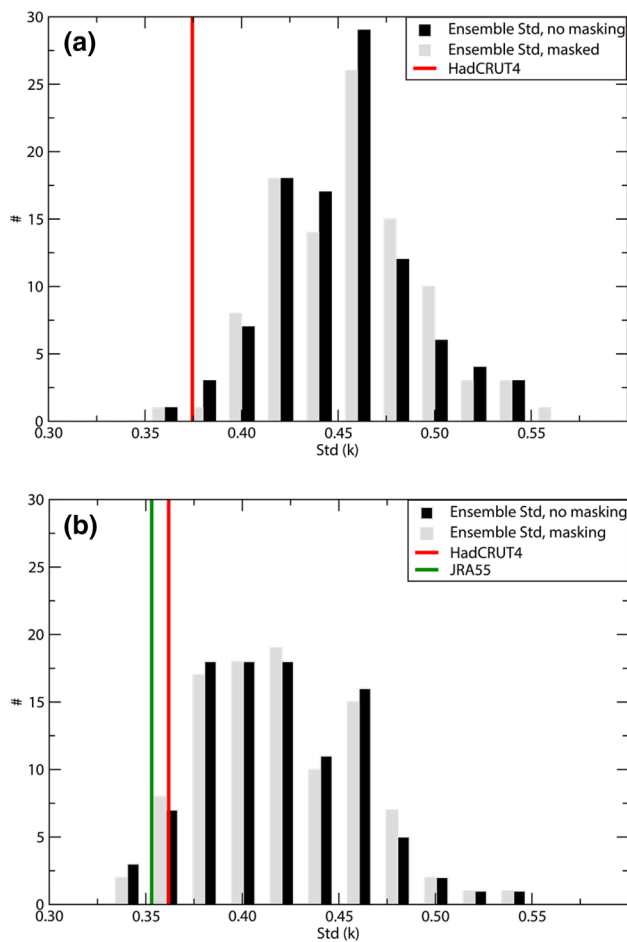


with time over the European region for the whole period of 1850–2005.

In order to compare the observed and the modelled time StD the same calculations as performed for the global data are repeated for the European region for both the full time series and the last 50-years of the time series. The results for the full period are shown in Fig. 5a and show that the distributions of the StDs are much more similar for the masked and unmasked ensemble data than was the case for

the global data reflecting the more consistent data coverage with time over Europe as discussed above. The HadCRUT4 data has a time StD at the low end of the ensemble range. Similar results are obtained for the last 50 years of the time series, shown in Fig. 5b, including for JRA55 which has a StD similar to that for HadCRUT4.

The effects of individual volcanic eruptions are not as clearly seen for Europe, as for the global means, since the natural temperature variability dominates, compared to the



**Fig. 5** Same as Fig. 3 but for the European region

Tropics (not shown). However, inspecting the individual seasons separately it appears that weak volcanic signals in the extra-tropics are mainly a winter phenomenon with the influence of the volcanic eruptions more noticeable in summer. The explanation for this is likely due to the higher level of internal variability of the temperature field in winter, primarily due to the more intense synoptic scale weather systems, and consequently the higher variance during winter. The temperature response to volcanic eruptions varies considerably between ensemble members but a response similar to observations can be seen in the ensemble mean, Fig. 2a.

This is clearly seen in Table 1, which shows the range of temperature trends in Europe for three 50-year periods, 1856–1905, 1906–1955 and 1956–2005, respectively, in terms of the mean trend and the maximum and minimum trends. In addition to the annual values, the values for winter (DJF) and summer (JJA) are also provided, but only in tabular form. Table 1 shows that there is no temperature increase during the first period for Europe, but an accelerating increase can be found thereafter for the other periods. It is interesting to note that the mean warming trend varies

**Table 1** Europe (35–70°N, 10W–40°E), 50-year trends (raw data) for three periods at the start, middle and end of the ensemble simulations for annual, DJF and JJA mean 2 m temperature, values in brackets indicate the ensemble member and p-value respectively. Unit °K/decade

|               | 1856–1905           | 1906–1955           | 1956–2005           |
|---------------|---------------------|---------------------|---------------------|
| <b>Annual</b> |                     |                     |                     |
| Min           | −0.207 (131, 0.025) | −0.104 (142, 0.115) | 0.015 (115, 0.837)  |
| Mean          | 0.001 (0.975)       | 0.102 (0.00)        | 0.191 (0.002)       |
| Max           | 0.241 (186, 0.00)   | 0.296 (184, 0.00)   | 0.407 (191, 0.00)   |
| <b>DJF</b>    |                     |                     |                     |
| Min           | −0.384 (102, 0.00)  | −0.237 (142, 0.041) | −0.145 (144, 0.072) |
| Mean          | 0.000 (0.984)       | 0.117 (0.00)        | 0.199 (0.00)        |
| Max           | 0.359 (186, 0.00)   | 0.454 (184, 0.00)   | 0.504 (152, 0.00)   |
| <b>JJA</b>    |                     |                     |                     |
| Min           | −0.220 (174, 0.00)  | −0.103 (165, 0.009) | 0.041 (169, 0.459)  |
| Mean          | −0.006 (0.899)      | 0.092 (0.00)        | 0.190 (0.002)       |
| Max           | 0.171 (134, 0.011)  | 0.227 (180, 0.00)   | 0.350 (166, 0.00)   |

little by season but the difference between the maximum and minimum trends is more than twice as large in DJF as in JJA. This means that a climate change signal is likely to be more noticeable in summer than in winter. The winter variance is in fact so large that there are some members that show a cooling trend for Europe in the period 1965–2005. Similar results have been found by Wigley and Jones (1981) for example.

Figure 6 depicts the annual trends over the European area and suggests significant variations in the local 50-year trends. Significance at the 5% level is indicated by the stippling where the average trends p-values have been combined using Fisher's method (Fisher 1934). The maximum and minimum trends for the period 1906–1955 show that both significant warming and cooling trends are possible with the largest differences occurring in the Scandinavian-Baltic region with either a cooling over the period of ca.  $-1^{\circ}\text{C}$  or alternatively a warming of some  $+3^{\circ}\text{C}$ . The winter values are about twice as large as the annual values (not shown).

Table 2 shows the same set of results for the Arctic (60°N–90°N). Here the mean warming trend is much larger than for Europe and much larger in winter than in summer. This is the well known Arctic amplification (Pithan and Mauritsen 2014). A contributing factor is presumably the higher ocean temperatures in the northern Norwegian Sea and Barents Sea (Fig. 1a), associated with the receding sea ice, and the effect this has on the atmospheric heat transport into the Arctic (Bengtsson et al. 2011). The variance in the winter Arctic circulation can be seen from the huge difference between maximum and minimum trends in the winter.

### 3.3 On the robustness of temperature trends

An ensemble climate integration with a sufficient number of members is a convenient tool to separate signal (external forcing) from noise (internal variability of the climate system). Irrespective of whether the external forcing used in this experiment is perfectly known or not the result is nevertheless valuable as long as the model is capable of representing the full spectrum of atmospheric processes over a long period of time as well as the interactions with the oceans and the land surfaces. Based on the results presented by Giorgetta et al. (2013) and references therein, we consider the model used here to be a suitable tool for the study we have conducted. The global temperature increase from 1850 to 2005 for the ensemble mean is  $0.9\text{ }^{\circ}\text{C}$  which happens to be close to the observed change while the model could in fact by chance have provided an increase in the range of  $0.65\text{--}1.10\text{ }^{\circ}\text{C}$  if only integrated once.

In order to obtain a more detailed assessment of the linear trends for 20 and 50 year periods, calculations of all possible 20 and 50 year trends have been undertaken for global, European and Arctic mean 2 m temperature fields. These trends are shown for both the raw data and the spline detrended data in Figs. 7 and 8 for 20 and 50 year periods, respectively. Although the results for both the raw and detrended data seem to indicate that the distributions have a Gaussian appearance, for all three regions, performing an Anderson–Darling normality test (Anderson and Darling 1954) (indicated by the p-values in the figures) for each distribution shows that this is in fact not the case. Also shown in the figures are the skew and kurtosis of the distributions which also indicate in general their non-Gaussian nature.

The 20 year trends (Fig. 7) have a broader distribution than those for the 50 year trends (Fig. 8) and the distributions for the raw data have a positive offset whilst the detrended data tend to be centered around zero for all three regions. Figure 9 shows the same diagnostics for the observed data from HadCRUT4, but excluding the separate Arctic region, which is poorly observed over the whole period. HadCRUT4 has at least a similar range of variability for the 20 and 50-year trends which are broadly of the same magnitude as for the model results. This means assuming that the ensemble simulation is realistic in amplitude then sizeable deviations are possible from the ensemble mean. If the ensemble mean has a weak warming trend as in the period prior to 1980 it is quite possible that internal variability might produce large deviations and that a negative trend might develop for several decades.

We have explored this in more detail by assessing the trends for the period 1910–1940, when the observed global temperature shows a considerable warming trend of  $0.13\text{ }^{\circ}\text{C}/\text{decade}$  whereas the ensemble mean has a warming trend of  $0.08\text{ }^{\circ}\text{C}/\text{decade}$ . The full ensemble indicates that any

value between 0 and  $0.19\text{ }^{\circ}\text{C}/\text{decade}$  could have occurred (Fig. 10a).

The trends for the period 1940–1970 are equally interesting (Fig. 10b) as during this period the observed global temperature trend was  $-0.025\text{ }^{\circ}\text{C}/\text{decade}$  while the ensemble mean hardly shows any cooling with the value staying close to  $0\text{ }^{\circ}\text{C}$ . For this period the full range of the ensemble falls between  $-0.08\text{ }^{\circ}\text{C}/\text{decade}$  and  $+0.09\text{ }^{\circ}\text{C}/\text{decade}$ . It therefore seems that assuming the external forcing used in the ensemble experiment was correct, the trend in the period 1940–1970 could equally well have been positive.

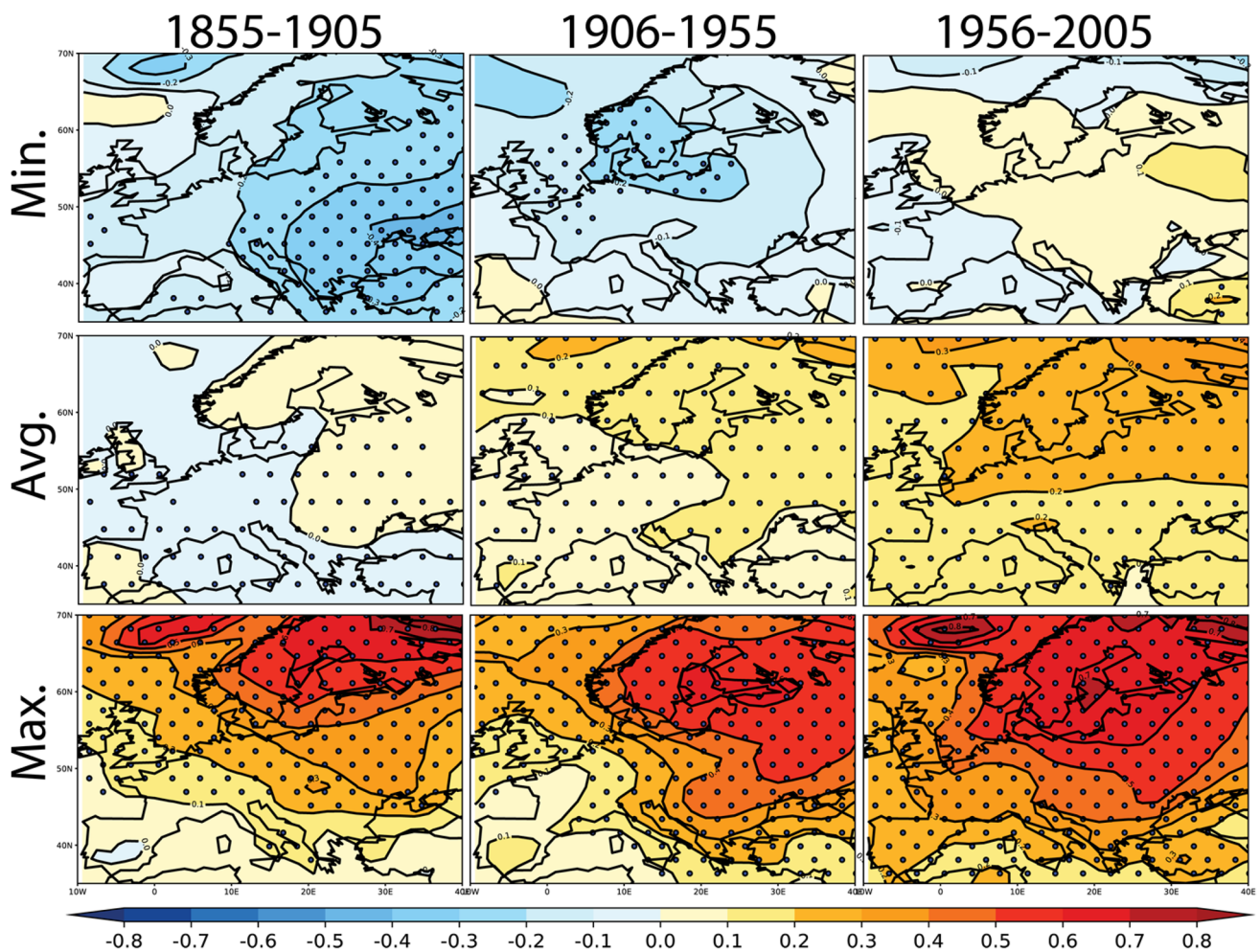
However, it cannot be excluded that the forcing of anthropogenic aerosols could have been larger than applied in this experiment. This might be inferred from the fact that the observed trends over the 1910–1940 period are larger than those over the 1940–1970 period and larger than in the model. Several previous studies have highlighted that the cessation of warming in the mid-twentieth century was likely due to radiative cooling associated with increasing sulphate aerosols (Wilcox et al. 2013; Tett et al. 2002). Also omitted from the model simulations are the poorly quantified long term solar variability (Kopp 2016) which may also have contributed to the observed forced response. Similarly, the internal variability of the model might be overestimated, as indeed is indicated. For this reason we must express caution in the interpretation of the result. Nevertheless, it seems obvious that a single climate experiment is insufficient and might lead to an incorrect interpretation of a modelling experiment. Assuming that changes in forcing are the only explanation for the differences between a model and observations might thus be misleading.

As can be further seen from Figs. 7 and 8 the variance for Europe and the Arctic is significantly larger than for the global mean and therefore as was shown in the studies of Delworth and Knutsson (2000) and Bengtsson et al. (2006a) different ensemble members show significant regional, decadal or multi-decadal trends caused by internal processes unrelated to external forcing.

### 3.4 On precipitation trends

The long-term change in global mean precipitation, is insignificant for the full 1850–2005 period, shown in Fig. 11a. What is clearly obvious is the marked fall in global precipitation coincident with volcanic eruptions. Unfortunately there are no reliable observations to show a similar association. We suggest that the fall in precipitation at the same time as volcanic eruptions is likely a consequence of aerosol emissions affecting the surface energy balance and thus reducing the moisture fluxes (Iles et al. 2013).

For a detailed assessment, precipitation trends have been calculated in a similar way as for the 2 m temperature trends. Here the focus is on the European region with the trends



**Fig. 6** Annual mean, 50-year, temperature (2 m) trends for Europe, (35–70°N, (10W–40°E), for three periods at the start (1855–1905), middle (1906–1955) and end (1956–2005) of the ensemble simula-

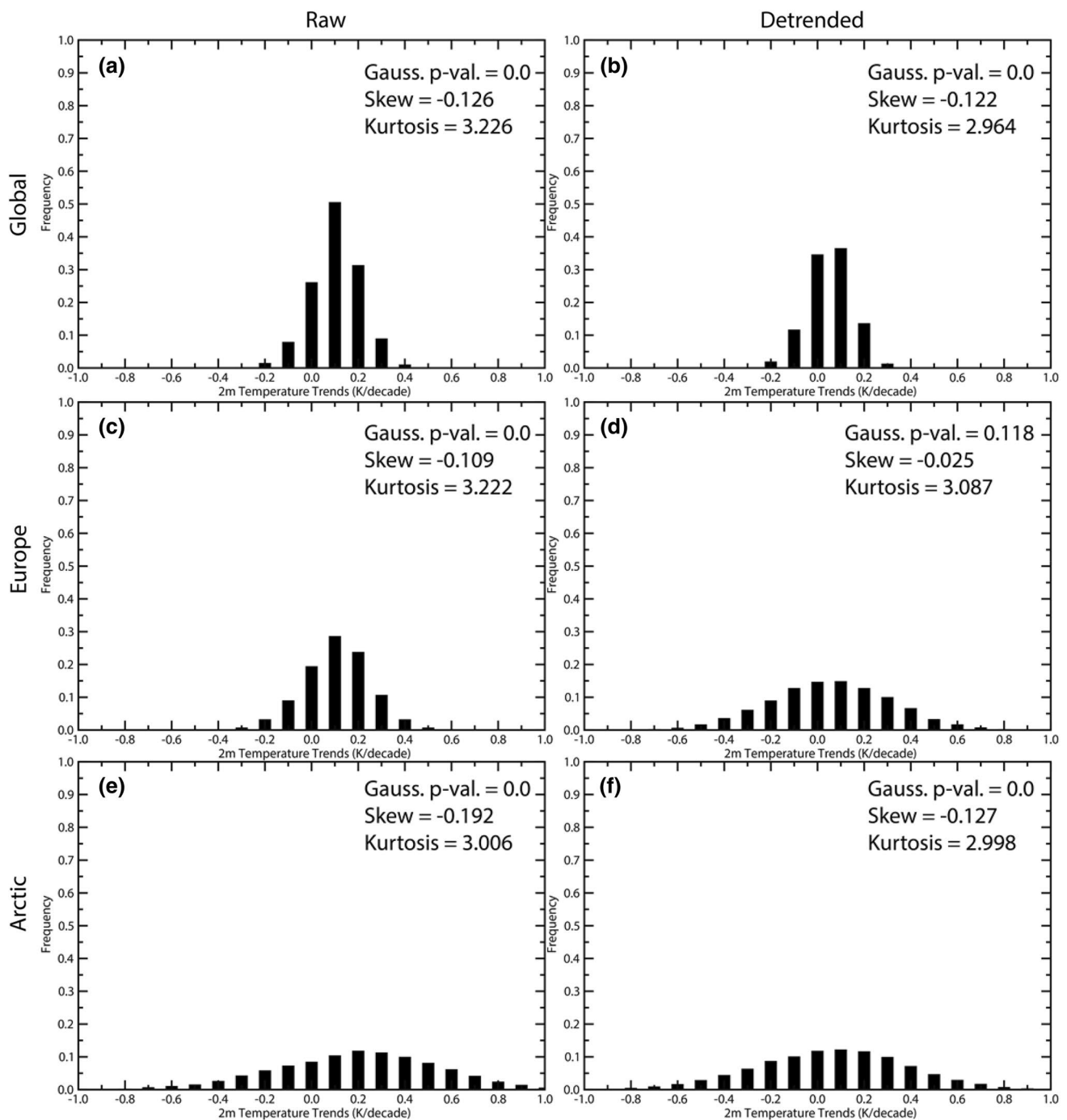
tions. Minimum trend (top), the mean trend (middle) and the maximum trend (bottom). Unit °K/decade, stippling indicates p values below 0.05. Area averaged values are shown in Table 1

**Table 2** Same as Table 1 but for the Arctic (60–90°N)

|               | 1856–1905           | 1906–1955           | 1956–2005          |
|---------------|---------------------|---------------------|--------------------|
| <b>Annual</b> |                     |                     |                    |
| Min           | −0.263 (174, 0.012) | −0.013 (142, 0.885) | 0.072 (115, 0.599) |
| Mean          | 0.017 (0.884)       | 0.178 (0.00)        | 0.318 (0.001)      |
| Max           | 0.276 (186, 0.00)   | 0.470 (162, 0.00)   | 0.490 (164, 0.00)  |
| <b>DJF</b>    |                     |                     |                    |
| Min           | −0.397 (112, 0.012) | −0.122 (142, 0.409) | 0.037 (144, 0.827) |
| Mean          | 0.027 (0.804)       | 0.242 (0.00)        | 0.420 (0.00)       |
| Max           | 0.456 (189, 0.00)   | 0.775 (162, 0.00)   | 0.761 (147, 0.00)  |
| <b>JJA</b>    |                     |                     |                    |
| Min           | −0.157 (174, 0.006) | 0.003 (130, 0.946)  | 0.040 (115, 0.479) |
| Mean          | 0.005 (0.900)       | 0.092 (0.00)        | 0.169 (0.00)       |
| Max           | 0.159 (158, 0.00)   | 0.245 (162, 0.00)   | 0.267 (181, 0.00)  |

calculated in the same way as in Sect. 3.3. The ensemble variation shown in Fig. 11b, shows that the variation of the precipitation for the European region is highly intriguing. First there is a slow but steady decrease for some 125 year after which it slowly starts to recover. The reason for this might be due to the fact that the climate of southern and northern Europe differs a lot and small changes in the position of the dominant storm track can create low frequency variations.

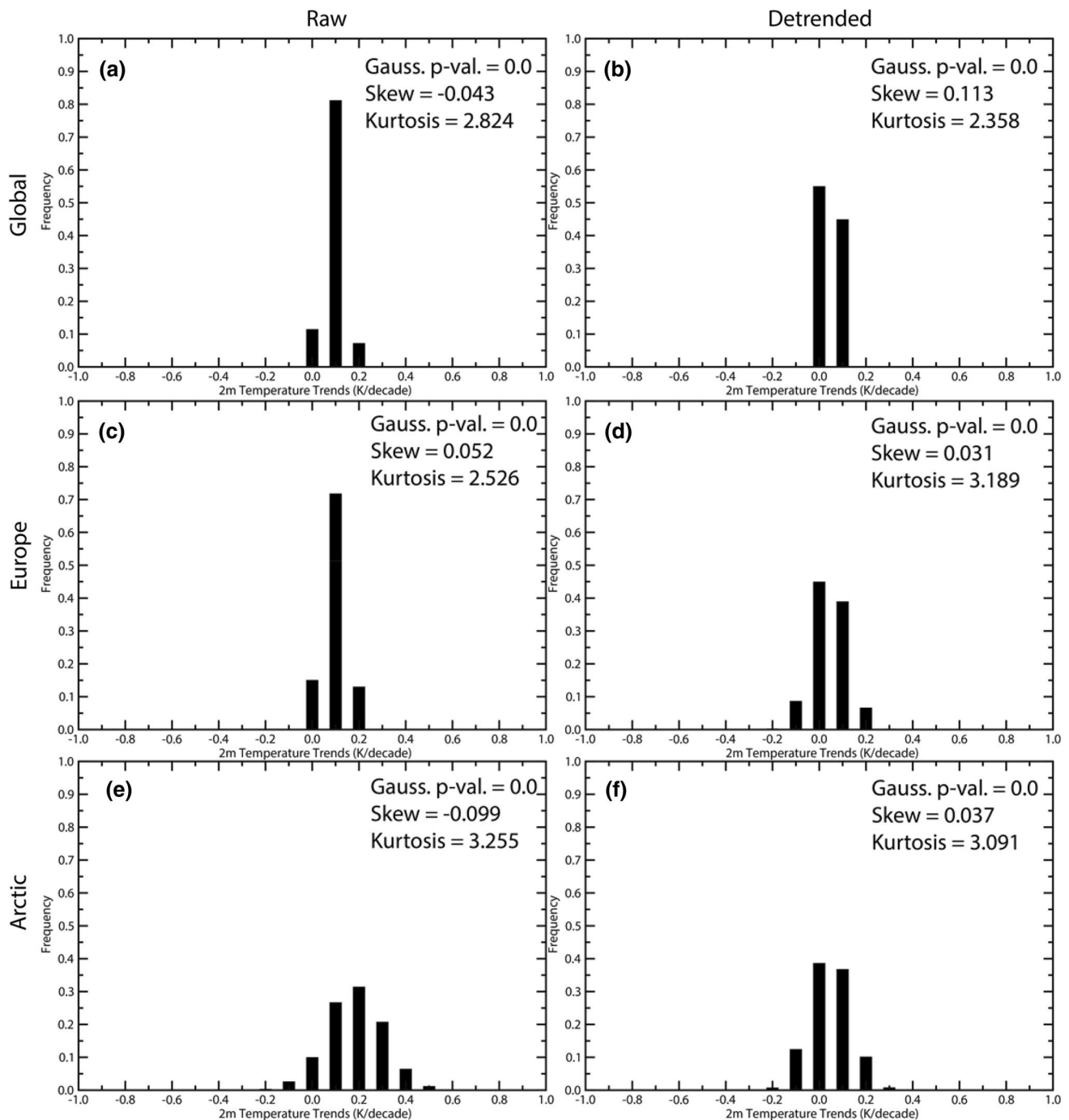
To obtain more detailed information on the variation of precipitation for the European area the data is divided in to the same three 50-year periods as in the previous section. Figure 12 shows the mean, the minimum and the maximum 50-year linear trends for the three periods, respectively. During the first two periods a minor drying trend can be seen but with a difference between northern and southern Europe. In the last period, 1956–2005, the trend towards higher precipitation in the northern part of Europe



**Fig. 7** Twenty year trends of 2 m temperature from the raw data (left column) and detrended using the spline fit (right column) for **a** and **b** global, **c** and **d** Europe, and **e** and **f** the Arctic

stands out more clearly. The high internal variability is clear from the minimum and maximum trends, suggesting that it is not feasible to relate any regional precipitation trends to changes in radiative forcing but that they are caused by random weather processes. For certain parts of the region, such as for central Europe, there is a range of more than 0.1 mm/day/decade (equivalent to ~200 mm

annual precipitation change over 50 years), between the wettest and the driest 50-year trend, that is significantly more than the mean effect which is about an order of magnitude less. However, few of the trends indicate any level of significance as indicated by the sparse stippling in Fig. 12. Numerical values are included in Table 3 as well



**Fig. 8** Same as Fig. 7 but for 50 year trends

as separate values for winter (DJF) and summer (JJA), as was done for surface temperature in Table 1.

In the Arctic on the other hand the situation is different with a systematically increasing precipitation trend (Fig. 11c). This is supported by theory (Held and Soden 2006), so that the Arctic net increase in water vapour

transport, and hence the increase in precipitation, scales well with the Clausius–Clapeyron relation (see also Bengtsson et al. 2011). It might therefore be expected that the increase in Arctic precipitation should broadly change proportional to the global temperature. This is indeed the case as can be

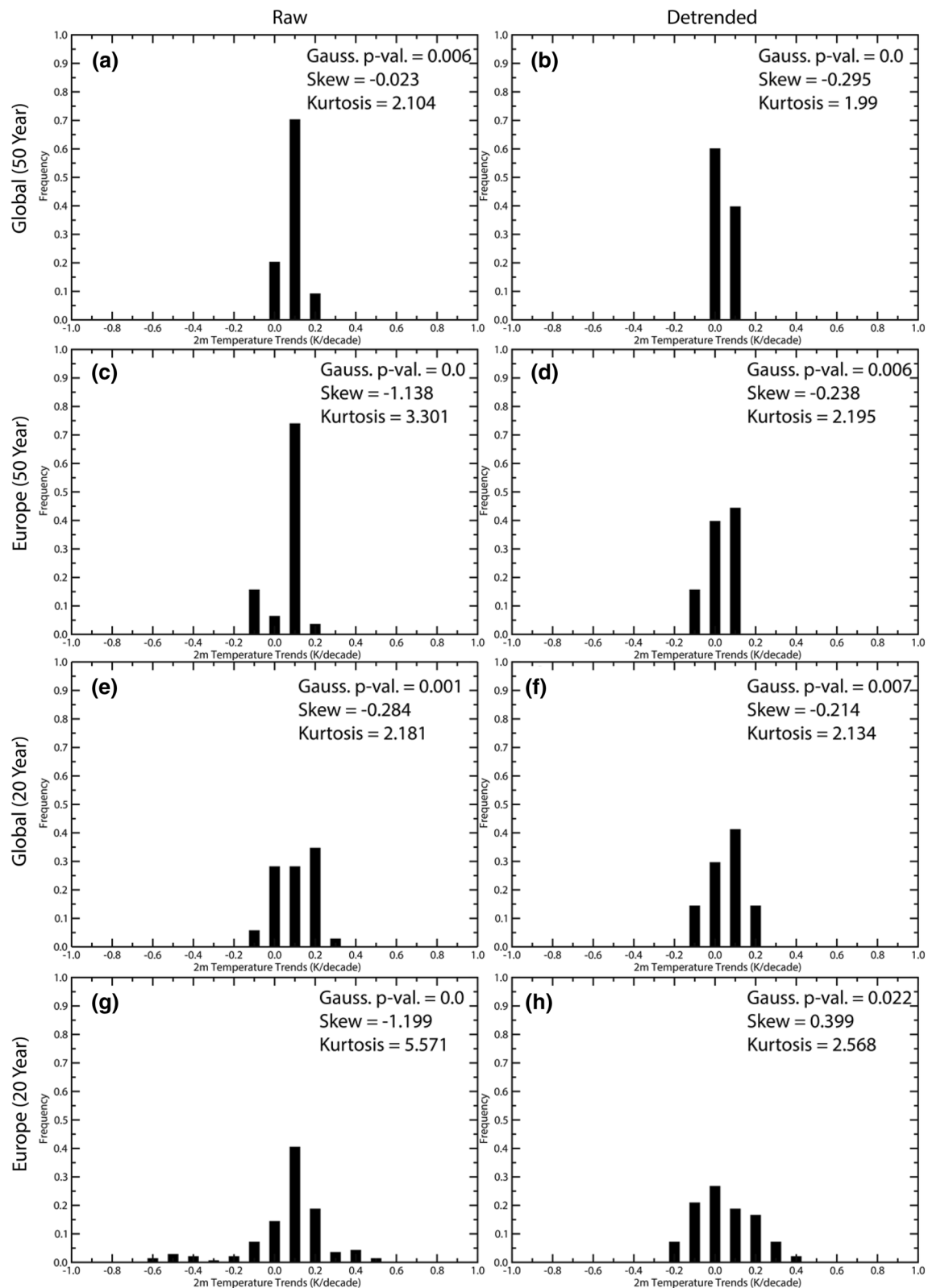
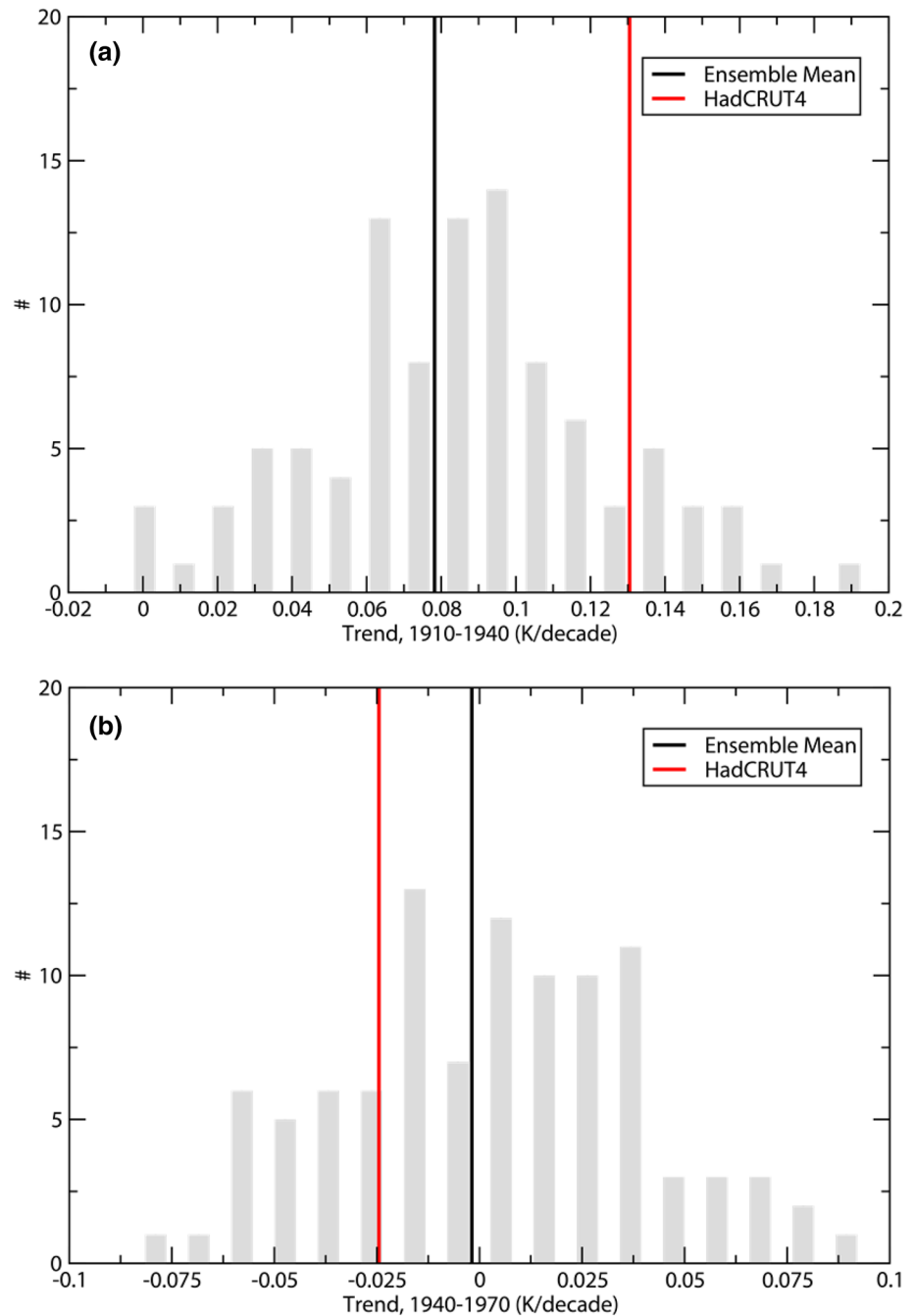


Fig. 9 Same as Figs. 7 and 8 but for HadCRUT4, for global and Europe only

**Fig. 10** Distribution of linear trends for the ensemble members (gray) and the values for the ensemble mean (black) and HadCRUT4 (red) for **a** the 1910–1940 period and **b** the 1940–1970 period. All trends have  $p$  values close to zero

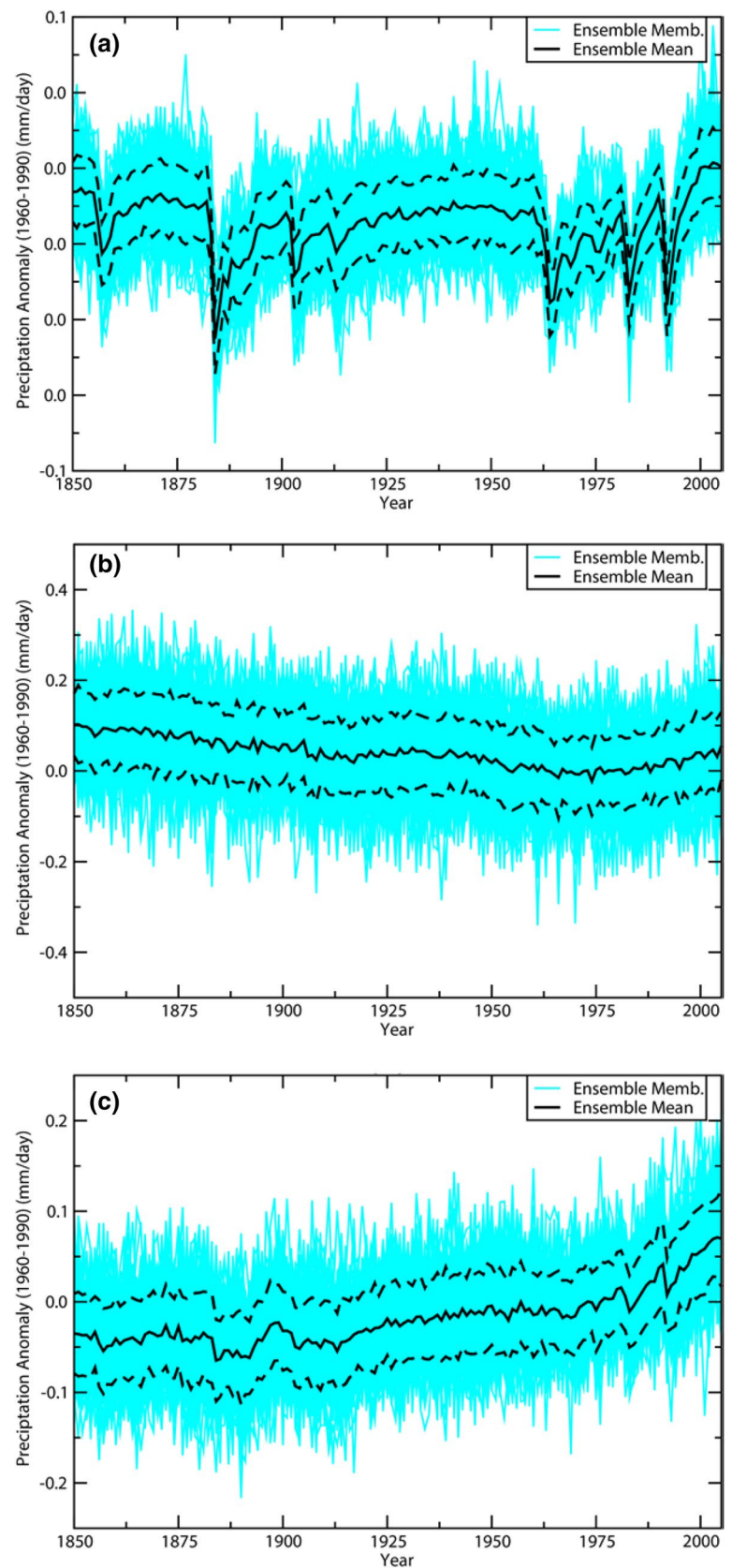


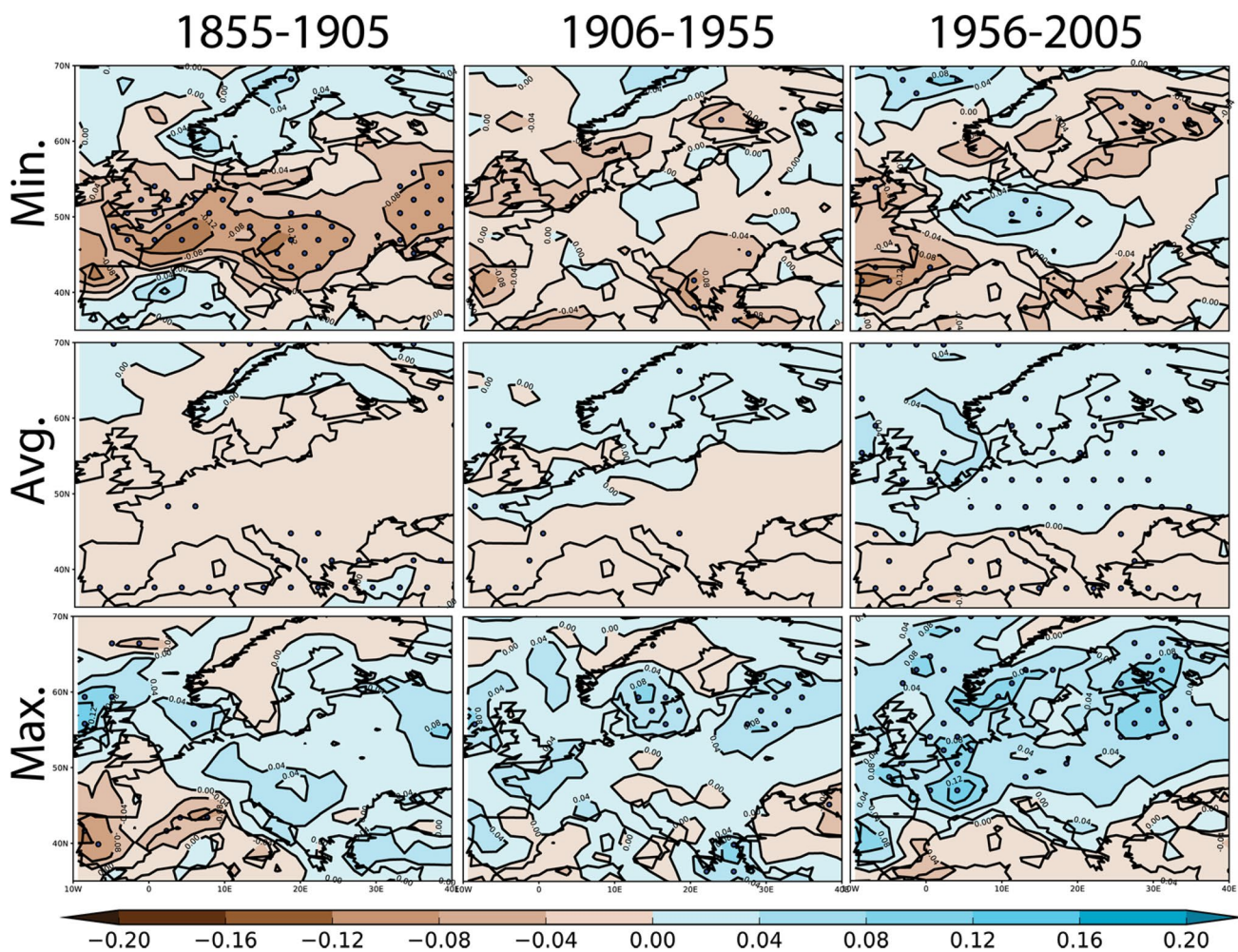
seen by comparing Fig. 2 showing the global temperature and Fig. 11c showing the Arctic precipitation.

Table 3 shows the range of precipitation trends for Europe for the same three 50-year periods used in Sect. 3.2 (c.f. Table 1). In addition to the annual values, values for winter (DJF) and summer (JJA) are also shown. Table 3 indicates that there is generally a very small change in the ensemble mean precipitation for each of the 50 year periods but the difference between the driest and wettest

50-year trends is about an order of magnitude larger. This means that over periods as long as 50 years there can be significant trends that either indicate a drying trend or a wetting trend. The difference of the two most extreme trends is more than ten times larger than the ensemble mean trend, meaning that precipitation trends are dominated by natural fluctuations, even on time scales of 50 years. The winter and summer precipitation give a similar picture. As can be seen from Fig. 12, there are

**Fig. 11** **a** Annual global mean precipitation anomaly as a function of time for all ensemble members (light blue), ensemble mean and  $\pm 1$  standard deviation (black). **b** Same as **a** but for the European region. **c** Same as **a** but for the Arctic





**Fig. 12** Same as Fig. 6 but for precipitation. Unit mm/day/decade. Area averaged values are shown in Table 3

differences between northern and southern Europe as the ensemble mean indicates more precipitation in northern Europe and less in southern Europe.

Table 4 shows the same results for the Arctic where a more distinct trend towards more precipitation for the full year as well as for winter and summer can be seen. The difference between the wettest and driest trends is much larger than the ensemble mean but less so than for the European area. See also Fig. 12.

## 4 Discussion

Examination of the model results indicate that the model has some systematic deficiencies. Observations are within the range of the ensemble experiments although there are indications that the variance of the model is larger than in the observational data. This is apparent by comparing the ensemble with HadCRUT4 and JRA55 which show that

practically all members of the ensemble have a higher variance. A similar result has been found for the NCAR Large Ensemble over North America (McKinnon et al. 2017). The disagreement between observational data and model results can be found both globally as well as for selected regions such as the European region. Furthermore, examination of the annual means (Figs. 2a, 4a) show that the model simulations show a stronger and more pronounced response to volcanic eruptions than can be seen from the observational data. To eliminate problems with the observational data in the early period an additional examination was done for the period 1956–2005 with essentially the same result.

The scientific questions posed in the introduction are now considered in turn.

(i) *How well can ensemble integrations with a modern coupled GCM be used to separate climate change signals in temperature and precipitation from the internal variability of these variables, and what are the implications for climate predictability?*

**Table 3** Europe (35–70°N, (10W–40E), 50-year trends (raw data) for three periods at the start, middle and end of the ensemble simulations for annual, DJF and JJA mean total precipitation, values in brackets indicate the ensemble member and p value respectively. Units mm/day/decade

|        | 1856–1905                           | 1906–1955                           | 1956–2005                           |
|--------|-------------------------------------|-------------------------------------|-------------------------------------|
| Annual |                                     |                                     |                                     |
| Min    | 1856–1905<br>–0.032 (167,<br>0.007) | 1906–1955<br>–0.021 (179,<br>0.011) | 1956–2005<br>–0.019 (122,<br>0.065) |
| Mean   | –0.011(0.00)                        | –0.002 (0.198)                      | 0.007 (0.009)                       |
| Max    | 0.009 (158, 0.288)                  | 0.019 (177, 0.087)                  | 0.032 (150, 0.010)                  |
| DJF    |                                     |                                     |                                     |
| Min    | –0.072 (117,<br>0.00)               | –0.053 (110,<br>0.009)              | –0.035 (194,<br>0.025)              |
| Mean   | –0.007 (0.072)                      | 0.004 (0.011)                       | 0.015 (0.00)                        |
| Max    | 0.045 (163, 0.033)                  | 0.054 (190, 0.003)                  | 0.058 (190, 0.00)                   |
| JJA    |                                     |                                     |                                     |
| Min    | –0.049 (105,<br>0.00)               | –0.038 (196,<br>0.002)              | –0.024 (141,<br>0.077)              |
| Mean   | –0.015 (0.00)                       | –0.005 (0.018)                      | 0.004 (0.134)                       |
| Max    | 0.013 (172, 0.437)                  | 0.022 (199, 0.112)                  | 0.034 (103, 0.00)                   |

**Table 4** Same as Table 3 but for the Arctic (60–90°N)

|        | 1856–1905              | 1906–1955              | 1956–2005              |
|--------|------------------------|------------------------|------------------------|
| Annual |                        |                        |                        |
| Min    | –0.022 (174,<br>0.00)  | –0.006 (142,<br>0.279) | 0.001 (169, 0.805)     |
| Mean   | 0.000 (0.913)          | 0.009 (0.00)           | 0.015 (0.001)          |
| Max    | 0.019 (134, 0.029)     | 0.023 (101, 0.00)      | 0.031 (166, 0.00)      |
| DJF    |                        |                        |                        |
| Min    | –0.032 (174,<br>0.00)  | –0.012 (142,<br>0.235) | –0.009 (106,<br>0.235) |
| Mean   | 0.001 (0.475)          | 0.010 (0.00)           | 0.013 (0.00)           |
| Max    | 0.028 (189, 0.019)     | 0.040 (184, 0.002)     | 0.036 (164, 0.00)      |
| JJA    |                        |                        |                        |
| Min    | –0.024 (146,<br>0.002) | –0.012 (126,<br>0.145) | –0.004 (198,<br>0.464) |
| Mean   | 0.000 (0.825)          | 0.007 (0.00)           | 0.013 (0.003)          |
| Max    | 0.017 (134, 0.023)     | 0.025 (196, 0.00)      | 0.025 (136, 0.009)     |

The temperature of the real atmosphere as well as the simulated temperature of the ensemble integration is exposed to three kinds of influence. Firstly, a long term effect essentially due to increasing greenhouse gases, secondly, short term effects mainly caused by volcanic eruptions and thirdly, internal processes in the climate system mainly due to weather events. We have decided here to separate the high frequency variations due to natural random processes from the slowly acting changes that we postulate to be due to external processes such as solar or greenhouse

effects. This has been achieved by removing the non-linear trend using a spline fit to the data that has the form rather similar to the change in radiative forcing due to the well mixed greenhouse gases (Bengtsson and Schwartz 2013). Another approach would be to consider the temperature of the ensemble mean as the signal and the deviation from the ensemble mean as the noise. The noise would then be due to internal change in the climate system. However, even the ensemble mean has some residual high frequency variability, especially associated with the volcanic events. Also, the spline approach is the only option for the observations with only a single realisation. The difference between the two methods are minor in spite of the influence from volcanic eruptions on the ensemble mean, though the second method only represents internal processes in the climate system.

It is clear that the magnitude of random processes must set limits for the deterministic inter-annual predictability. The large internal variability, as we have found in this study, implies that decadal type climate predictability will be a challenge if at all feasible. This was recently pointed out by von Känel et al. (2017) and this study supports this finding.

(ii) *What is the typical variance of global and regional multi-annual linear temperature trends and what is the likelihood of a long-term temperature hiatus? Are multi-decadal temperature anomalies in the twentieth century caused by internal processes or by external effects?*

The long-term warming trend (Fig. 1a) shows a pattern that is similar to that observed (see “<https://data.giss.nasa.gov/gistemp/maps/>”). The warming pattern is mostly larger over land than over the ocean areas and also more pronounced in the Northern Hemisphere and particularly in the Arctic. The variance of the ensemble varies significantly and in many parts of the world the noise dominates over the signal. Globally individual ensemble members can significantly deviate from the ensemble mean on a time scale of 20 years (Fig. 7a) which might well be an explanation for the hiatus observed in the present century. In some regions significant reverse trends as long as 50 years are possible (e.g. Fig. 6; Table 1). The ensemble experiment supports the hypothesis that the multi-decadal anomalies in the twentieth century were strongly influenced by internal processes in the climate system.

(iii) *Can any trend in the standard deviation and extremes of temperature and precipitation during the last 150 years be identified or are basic climate statistical parameters stationary?*

The standard deviation between members of the ensemble shows a decrease with time at middle and high latitudes with the largest decrease at 60–90°N in DJF (not shown). In other regions the decrease is minor. In the tropical belt

20°S–20°N there is a slight increase. A reduction in standard deviation in the Arctic seems plausible as a warmer Arctic ocean is expected to reduce available potential energy and thus provide less energy for the extra-tropical depressions (Bengtsson et al. 2006b; Zappa et al. 2013).

Global precipitation is virtually unchanged during the experiment but shows a marked reduction during major volcanic eruptions (Fig. 11a). However, as can be seen from Fig. 1b there are regional precipitation changes with generally less precipitation at subtropical and middle latitudes and increased precipitation at high latitudes and in some tropical areas.

## 5 Conclusions and summary

*Based on this ensemble study we conclude that internal processes in the climate system have played an important role in influencing decadal and multi-decadal temperature trends.*

Ensemble weather prediction studies have clearly indicated the limitation of the forecast skill of synoptic scale processes to be of the order of at most a few weeks (Lorenz 1982). As climate is nothing other than the envelope of all possible weather that has occurred or might occur in a certain time and space, it also follows that a climate prediction, in the form of an integration extended to time scales from years to centuries, can also be affected by chaotic processes, as shown in this study. However, if this is the case, it is also logical that the history of weather and climate is also just one unique realization of all possible weather and climate that might have occurred in the past. The fact is that quite another set of weather and climate could have occurred in the past even if the external processes of radiative forcing from greenhouse gases, aerosols and solar variability etc. would have been known exactly! In other words it is not sufficient to know the observed past climate but we also need to know what other climates that might have occurred over the same period of time. This knowledge is in fact required for an in depth understanding of the dynamics of climate and requires ensemble integrations. So even if we know the past weather and climate exactly, as well as the external radiative forcing and its variation in time, this is not sufficient knowledge if we wish to understand how climate might evolve in the future.

This is elucidated by the following example. Assume that a scientist decided to empirically determine the transient climate sensitivity (TCS) using the change in near-surface temperature for the period 1956–2005. Using a value for the net increase in external forcing by say  $1.5 \text{ W/m}^2$  during the period (a possible realistic estimate), one would then

arrive at a value of TCS of  $0.64^\circ\text{C}/1.5 \text{ W/m}^2 = 0.43$ . However, the scientist could equally well have arrived at another value of TCS in the range 0.29–0.53 if we assume that the range of internal variability could have been the same as in the present ensemble simulation for the period 1956–2005. The corresponding realized warming for doubled  $\text{CO}_2$  would then be  $1.07^\circ\text{C} - 1.95^\circ\text{C}$  or a factor of 1.8. Similar values have been obtained by Otto et al. (2013). For the equilibrium climate sensitivity the values would be proportionally larger by 39% (Bengtsson and Schwartz 2013, their Table 1).

So even in a case of perfect knowledge of external forcing TCS and presumably also equilibrium climate sensitivity could only be determined by a factor of about 1.8 if global data for a period of 50 years is used. The use of shorter periods or smaller domains would increase the range further.

With relevance to climate change studies it should be pointed out that this range of uncertainty is only a consequence of unpredictable processes and not to additional uncertainties caused by model deficiencies such as related to physical parameterizations and model resolution.

The study has suggested that only a minor global increase in ensemble mean precipitation occurs but that systematic regional changes occur in agreement with theory, showing that horizontal transport of water vapour follows the Clausius–Clapeyron relation. This can be clearly seen in the Arctic precipitation that shows a systematic increase that is broadly proportional to the increase in global water vapour and thus closely related to the global temperature increase.

The study has also shown considerable stochastic variability in precipitation that noticeably affects regional precipitation trends on a time scale of 50-years. Model precipitation shows a strong response to volcanic aerosols with a marked reduction in precipitation. This can be seen in all ensemble members. It is suggested that this is caused by diminished solar radiation leading to reduced evaporation. Unfortunately, present precipitation observations are not reliable enough to verify this finding.

Finally, we wish to highlight that the present study has been undertaken by a single climate model. We have for example noted that the temporal variance of the majority of ensemble members is larger than what can be inferred from available observations. The result of the study must be assessed with that in mind. We have no simple explanation to this but it might be that the model projects the variance on larger scales than nature as a consequence of limited resolution. We would consequently encourage other modeling groups to undertake similar studies which will hopefully make use of the latest high resolution models coupled models (Haarsma et al. 2016). Intuitively we might have expected the opposite and that reality might expose a higher level of variance than the climate model.

**Acknowledgements** We are grateful to the Max Planck Institute for Meteorology for making data from the ensemble experiment available and the Japan Meteorological Agency (JMA) for making the Japanese 55-year Reanalysis (JRA-55) data available.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Anderson TW, Darling DA (1954) A test of goodness-of-fit. *J Am Stat Assoc* 49:765–769. <https://doi.org/10.2307/2281537>
- Bengtsson L, Schwartz SE (2013) Determination of a lower bound on Earth's climate sensitivity. *Tellus B* 65:21533
- Bengtsson L, Hodges KI, Roeckner E, Brokopf R (2006a) On the natural variability of the pre-industrial European climate. *Clim Dyn* 27:743–760
- Bengtsson L, Hodges KI, Roeckner E (2006b) Storm Tracks and Climate Change. *J Clim* 19:3518–3543
- Bengtsson L, Hodges KI, Koumoutsaris S, Zahn M, Keenlyside N (2011) The changing atmospheric water cycle in Polar Regions in a warmer climate. *Tellus A* 63:907–920. <https://doi.org/10.1111/j.1600-0870.2011.00534.x>
- Cowan K, Way RG (2014) Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart J R Meteorol Soc* 140:1935–1944
- Dai A, Bloecker CE (2018) Impacts of internal variability on temperature and precipitation trends in large ensemble simulations by two climate models. *Clim Dyn*. <https://doi.org/10.1007/s00382-018-4132-4>
- Delworth TL, Knutson TR (2000) Simulation of early 20th century global warming. *Science* 287:2246–2250
- Deser C, Phillips AS, Bourdette V, Teng H (2012a) Uncertainty in climate change projections: the role of internal variability. *Clim Dyn* 38:527–546
- Deser C, Knutti R, Solomon S, Phillip AS (2012b) Communication of the role of natural variability in future North American climate. *Nat Clim Change* 2:775–779
- Deser C, Phillips AS, Alexander MA, Smoliak BV (2014) Projecting North American Climate over the next 50 years: uncertainty due to internal variability. *J Clim* 27:2271–2296
- Deser C, Terray L, Phillips AS (2016) Forced and internal components of winter air temperature trends over North America during the past 50 years: mechanisms and implications. *J Clim* 29:2237–2258
- Driscoll S, Bozzo A, Gray LJ, Robock A, Stenchikov G (2012) Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. *J Geophys Res*. <https://doi.org/10.1029/2012JD017607>
- Fisher RA (1934) Statistical methods for research workers, 4th edn. Oliver & Boyd, Edinburgh
- Fyfe JC, Meehl GA, England MH, Mann ME, Santer BD, Flato GM, Gregory M, Hawkins E, Gillett NP, Xie S-P, Kosaka Y, Swart NC (2016) Making sense of the early-2000s warming slowdown. *Nat Clim Change* 6:224–228
- Giorgetta MA et al (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J Adv Model Earth Syst* 5:572–597
- Haarsma RJ, Roberts MJ, Vidale PL, Senior CA, Bellucci A, Bao Q, Chang P, Corti S, Fučkar NS, Guemas V, von Hardenberg J, Hazeleger W, Kodama C, Koenigk T, Leung LR, Lu J, Luo J-J, Mao J, Mizielinski MS, Mizuta R, Nobre P, Satoh M, Scoccimarro E, Semmler T, Small J, von Storch J-S (2016) High resolution model intercomparison project (HighResMIP v1.0) for CMIP6. *Geosci Model Dev* 9:4185–4208
- Hawkins E, Smith RS, Gregory JM, Stainforth DA (2016) Irreducible uncertainty in near-term climate projections. *Clim Dyn* 46:3807. <https://doi.org/10.1007/s00382-015-2806-8>
- Held IM, Soden BJ (2006) Robust responses of the hydrological cycle to global warming. *J Clim* 19:5686–5699
- Hu S, Fedorov AV (2017) The extreme El Niño of 2015–2016 and the end of global warming hiatus. *Geophys Res Lett* 44:3816–3824
- Hunt BG, Elliot TI (2006) Climatic Trends. *Clim Dyn* 26:567–585
- ICPO (1998) CLIVAR Initial Implementation Plan, WCRP No. 103, WMO/TS No. 869, ICPO No. 14. June 1998
- Iles CE, Hegerl GC, Schurer AP, Zhang X (2013) The effect of volcanic eruptions on global precipitation. *JGR Atmos* 118:8770–8786. <https://doi.org/10.1002/jgrd.50678>
- Kay JE, et al (2015) The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bull Am Meteor Soc* 96:1333–1349
- Kinne S, O'Donnell D, Stier P, Kloster S, Zhang K, Schmidt H, Rast S, Giorgetta M, Eck TF, Stevens B (2013) MAC-v1: a new global aerosol climatology for climate studies. *J Adv Model Earth Syst* 5:704–740
- Kobayashi S, et al (2015) The jra-55 reanalysis: general specifications and basic characteristics. *J Meteorol Soc Jpn* 93, 5–48. <https://doi.org/10.1175/JCLI-D-14-00005.1>
- Kopp G (2016) Magnitudes and timescales of total solar irradiance variability. *J Space Wea Space Clim* 6:A30. <https://doi.org/10.1051/swsc/2016025>
- Lorenz E (1964) The problem of deducing the climate from the governing equations. *Tellus* 16:1–11
- Lorenz E (1982) Atmospheric predictability experiments with a large numerical model. *Tellus* 34:505–513
- Lyapunov AM (1892) The general problem of the stability of motion (in Russian). *Kharkov Mathematical Society, Collected Works II*, 7, pp 250
- Lyapunov AM (1992) The general problem of the stability of motion. *Int J Control* 55:531–773 (translated into English by A. T. Fuller)
- Machete RL, Smith LA (2016) Demonstrating the value of larger ensembles in forecasting physical systems. *Tellus A* 68:28393. <https://doi.org/10.3402/tellusa.v68.28393>
- McKinnon KA, Poppick A, Dunn-Sigouin E, Deser C (2017) An “Observational Large Ensemble” to compare observed and modeled temperature trend uncertainty due to internal variability. *J Clim* 30:7585–7598
- Morice CP, Kennedy JJ, Rayner NA, Jones PD (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. *J Geophys Res* 117:D08101. <https://doi.org/10.1029/2011JD017187>
- Otto A, Otto FEL, Boucher O, Church J, Hegerl G, Forster PM, Gillett NP, Gregory J, Johnson GC, Knutti R, Lewis N, Lohmann U, Marotzke J, Myhre G, Shindell D, Stevens B, Allen MR (2013) Energy budget constraints on climate response. *Nat Geosci* 6:415–416
- Pithan F, Mauritsen T (2014) Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nat Geosci* 7:181–184
- Poincaré H (1892) *Methodes nouvelles de la mecanique celeste*. Gauthier-Villars, Paris (3 vols)
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>

- Robock A (2000) Volcanic eruptions and climate. *Rev Geophys* 38:191–219
- Santer B, Bonfils C, Painter JF, Zelinka M, Mears C, Solomon S, Schmidt G, Fyfe JC, Cole J, Nazarenko L, Taylor KE, Wentz FJ (2014) Volcanic contribution to decadal changes in tropospheric temperature. *Nat Geosci*. <https://doi.org/10.1038/ngeo2098>
- Sigmond M, Fyfe JC (2016) Tropical Pacific impacts on cooling North American winters. *Nat Clim Change* 6:970–975
- Tett SF, et al (2002) Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J Geophys Res* 107:4306. <https://doi.org/10.1029/2000JD000028>
- Thompson DWJ, Barnes EA, Deser C, Foust WE, Phillips AS (2015) Quantifying the role of internal climate variability in future climate trends. *J Clim* 28:6443–6456
- Trenberth KE (2015) Has there been a hiatus? *Science* 349(6249):691–692
- von Känel L, Frölicher TL, Gruber N (2017) Hiatus-like decades in the absence of equatorial Pacific cooling and accelerated global ocean heat uptake. *Geophys Res Lett*. <https://doi.org/10.1002/2017GL073578>
- Wallace JM, Fu Q, Smoliak BV, Lin P, Johanson CM (2012) Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. *Proc Natl Acad Sci* 109:14332–14337
- Wigley TML, Jones PD (1981), Detecting CO<sub>2</sub>-induced climatic change. *Nature* 292:205–208
- Wilcox LJ, Highwood EJ, Dunstone NJ (2013) The influence of anthropogenic aerosol on multi-decadal variations of historical global climate. *Environ Res Lett* 8:024033
- Zappa G, Shaffrey LC, Hodges KI, Sansom PG, Stephenson DB (2013) A multimodel assessment of future projections of north atlantic and european extratropical cyclones in the CMIP5 climate models. *J Clim* 26:5846–5862