



Probabilistic forecasts of near-term climate change: verification for temperature and precipitation changes from years 1971–2000 to 2011–2020

Jouni Räisänen¹

Received: 7 July 2021 / Accepted: 27 January 2022 / Published online: 12 February 2022
© The Author(s) 2022

Abstract

In year 2006, Räisänen and Ruokolainen proposed a resampling ensemble technique for probabilistic forecasts of near-term climate change. Here, the resulting forecasts of temperature and precipitation change from years 1971–2000 to 2011–2020 are verified. The forecasts of temperature change are found to be encouragingly reliable, with just 9% and 10% of the local annual and monthly mean changes falling outside the 5–95% forecast range. The verification statistics for temperature change represent a large improvement over the statistics for a surrogate no-forced-change forecast, and they are largely insensitive to the observational data used. The improvement for precipitation changes is much smaller, to a large extent due to the much lower signal-to-noise ratio of precipitation than temperature changes. In addition, uncertainty in observations is a major complication in verification of precipitation changes. For the main source of precipitation data chosen in the study, 20% and 15% of the local annual and monthly mean precipitation changes fall outside the 5–95% forecast range.

Keywords Climate change · Probabilistic forecast · Verification · Temperature · Precipitation · CMIP3

1 Introduction

The typical accuracy of weather forecasts is easy to assess using verification data collected from past forecasts (Nurmi 2003; Jolliffe and Stephenson 2011). However, such an assessment for projections of climate change is far more difficult, due to at least three reasons:

1. The time scale of the projections limits the number of meaningful verification cases. Century-scale projections of the effect that large increases in greenhouse gas concentrations might have on climate (Collins et al. 2013) are not directly verifiable at all, since the target period of these projections has not been reached.
2. For climate change projections on shorter (decadal-to-multidecadal) time scales and therefore weaker anthropogenic forcing, the forced change might still be heavily

masked by internal variability (e.g., Deser et al. 2012). Thus, an imperfect simulation of the observed climate change would be expected even in a perfect model with perfectly specified forcing.

3. The forcing, comprising anthropogenic changes in the atmospheric composition and land use as well as solar variability and volcanic eruptions, may differ between the simulations and the real world. This is both due to the scientific uncertainty and technical difficulty in describing some types of forcing in climate models (e.g., aerosols and land use change) and because the time evolution of both anthropogenic and natural forcing is only partly predictable in advance. Therefore, a climate model with poorly specified forcing might produce a bad projection of climate change even if its sensitivity to the forcing were correct. Conversely, error compensation between the specification of the forcing and the model response to the forcing might result in a good simulation of climate change for the wrong reasons (Knutti 2008).

✉ Jouni Räisänen
jouni.raisanen@helsinki.fi

¹ Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, P.O. Box 64, 00014 University of Helsinki, Finland

Due to the signal-to-noise issue in the verification of short-term climate change projections, climate models have traditionally been evaluated in hindcast mode, by comparing simulated and observed climate changes during the late

nineteenth to early twenty-first century instrumental era (Räisänen 2007; Flato et al. 2013; van Oldenborgh et al. 2013). However, the knowledge of the observed climate change may affect the choices that the modelers make when developing their models and specifying, for example, the aerosol forcing in the simulations (Knutti 2008; Schmidt et al. 2017; Gettelman et al. 2019; Mauritsen et al. 2019). Therefore, hindcast verification is not equivalent to forecast verification.

Nevertheless, the continuing increase of atmospheric greenhouse gas concentrations together with the aging of the earliest climate model projections is gradually making a forecast mode verification of these projections more meaningful. Rahmstorf et al. (2007) compared the observed evolution of global mean temperature, global sea level and atmospheric CO₂ concentration in years 1990–2007 with the projections presented in the third Intergovernmental Panel on Climate Change assessment (IPCC 2001). They found the increase in CO₂ concentration to closely track the central IPCC estimate, whereas the observed warming since 1990 was in the upper half, and the increase in sea level close to the upper bound of the projections. By contrast, Frame and Stone (2013) found the central business-as-usual (Scenario A) temperature projection in IPCC (1990) to exceed the observed warming in 1990–2010 by approximately 50%, although this overestimate was affected by a too rapid increase in radiative forcing. More recently, Hausfather et al. (2020) reviewed 17 projections of global mean temperature change, published between 1970 and 2007 and representing projection periods varying from 1970–2000 to 2007–2017. Ten of these projections were consistent with observations, in the sense that the 95% confidence interval for the difference between the observed and projected warming included zero. Of the remaining seven projections, four overestimated and three underestimated the warming. When accounting for the differences between the assumed and realized radiative forcing, 14 out of the 17 projections were judged to be consistent with observations.

Nearly all forecast mode comparisons between projected and observed climate change have focused on the global mean temperature, which has a higher signal-to-noise ratio between the forced change and internal variability than local climate changes. As an exception, Stouffer and Manabe (2017) compared the geographical and seasonal patterns of temperature change from 1961–1990 to 1991–2015 with the idealized CO₂-only climate change simulations first published by Stouffer et al. (1989). Their study revealed a good agreement on many large-scale features, such as generally larger warming over land than ocean, a maximum of warming in high northern latitudes with much less warming over the Southern Ocean, and a general although not detailed consistency in the distribution of zonally averaged monthly mean temperature

changes. The simulation also included a pronounced minimum of warming over the northern North Atlantic, attributed in part to the weakening of the Atlantic thermohaline circulation. This minimum was less pronounced in the observed temperature difference between 1961–1990 and 1991–2015, but it stands out clearly in trends calculated for the longer 1901–2012 period (Hartmann et al. 2013, Fig. 2.21). Still, the comparison in Stouffer and Manabe (2017) was qualitative only, as dictated by the much stronger radiative forcing in the idealized model simulation than in the real world.

The impacts of climate change are not determined by the global mean warming alone. Therefore, more research is needed on how well models have been able to predict recently observed local-to-regional climate changes. To this end, this study verifies the grid box scale climate change forecasts of Räisänen and Ruokolainen (2006, hereafter RR06) for changes in annual and monthly mean temperature and precipitation from the years 1971–2000 to 2011–2020. Importantly, these forecasts were formulated in probabilistic terms, considering the uncertainties associated with both climate model response to anthropogenic forcing and internal climate variability. Such a probabilistic verification remains meaningful even when the signal-to-noise ratio of the observed changes is relatively low. The use of the word “forecast” here follows the wording in RR06. However, unlike weather forecasts and seasonal-to-decadal climate predictions (Kirtman et al. 2013; Suckling 2018), these projections do not use information on the initial state of the climate system.

In the following, the methods used to generate the probabilistic climate change forecasts of RR06 are first outlined and the data sets and statistics applied in their verification are introduced (Sect. 2). After this, the verification results for temperature (Sect. 3) and precipitation change (Sect. 4) are described. The results for temperature change are particularly encouraging, the forecast being both reliable in a probabilistic sense and much better than a surrogate forecast assuming no forced climate change. The conclusions are given in Sect. 5. Some additional information is included in the Supplementary material.

2 Data and methods

This section first summarizes the resampling ensemble technique developed by RR06. After this, the observational data sets used for estimating the temperature and precipitation changes from 1971–2000 to 2011–2020 are introduced. Finally, the verification statistics used for evaluating the probabilistic climate change forecasts are described.

2.1 The resampling ensemble technique

The resampling ensemble technique was developed by RR06 to generate probabilistic climate change forecasts that include the two main uncertainties in near-term climate change: (i) uncertainty in climate models and (ii) internal climate variability (Hawkins and Sutton 2009, 2010). The effects of the emission scenario uncertainty can be explored by making probabilistic forecasts separately for different scenarios. However, following the main focus in RR06, only the Special Report on Emission Scenarios (Nakićenović and Swart 2000) A1B scenario (SRES A1B) is used in this study.

Climate modelling uncertainty was represented in RR06 using 21 models in the third phase of the Coupled Model Intercomparison Project, CMIP3 (Meehl et al. 2007). Following the indistinguishable ensemble paradigm (Annan and Hargreaves 2010), which assumes that climate changes in the real world belong to the same population as those in the model simulations, the same weight was given to all 21 models. For each model, only one realization of the simulated twentieth-to-twenty-first century climate was used.

Simulated climate changes between two periods of time represent a combination of forced response and internal variability. Thus, the distribution of climate changes from (e.g.) 1971–2000 to 2011–2020 among 21 simulations from different models implicitly represents both the modelling uncertainty and internal climate variability. Still, a sample of 21 would be too small for meaningful probabilistic forecasts. Therefore, RR06 used resampling in time to increase the sample size. The resampling algorithm is detailed below and is further motivated in RR06.

1. The change in the 21-model mean global mean temperature G between the actual baseline and projection periods was calculated. For the SRES A1B scenario, the average warming from 1971–2000 to 2011–2020 is $\Delta G = 0.622^\circ\text{C}$.
2. Running 30-year (G_{30}) and 10-year means (G_{10}) were calculated from the time series of G from 1901 to 2098.
3. Starting from 1931 to 1940, a 10-year period (P_{10}) was stepped forward with 5-year steps. For each choice of P_{10} , that earlier 30-year period (P_{30}) was identified for which the corresponding multi-model mean global mean temperature difference $G_{10}-G_{30}$ was the closest to the target value ΔG .
4. If the found minimum of $|G_{10} - G_{30} - \Delta G|$ was smaller than 0.03°C , P_{30} and P_{10} were added to the list of analogue periods. The resulting list includes 20 pairs P_{30}/P_{10} , ranging from 1906–1935/1991–2000 to 2051–2080/2086–2095 (Table S1). Among these 20 pairs, the multi-model global mean warming from P_{30} to P_{10} varies from 0.608 to 0.636°C .

5. It was assumed that the probability distribution of climate changes only depends on the multi-model global mean warming. Under this assumption, the changes between all the 20 found pairs of periods belong to the same statistical population as those from 1971–2000 to 2011–2020. This allowed increasing the nominal sample size for the probabilistic forecast from 21 to $21 \times 20 = 420$.

The assumption named in step 5 is a variant of the widely used pattern scaling assumption, which states that local climate changes should scale linearly with the global mean warming (Tebaldi and Arblaster 2014). This assumption is not precise. The regional patterns of forced climate change depend on the mixture of forcing agents that collectively cause the global mean temperature change (Shiogama et al. 2013), and they may also change with time as the state of the climate system evolves (Hawkins et al. 2014; see also Fig. 2 in RR06). Assumption 5 also presupposes that the uncertainty in climate change between two periods of time is independent of the timing of these periods as far as the multi-model global mean warming between them is the same. The rationale is that (i) under the pattern scaling assumption, intermodel differences in forced climate change should also be proportional to the global warming between these two periods, and (ii) as far the mean climate does not change too much, internal climate variability should broadly retain its magnitude. At least the latter assumption may sometimes fail, particularly for scenarios with strong greenhouse gas forcing and (considering temperature) specifically in sea areas where melting of ice curtails the variability of air temperature (e.g., LaJoie and DelSole 2016). Nonetheless, cross verification between the 21 CMIP models indicated that the resampling should clearly improve the resulting probabilistic climate change forecasts over those obtained by just using the changes between the nominal baseline and forecast periods (Table 2 in RR06).

For comparison with the actual forecasts, surrogate “no-forced-change” forecasts were constructed by re-centering the actual 420-member forecast ensemble to zero.¹ These forecasts thus have zero mean but the same width as the actual forecast distributions. In fact, the width of these distributions should have been slightly reduced, because intermodel differences in the forced response (which would be absent for constant forcing) widen the actual forecast distributions from what results from internal variability alone. However, for the time periods considered in this study, the variance in the probabilistic forecasts is dominated by

¹ RR06 used the CMIP3 preindustrial control simulations for the same purpose, but the data from these simulations were lost before the present study.

Table 1 The observational data sets used in this study

Temperature	References	Coverage	Shorthand
ERA5 reanalysis	Hersbach et al. (2020)	Global	ERA5
HadCRUT5.0.1.0	Morice et al. (2021)	99.6%	HadCRUT
GISTEMP v4	Lenssen et al. (2019)	99.6%	GISTEMP
NOAAGlobalTemp v5	Zhang et al. (2019)	92.6%	NOAA
Berkeley Earth land/ocean temperature record	Rohde and Hausfather (2020)	Global	Berkeley Earth
Precipitation	Reference	Coverage	Shorthand
ERA5 reanalysis	Hersbach et al. (2020)	Global	ERA5
JRA-55 reanalysis	Kobayashi et al. (2015)	Global	JRA-55
NCEP-NCAR reanalysis 1	Kalnay et al. (1996)	Global	NCEP
GPCC full data monthly product version	Schneider et al. (2020)	Land at 60° S–90° N	GPCC
CRU TS v4.05	Harris et al. (2020)	Land at 60° S–90° N	CRU

NCEP-NCAR National Centers for Environmental Prediction—National Center for Atmospheric Research, *GPCC* Global Precipitation Climatology Centre, *CRU TS* Climatic Research Unit gridded Time Series

internal variability. The estimated, globally averaged contribution of intermodel differences to this variance ranges from only 5% for monthly mean precipitation changes to 29% for annual mean temperature changes (Supplementary material, Sect. S2).

2.2 Data sets

For evaluating real-world climate changes from 1971–2000 to 2011–2020, five different data sets were considered for both temperature and precipitation. These data sets, together with their main references, spatial coverage and shorthand notations are listed in Table 1. Further details are given in Supplementary Tables S3–S4. All the observational data were regridded to the $2.5^\circ \times 2.5^\circ$ latitude–longitude grid of RR06 by using conservative remapping.

The main findings for temperature change turned out to be robust to the choice of the data set. The European Centre for Medium Range Weather Forecasts ERA5 reanalysis is used as the main data source for temperature in this article.

For precipitation change, observational uncertainty is a much larger issue. The gauge-based GPCC and CRU data sets were judged to be more reliable than the three reanalyses listed in Table 1, but they only cover land areas and exclude Antarctica. Therefore, for most of the analysis in this article, GPCC was used over land at 60° S–90° N and ERA5 elsewhere. GPCC was preferred over CRU because it includes a larger number of rain gauge observations since the 1990s (Dai and Zhao 2017).

The analyzed or reanalyzed temperature and precipitation changes will be referred to as observed changes in this article, except where it is necessary to stress which data set is used. This wording is used for simplicity, not to undermine the uncertainty in especially precipitation change.

2.3 Verification statistics

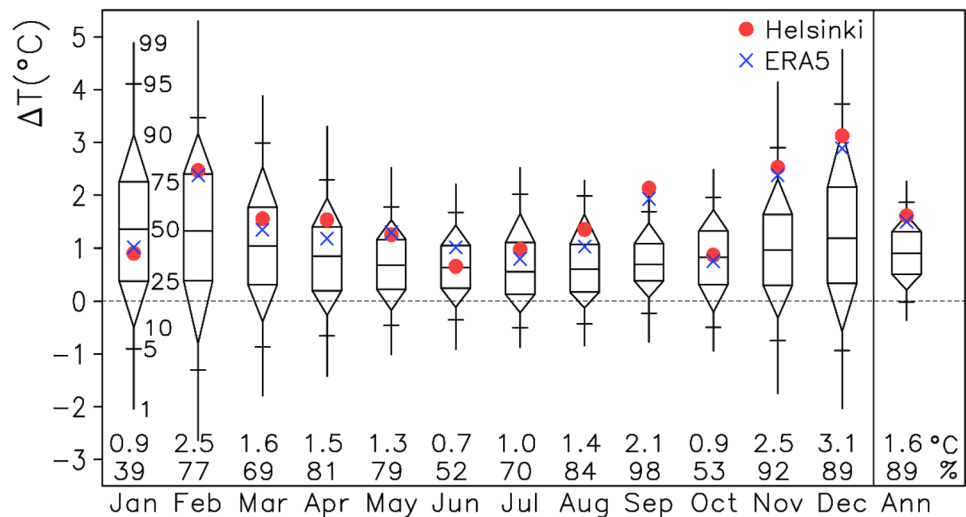
In addition to pointwise comparison with observed climate changes, percent rank histograms and the continuous ranked probability score (CRPS) together with the corresponding skill score (CRPSS) are used to characterize the quality of the probabilistic forecasts. For the percent rank histograms, the local cumulative probabilities of the observed changes are first calculated as the fraction of realizations in the forecast ensemble (out of 420) for which the change is smaller than observed. Then these probabilities are assigned to 5% classes (0–5%, 5–10% ... 95–100%) and the frequencies in each class are averaged over the global area and (for monthly changes) over the 12 months. Probability values falling at the class boundaries (e.g., $21/420 = 0.05$) are counted with half-weight in both the lower and the upper class.

For a reliable probabilistic forecast system, the percent rank histograms should be flat, with an equal frequency of verification cases in all parts of the forecast distribution. However, of two forecast systems producing equally flat histograms, the one with narrower (or, sharper) forecast distributions would be more useful (Gneiting et al. 2007). CRPS (Candille and Talagrand 2005) is a summary measure affected by both the reliability and the sharpness of the probabilistic forecasts. It can be written as

$$\overline{\text{CRPS}} = \overline{\int_{-\infty}^{\delta} F^2 dx + \int_{\delta}^{\infty} (1 - F)^2 dx} \quad (1)$$

where δ is the observed change, F is the cumulative probability from the forecast and the overbar indicates averaging over all relevant cases—here over the global area and (for the monthly changes) over the 12 months. For a deterministic forecast, F would jump from 0 to 1 at the forecast value,

Fig. 1 Changes in monthly and annual mean temperature from 1971–2000 to 2011–2020 at station Helsinki Kaisaniemi (red dots) and in the ERA5 reanalysis at (60° N, 25° E) (blue crosses), together with the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 99th percentiles of the forecast distribution (box plots). The numbers in the two bottom rows give the observed temperature change at Helsinki Kaisaniemi and its percent rank in the forecast distribution



making CRPS equal to the mean absolute difference between the forecast and the verifying observation.

The continuous ranked probability skill score

$$\text{CRPSS} = 1 - \frac{\text{CRPS}}{\text{CRPS}_{\text{ref}}} \quad (2)$$

requires a reference forecast. Here, CRPS_{ref} is CRPS for the surrogate no-forced-change probabilistic forecast. Thus, CRPSS serves to quantify how much better or worse the actual forecast was than the no-forced-change forecast.

CRPS and CRPSS were also computed in cross-verification mode, in which climate changes in each of the 21 CMIP3 models were compared, in turn, with the probabilistic forecasts obtained from the other 20 models. Then, the number of CMIP3 models was calculated for which the cross-verification scores were better (that is, CRPS lower and CRPSS higher) than the real-world ones. These numbers are denoted as $n\text{-CRPS}$ and $n\text{-CRPSS}$, and their possible values range from 0 (real-world forecasts better than all forecasts in cross verification) to 21 (real-world forecasts worse than all forecasts in cross verification).

3 Results for temperature

As an introductory example, monthly and annual mean temperature changes in the grid box (60° N, 25° E) (same as used for illustration in RR06) are shown in Fig. 1. In this case, two observational estimates are given, one based on local measurements at station Helsinki Kaisaniemi (60° 11' N, 24° 57' E; data available from <https://en.ilmatieteenlaitos.fi/download-observations>) and the other on the ERA5 reanalysis. The figure highlights several points of interest:

1. The local station observations and the ERA5 reanalysis agree well with each other. This reflects the large spatial scale of temperature variability (Hansen and Lebedeff 1987) and the assimilation of near-surface temperature observations in ERA5 (Hersbach et al. 2020).
2. The observed annual mean warming at this location (station observations 1.6 °C, ERA5 1.5 °C) exceeds the median of the probabilistic forecast (0.9 °C) but is still well within the forecast distribution (at the 89th percentile for the station observations).
3. Following the statistical expectation for a reliable probabilistic forecast system, the temperature changes in 11 months out of 12 fall in the 5–95% forecast range. In September, the observed warming exceeds the 95th percentile of the forecast.
4. The forecast distributions for the monthly mean temperature changes are wide, allowing for at least a 10% chance of cooling in the individual months. Furthermore, as is typical for Northern Hemisphere high-latitude areas, both the median warming and the width of the distribution are larger in winter than in summer. The forecast distribution for the annual mean temperature change is narrower than those for the monthly changes, leading to a larger forecasted probability of warming (in this case, 95%). This mainly reflects the smaller internal variability in annual than monthly mean temperatures.

A global comparison between the observed (ERA5) and forecasted annual mean temperature changes is provided in Fig. 2; selected results for the individual months are included in Figs. S1–S2. The globally averaged temperature change in ERA5 is the same as the mean warming in the probabilistic forecast (0.62 °C), but there are substantial differences between the geographical details of the observed and the forecasted change (Fig. 2a, b). The observed warming is

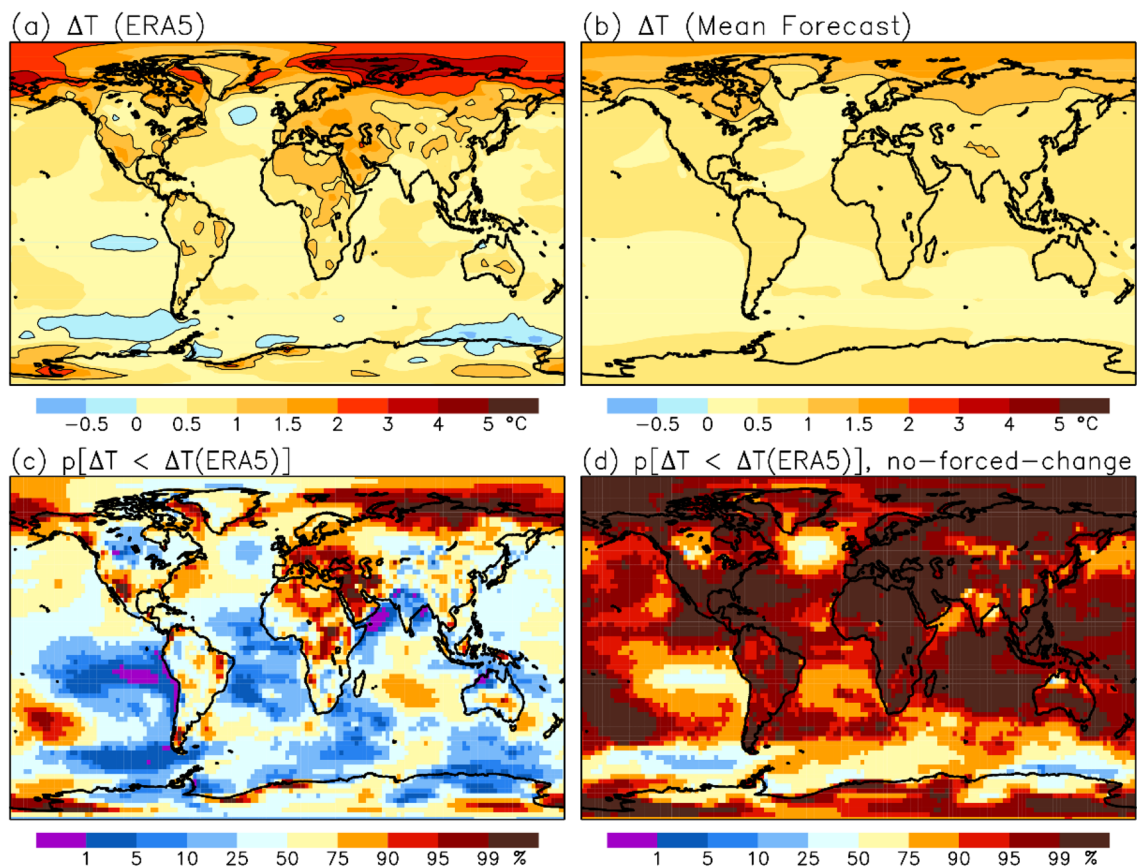


Fig. 2 Top: changes in annual mean temperature from 1971–2000 to 2011–2020 **a** in ERA5 and **b** from the mean of the forecast distribution. Bottom: percent rank of the ERA5 temperature change within

(c) the actual forecast distribution (d) and the surrogate no-forced-change distribution

more strongly focused in the Northern Hemisphere and has much larger spatial variability than the mean projection. The observed change varies from local cooling of 0.7 °C over the Southern Ocean to a warming of 5.2 °C over the Barents Sea, whereas the projected warming only varies within the range 0.2–1.9 °C. There are, however, many qualitative similarities in the geographical patterns, such as the maximum of warming in the Arctic and the minima in the Southern Ocean and northern North Atlantic. Therefore, the spatial correlation between the two fields is high ($r=0.82$).

Proceeding to a probabilistic view on verification, the percent rank of the observed temperature changes within the forecast distribution is shown in Fig. 2c. For example, over the eastern tropical Pacific that has received attention for its role in the recent “global warming hiatus” (Kosaka and Xie 2013; England et al. 2014; Stolpe et al. 2020), the observed change is in the lower end of the forecast distribution. On the other hand, in northern Siberia and in individual grid boxes in a few other regions, the observed warming exceeds the 99th percentile of the forecast. Despite these specific areas where the observed change was poorly anticipated

by the probabilistic forecast, the distribution of the percent ranks as a whole is reasonably well-balanced (as quantified below). This is clearly not the case when the observed changes are compared with the no-forced-change forecast (Fig. 2d). In large parts of the world, the observed temperature change exceeds the 95th or even the 99th percentile of this distribution.

To evaluate the reliability of the forecasts, percent rank histograms for annual (Fig. 3a) and monthly (Fig. 3b) temperature changes were constructed as described in Sect. 2.3. For comparison, the same analysis was repeated in cross-verification mode, comparing the changes in each of the 21 CMIP3 models against the forecasts obtained from the remaining 20 models. In Fig. 3a, b, the approximate 25th and 75th percentiles of the cross-verified rank fractions are shown with dashed red lines, whereas the dashed purple lines show the lowest and highest fractions.

The percent rank histograms for temperature change are slightly bottom-heavy, with the observed annual (monthly) mean warming falling below the median forecast in 57% (56%) of cases (Fig. 3a, b). The apparent discrepancy with

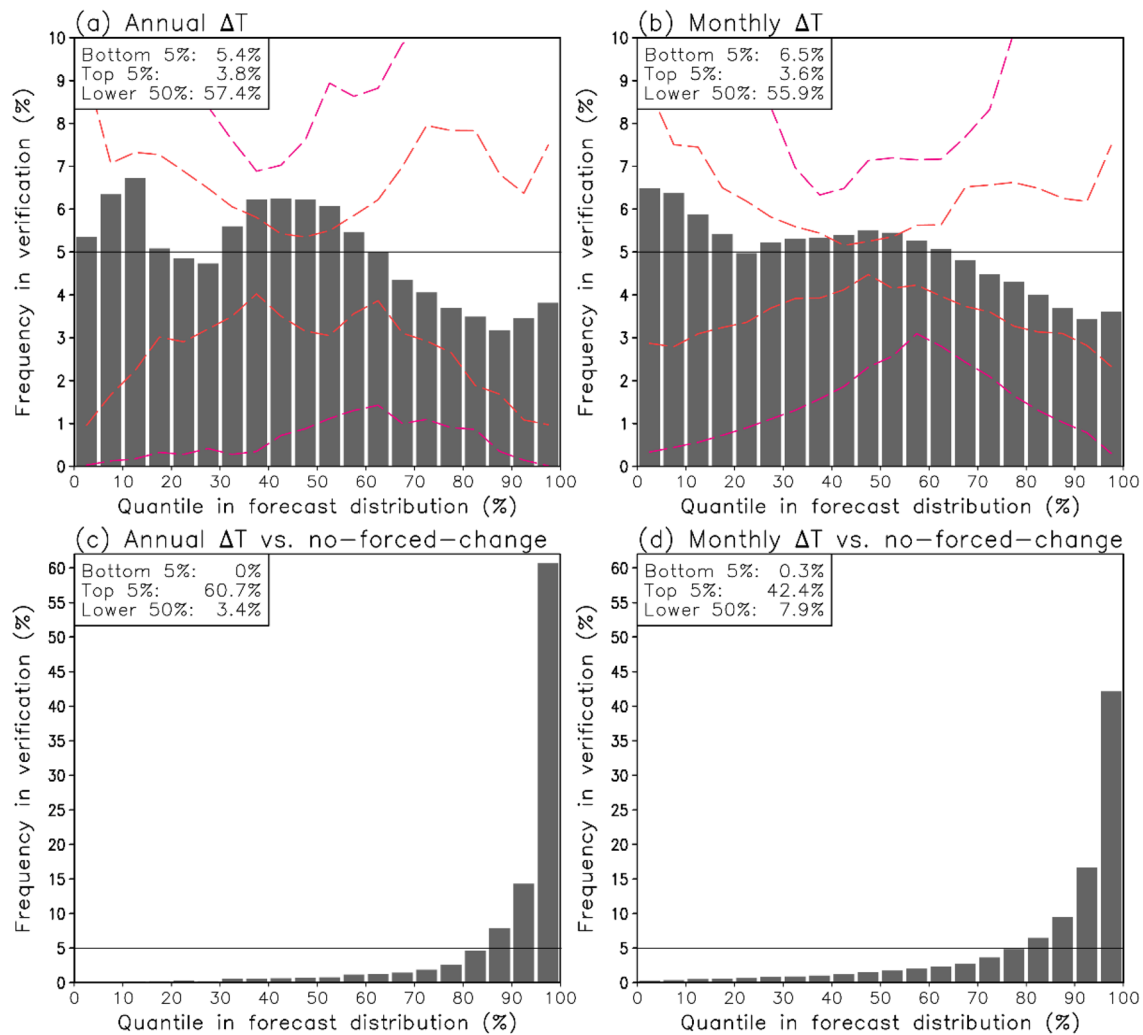


Fig. 3 Frequency histogram of the percent rank of the ERA5 annual (left) and monthly (right) mean temperature changes within the actual forecast distribution (top) and the surrogate no-forced-change distribution (bottom), using 5% bin width. Selected statistics are given in

the top-left corner of the figure panels. In **a, b**, the red lines indicate the approximate 25–75% range in cross verification between CMIP3 models, and the purple lines the minimum and maximum among the 21 models

the mentioned close agreement in global average warming between ERA5 and the mean forecast is explained by the asymmetric distribution of the observed changes. There are many areas where the observed warming is slightly smaller than the mean or the median forecast, but in terms of the global mean temperature change, this is compensated by the much stronger Arctic maximum in the observed warming. Accordingly, the fraction of changes in the bottom 5% of the forecast distribution is larger than that in the top 5% (see the numbers in Fig. 3a, b), but the combined fraction of these cases (9% for annual and 10% for monthly mean changes) follows the statistical expectation. This is the case even for the 1st and 100th percentile of the forecast distribution, which are together populated by 2.1% and 2.3% of the annual and monthly mean changes (not shown). Compared with the cross-verification results, the actual percent rank

histograms are by no means unusual. The actual frequencies fall in most cases between the 25th and 75th percentiles of those found in the cross verification.

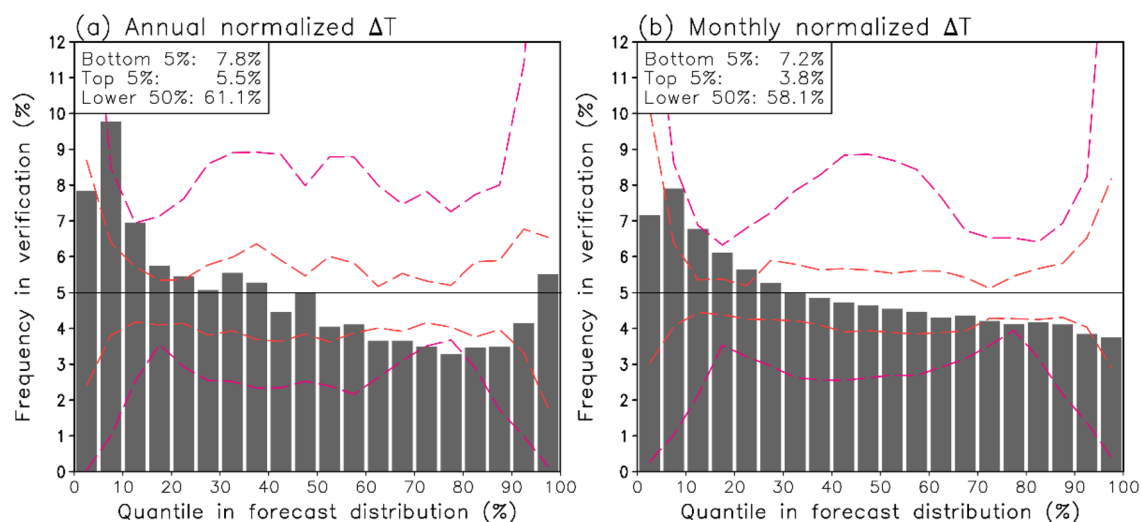
The percent rank histograms become dramatically different when comparing the observed temperature changes against the no-forced-change forecast, with a strong overrepresentation of the uppermost bins (Fig. 3c, d). However, this is slightly less pronounced for the monthly (Fig. 3d) than annual mean (Fig. 3c) temperature changes, due to the wider forecast distributions on the monthly time scale.

Similar results were obtained when using four alternative observational data sets. Just as for ERA5, a slight majority of the annual and monthly mean temperature changes in the HadCRUT, GISTEMP, NOAA and Berkeley Earth analyses fall in the lower half of the forecast distribution (row 1 in Table 2). The fraction of temperature changes within either

Table 2 Verification statistics for temperature change

Data set	Annual					Monthly				
	E	H	G	N	B	E	H	G	N	B
Coverage (%)	100	99.6	99.6	92.6	100	100	99.6	99.6	92.6	100
Lower 50% (%)	57.4	57.7	52.8	55.8	58.3	55.9	55.4	54.2	56.8	55.3
Outside 5–95% (%)	9.2	5.6	11.9	11.6	7.6	10.1	6.7	9.4	9.0	8.9
CRPS (°C)	0.170	0.148	0.167	0.150	0.159	0.257	0.212	0.231	0.225	0.235
n-CRPS	11	4	11	7	5	11	1	5	8	6
CRPSS (0.001)	658	691	661	674	671	495	551	526	518	516
n-CRPSS	9	6	9	9	8	10	8	9	10	10

Data sets: *E*ERA5, *H*HadCRUT, *G*GISTEMP, *N*NOAA, *B*Berkeley Earth. The first two data rows give the frequency of verification cases below the median and in either the bottom or top 5% of the forecast distribution. See the text for the definition of the other statistics. CRPS for NOAA is biased low due to lack of data in polar areas

**Fig. 4** As Fig. 3a, b, but for temperature changes normalized by the global annual mean warming

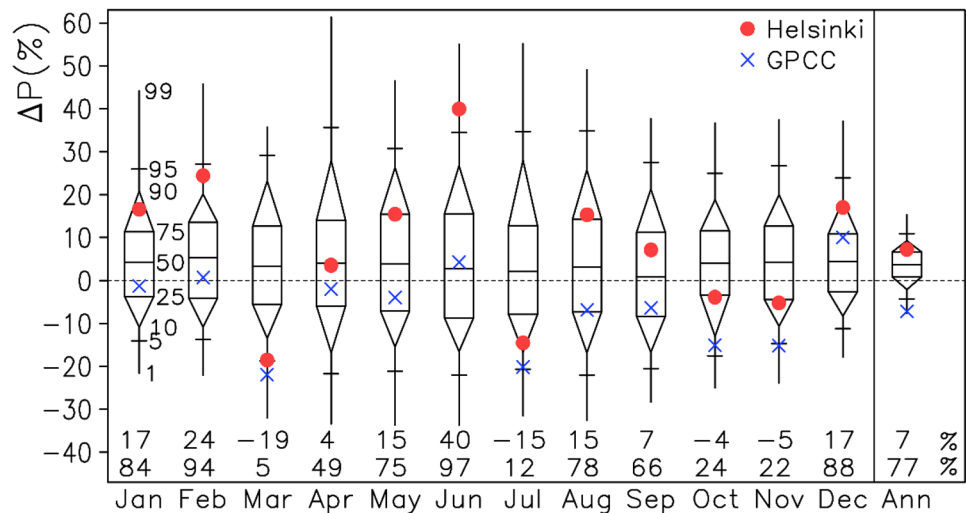
the bottom or top 5% of the forecast distribution is otherwise close to 10%, but lower for HadCRUT (5.6% for annual and 6.7% for monthly changes; row 2). The distribution of temperature changes is smoother in HadCRUT than in the other data sets, with less extreme local minima and maxima (Fig. S3). Accordingly, the changes in this data set fall less commonly in the tails of the forecast distribution.

The results for CRPS and CRPSS (Eqs. 1–2) are given in the lower half of Table 2. Regardless of the data set, CRPSS is higher for annual (0.658–0.691) than monthly mean temperature changes (0.495–0.551), reflecting the higher signal-to-noise ratio of the former. The lowest CRPS and highest CRPSS are found when using HadCRUT for verification, apparently because of the muted variability in this data set. Moreover, *n*-CRPS (1–11 depending on data set and time resolution) and *n*-CRPSS (6–10) are both in the midrange or lower half of their possible values. Thus, in addition to being far more accurate than the surrogate no-forced-change

forecasts, the probabilistic forecasts for temperature change are about as accurate as expected from the variation between the CMIP3 models.

The good verification statistics for temperature change reflect, in part, the close agreement in global mean warming between the mean forecast (0.62 °C) and the observations (also 0.62 °C for ERA5). For comparison, the global mean temperature changes in the 21 individual CMIP3 models vary from 0.34 °C to 1.01 °C. To check how well the probabilistic forecast performs for the geographical patterns of temperature change, the forecasted and ERA5 temperature changes were normalized by their global annual mean. Unsurprisingly, this suppresses the variability in the probabilistic forecast, thus slightly increasing the frequency of verification cases in the tails of the distribution (Fig. 4). Still, the fraction of annual (monthly) mean changes that fall outside the 5–95% forecast range only increases from 9 to 13% (10–11%). These findings concur qualitatively,

Fig. 5 As Fig. 1, but for changes in precipitation. The blue crosses represent the precipitation changes at (60° N, 25° E) in the GPCC analysis



but not quantitatively, with those of van Oldenborgh et al. (2013). When comparing CMIP5 hindcasts against observations in years 1950–2011, they found the CMIP5 ensemble to produce a nearly flat rank histogram for absolute temperature trends. However, in their study, there was a much larger increase in the fraction of trends falling in the extreme ends of the CMIP5-based distribution when the variations in global warming were factored out. This difference may reflect both the different periods and different data sets used in the two studies.

4 Results for precipitation

Figure 5 repeats the analysis of Fig. 1 for precipitation changes in the grid box (60° N, 25° E). Just as for temperature, the changes inferred from the local station observations fall in most cases well within the forecast distribution, although the 19% decrease in March is at the 5th percentile and the 40% increase in June at the 97th percentile of the forecast. However, there are two important differences between the two variables:

1. The signal-to-noise ratio is lower for precipitation than temperature changes. The annual mean forecast indicates an 80% probability of increasing precipitation, compared with 95% probability of warming (see also Fig. 4 in RR06). Furthermore, although the median forecast is positive throughout the year, the forecasted probability of increasing monthly precipitation reaches, at most, 65% in January and December.
2. There is a major discrepancy between the local station observations and the second observational estimate (GPCC) included in Fig. 5, with systematically more negative changes in the latter. This appears to be

a local anomaly in the GPCC data set, contrasting with an increase in annual precipitation in most of northern Europe (Fig. 6a). Nevertheless, this example illustrates the complication that arises from observational uncertainty when evaluating forecasts of precipitation change.

The observed (GPCC over land at 60°S–90°N and ERA5 elsewhere) and forecasted annual mean precipitation change are compared in Fig. 6; some results for the individual months are shown in Figs. S4–S5. The observed change (Fig. 6a) is much more spatially variable than the mean forecast (Fig. 6b). Some common features still stand out, including for example drying in southern Africa and southern North America as well as a general increase in precipitation in high latitudes. The Pearson product-moment correlation between the two fields is 0.19 and the Spearman rank correlation (which is less sensitive to large percent changes in arid areas) is 0.29. Figure 6c suggests an excess of areas where the observed change falls in the upper tail of the forecast distribution, particularly over the oceans where ERA5 data are used. This tendency becomes more pronounced when comparing the observed change with the no-forced-change distribution (Fig. 6d), which excludes the mostly positive mean changes in the actual forecast. However, the contrast between the percent rank fields in Fig. 6c, d is much smaller than that for temperature change (Fig. 2c, d).

Figure 7a confirms that larger than expected fractions of the observed annual mean precipitation changes fall above the median (61%) and the 95th percentile (15%) of the forecast distribution. The fraction of changes in the lower tail (5% below the 5th percentile) follows the expectation. For the monthly precipitation changes (Fig. 7b), for which the forecast distributions are wider than for the annual mean change, the excess of upper-end values is smaller. On the

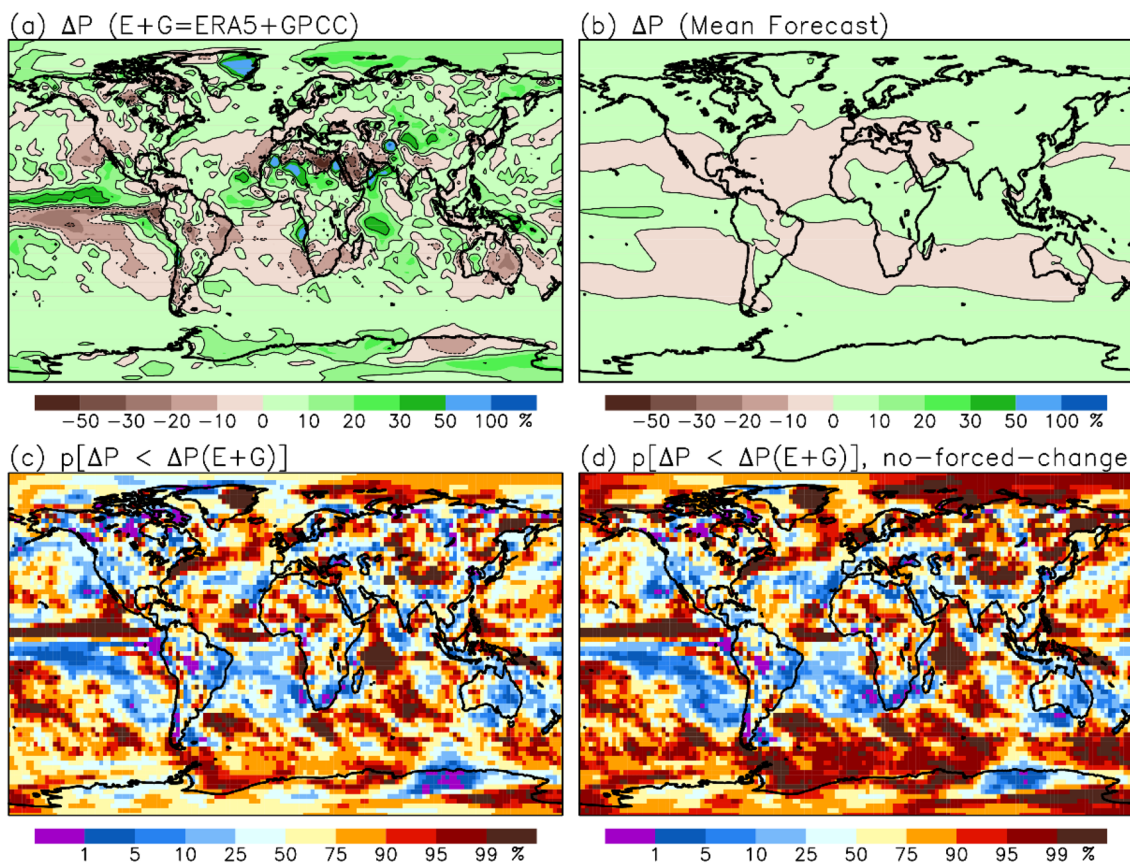


Fig. 6 As Fig. 2 but for changes in mean annual precipitation. The observational estimate **a** combines GPCC (land areas at 60° S–90° N) with ERA5 (oceans and Antarctica). In **a**, **b**, contours are drawn at 0, $\pm 10\%$ and $\pm 30\%$

other hand, the excess of upper-end values is exacerbated when replacing the actual forecast with the no-forced-change forecast (Fig. 7c, d), although the difference is small for the monthly changes.

However, the results in Figs. 6, 7 must be put in the context of observational uncertainty. ERA5 suggests a 4.8% increase in ocean mean precipitation from 1971–2000 to 2011–2020, far above the range of -0.1 to 1.6% in the 21 CMIP3 models (Table S5). Hersbach et al. (2020) likewise report an apparently excessive increase in ocean precipitation in ERA5 beginning in the 1990s. Over land, the mean change in GPCC (0.5%) is well within the range of the CMIP3 simulations (-0.7 to 3.5%) and in good agreement with CRU (0.6%) (Table S5, the quoted land means exclude Antarctica). The land area rank correlation between GPCC and CRU (0.59) is also higher than those for other pairs of data sets (Table S6), but many differences in the geographical details of precipitation change occur even between GPCC and CRU (Fig. S6).

Verification statistics of precipitation change against different observational data sets are reported for the global

area in Table 3 and separately for land at 60° S–90° N in Table 4. In the global domain, more than 50% of the changes in all three reanalyses exceed the median forecast (row 1 in Table 3). Yet, the fraction of changes falling outside the 5–95% forecast range is much larger for JRA-55 and NCEP than for ERA5 (row 2 in Table 3), which most likely reflects larger inhomogeneities in these older reanalyses. In NCEP, in particular, spuriously large positive and negative precipitation changes are widespread (Fig. S6e) and the changes are poorly correlated with ERA5 and JRA-55 (Table S6). Similar conclusions for the fraction of reanalysed precipitation changes in the tails of the forecast distribution also hold for land areas (row 2 in Table 4). However, excluding the annual mean in NCEP, the reanalysed precipitation changes over land fall more commonly below than above the median forecast (row 1 in Table 4). This is most striking for ERA5, in which the annual mean precipitation change is below the median forecast in 66% of the non-Antarctic land area, and the area mean change for this domain is -3.4% (Table S5). The reasons for this apparently unrealistic decrease in land precipitation in ERA5 are not clear (Hersbach et al. 2020).

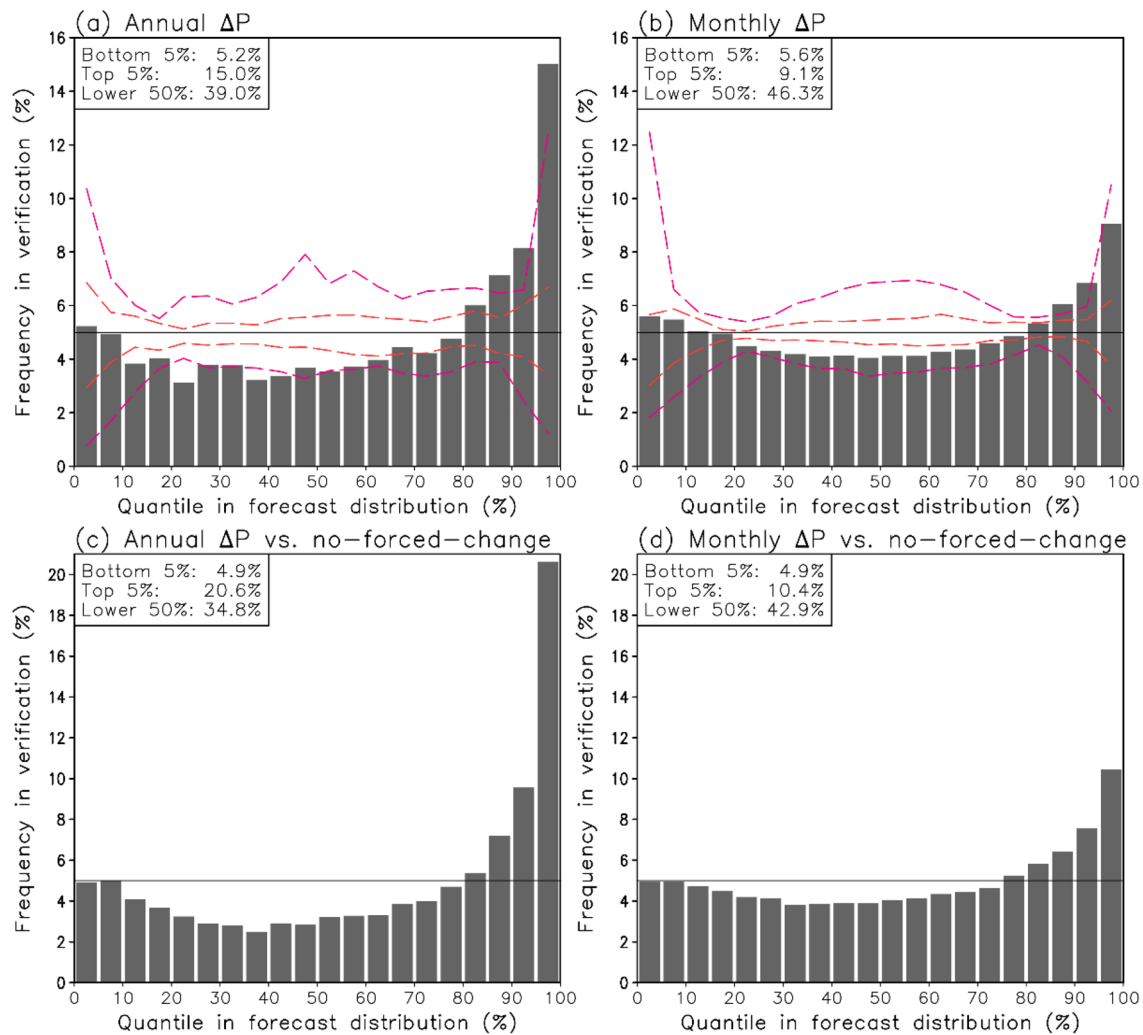


Fig. 7 As Fig. 3, but for changes in precipitation

Table 3 Global verification statistics for precipitation change

Data set	Annual				Monthly			
	E+G	E	J	N	E+G	E	J	N
Lower 50% (%)	39.0	43.7	46.7	43.0	46.3	49.1	49.3	47.5
Outside 5–95% (%)	20.3	22.2	32.8	52.7	14.7	14.8	21.2	29.6
CRPS (%)	5.80	5.89	9.79	12.74	13.50	13.23	17.00	19.56
n-CRPS	19	19	21	21	18	17	20	21
CRPSS (0.001)	80	61	44	5	19	11	12	6
n-CRPSS	9	11	13	18	9	13	13	20

Data sets: *E* + *G* GPCC over land at 60° S–90° N and ERA5 elsewhere, *E* ERA5, *J* JRA-55, *N* NCEP. In the calculation of CRPS, an upper limit of 200% was applied to precipitation changes

Verification against the gauge-based GPCC and CRU data sets gives a more positive impression of the forecast performance for precipitation in land areas. Only 14.2% (13.2%) of the annual and 10.3% (7.4%) of the monthly precipitation changes in GPCC (CRU) fall outside the

5–95% forecast range (row 2 in Table 4). However, the magnitude of precipitation changes in these gauge-based analyses might be reduced by relaxation to climatology in areas with insufficient station coverage (Becker et al. 2013; New et al. 2000; Harris et al. 2020). This issue is

Table 4 Verification statistics for precipitation change, land at 60°S–90°N

Data set	Annual					Monthly				
	E	J	N	G	C	E	J	N	G	C
Lower 50% (%)	66.1	55.2	47.4	46.9	47.7	60.5	54.5	50.3	49.0	47.3
Outside 5–95% (%)	27.7	39.2	52.3	14.2	13.2	15.5	20.7	28.8	10.3	7.4
CRPS (%)	7.28	14.15	16.76	5.77	4.91	17.15	21.36	26.07	15.41	12.56
n-CRPS	19	21	21	16	8	14	21	21	11	2
CRPSS (0.001)	−6	0	3	70	76	−7	−1	7	22	26
n-CRPSS	15	15	15	8	7	21	19	13	6	4

Data sets: ERA5, JJA-55, NCEP, GPCC, CRU. In the calculation of CRPS, an upper limit of 200% was applied to precipitation changes

particularly pertinent to CRU, which is based on a smaller number of station records than GPCC in the most recent decades (Dai and Zhao 2017). Conversely, inhomogeneities due to unaccounted-for changes in station network and instrumentation would have the opposite effect. Inhomogeneities that have a similar effect throughout the year should have a relatively larger impact on the distribution of annual than monthly precipitation changes, due to the smaller variability in annual precipitation. Consistent with this, the fraction of changes in the bottom and top 5% of the forecast distribution is larger for annual than monthly precipitation in all the observational data sets included in Tables 3 and 4.

As expected from the excess of verification cases in the tails of the forecast distribution, the CRPS values for precipitation change are mostly larger than those found in cross verification between the CMIP3 models (n -CRPS = 17–21 in the global domain, row 4 in Table 3). As an exception, n -CRPS = 2 for monthly precipitation over land verified against CRU (row 4 in Table 4); however, as discussed above, this data set may underestimate the magnitude of precipitation changes. Nevertheless, the forecast appears to have some skill in comparison with the reference no-forced-change forecast. For the favored choice of data sets (GPCC over non-Antarctic land, ERA5 elsewhere) in the global domain, CRPSS = 0.080 for annual and 0.019 for monthly precipitation, while the corresponding values in verification against GPCC over land are 0.070 and 0.022 (row 5 in Tables 3 and 4). While much lower than the corresponding values for temperature (Table 1), these values are in the upper half of those found in CMIP3 cross verification [bottom row of Tables 3 (columns E + G) and 4 (column G)]. From this perspective, the probabilistic forecast performs at its expected level of skill, although this level is modest particularly for monthly precipitation changes due to their very low signal-to-noise ratio.

5 Conclusions

In RR06, probabilistic forecasts of temperature and precipitation change between the periods 1971–2000 and 2011–2020 were presented. Here, these forecasts were verified against several observational data sets. Although several earlier studies have verified the ability of climate models to forecast changes in the global mean temperature in advance (Hausfather et al. 2020), this study is, to the best knowledge of the author, the first quantitative verification of forecasts of grid box scale climate changes. Due to the relatively low signal-to-noise ratio of the local climate changes, such a verification is currently much more meaningful in probabilistic than in deterministic terms. The results were encouraging, but with a distinct contrast between temperature and precipitation:

1. The forecasts of temperature change were found to be statistically reliable, with just 9% (10%) of the annual (monthly) mean changes in ERA5 falling beyond the 5–95% forecast range. The continuous ranked probability skill score (CRPSS) against a surrogate no-forced-change forecast was 0.658 for annual and 0.495 for monthly temperature changes when using ERA5 for verification. These results were largely insensitive to the choice of the verifying data set.
2. The forecasts of precipitation change are more difficult to verify due to the uncertainty in observations. The primary data sets selected in this study (GPCC over land at 60° S–90° N, ERA5 elsewhere) suggest that 20% (15%) of annual (monthly) precipitation changes fell outside the 5–95% forecast range, but such a result might also reflect inhomogeneity in the verifying data. Regardless, the skill of local precipitation change forecasts is limited by the low signal-to-noise ratio between forced change and internal variability. Reflecting this, the CRPSS values found for annual (0.080) and monthly precipitation (0.019) were much lower than those for temperature, yet

consistent with the values obtained in cross verification between the CMIP3 models.

As noted in Sect. 2.1, the variance within the RR06 probabilistic forecasts for the climate changes between 1971–2000 and 2011–2020 is strongly dominated by internal variability. In forecasts for later future periods, intermodel differences become gradually more important. I hope to revisit the verification of these forecasts in the early 2030s, once observations for the decade 2021–2030 are available.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-022-06182-8>.

Author contributions This research was entirely conducted by the first author.

Funding Open Access funding provided by University of Helsinki including Helsinki University Central Hospital. The Institute for Atmospheric and Earth System Research (INAR) is supported by Academy of Finland Flagship funding (Grant No. 337549).

Data and code availability The data and the Grid Analysis and Display System (GrADS) scripts needed to reproduce the figures in this article are available at <https://doi.org/10.5281/zenodo.5076067>.

Declarations

Conflict of interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Annan JD, Hargreaves JC (2010) Reliability of the CMIP3 ensemble. *Geophys Res Lett* 37:L02703. <https://doi.org/10.1029/2009GL014194>
- Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Schamm K, Schneider U, Ziese M (2013) A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. *Earth Syst Sci Data* 5:71–99. <https://doi.org/10.5194/essd-5-71-2013>
- Candille G, Talagrand O (2005) Evaluation of probabilistic prediction systems for a scalar variable. *Q J R Meteorol Soc* 131:2131–2150. <https://doi.org/10.1256/qj.04.71>
- Collins M et al (2013) Long-term climate change: projections, commitments and irreversibility. In: Stocker TF et al (eds) *Climate change 2013: the physical science basis*. Cambridge University Press, pp 1029–1136
- Dai A, Zhao T (2017) Uncertainties in historical changes and future projections of drought. Part I: estimates of historical drought changes. *Clim Change* 144:519–533. <https://doi.org/10.1007/s10584-016-1705-2>
- Deser C, Phillips A, Bourdette V, Teng H (2012) Uncertainty in climate change projections: the role of internal variability. *Clim Dyn* 38:527–546. <https://doi.org/10.1007/s00382-010-0977-x>
- England MH, McGregor S, Spence P, Meehl GA, Timmermann A, Cai W, Gupta AS, McPhaden MJ, Purich A, Santoso A (2014) Recent intensification of wind-driven circulation in the Pacific and the ongoing warming. *Nat Clim Change* 4:222–227. <https://doi.org/10.1038/nclimate2106>
- Flato G et al (2013) Evaluation of Climate Models. In: Stocker TF et al (eds) *Climate change 2013: the physical science basis*. Cambridge University Press, pp 741–866
- Frame D, Stone D (2013) Assessment of the first consensus prediction on climate change. *Nat Clim Change* 3:357–359. <https://doi.org/10.1038/nclimate1763>
- Gottelman A, Hannay C, Bacmeister JT, Neale RB, Pendergrass AG, Danabasoglu G, Lamarque J-F, Fasullo JT, Bailey DA, Lawrence DM, Mills MJ (2019) High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophys Res Lett* 46:8329–8337. <https://doi.org/10.1029/2019GL083978>
- Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. *J R Stat Soc B* 69:243–268. <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
- Hansen JE, Lebedeff S (1987) Global trends of measured surface air temperature. *J Geophys Res* 92:13345–13372. <https://doi.org/10.1029/JD092iD11p13345>
- Harris I, Osborn TJ, Jones P, Lister D (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data* 7:109. <https://doi.org/10.1038/s41597-020-0453-3>
- Hartmann DL et al (2013) Observations: atmosphere and surface. In: Stocker TF et al (eds) *Climate change 2013: the physical science basis*. Cambridge University Press, pp 159–254
- Hausfather Z, Drake HF, Abbott T, Schmidt GA (2020) Evaluating the performance of past climate model projections. *Geophys Res Lett* 47:e2019GL085378. <https://doi.org/10.1029/2019GL085378>
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90:1095–1107. <https://doi.org/10.1175/2009BAMS2607.1>
- Hawkins E, Sutton R (2010) The potential to narrow uncertainty in projections of regional precipitation change. *Clim Dyn* 37:407–418. <https://doi.org/10.1007/s00382-010-0810-6>
- Hawkins E, Joshi M, Frame D (2014) Wetter then drier in some tropical areas. *Nat Clim Change* 4:646–647. <https://doi.org/10.1038/nclimate2299>
- Hersbach H et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- IPCC (1990) *Climate change: the IPCC scientific assessment*. In: Houghton JT et al (eds) *Report prepared for intergovernmental panel on climate change by working group I*. Cambridge University Press, p 410
- IPCC (2001) *Climate change 2001: the scientific basis*. In: Houghton JT et al (eds) *Contribution of working group I to the third assessment report of the intergovernmental panel on climate change*. Cambridge University Press, p 881
- Jolliffe T, Stephenson DB (eds) (2011) *Forecast verification: a practitioner's guide in atmospheric science*, 2nd edn. Wiley, p 292

- Kalnay E et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–472. [https://doi.org/10.1175/1520-0477\(1996\)077%3c0437:TNYRP%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077%3c0437:TNYRP%3e2.0.CO;2)
- Kirtman B et al (2013) Near-term climate change: projections and predictability. In: Stocker TF et al (eds) *Climate change 2013: the physical science basis*. Cambridge University Press, pp 953–1028
- Knutti R (2008) Why are climate models reproducing the observed global surface warming so well? *Geophys Res Lett* 35:L18704. <https://doi.org/10.1029/2008GL034932>
- Kobayashi S et al (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *J Meteor Soc Japan* 93:5–48. <https://doi.org/10.2151/jmsj.2015-001>
- Kosaka Y, Xie SP (2013) Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* 501:403–407. <https://doi.org/10.1038/nature12534>
- LaJoie E, DelSole T (2016) Changes in internal variability due to anthropogenic forcing: a new field significance test. *J Clim* 29:5547–5560. <https://doi.org/10.1175/JCLI-D-15-0718.1>
- Lenssen N, Schmidt G, Hansen J, Menne M, Persi A, Ruedy R, Zyss D (2019) Improvements in the GISTEMP uncertainty model. *J Geophys Res Atm* 124:6307–6326. <https://doi.org/10.1029/2018JD029522>
- Mauritsen T et al (2019) Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO₂. *J Adv Model Earth Syst* 11:998–1038. <https://doi.org/10.1029/2018MS001400>
- Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE (2007) The WCRP CMIP3 multimodel dataset: a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–1394. <https://doi.org/10.1175/BAMS-88-9-1383>
- Morice CP, Kennedy JJ, Rayner NA, Winn JP, Hogan E, Killick RE, Dunn RJH, Osborn TJ, Jones PD, Simpson IR (2021) An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *J Geophys Res Atm* 126:e2019JD032361. <https://doi.org/10.1029/2019JD032361>
- Nakićenovic N, Swart R (eds) (2000) *Special report on emissions scenarios: a special report of working group III of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, p 599
- New M, Hulme M, Jones PD (2000) Representing twentieth-century space–time climate variability. Part II: development of 1901–96 monthly grids of terrestrial surface climate. *J Clim* 13:2217–2238. [https://doi.org/10.1175/1520-0442\(2000\)013%3c2217:RTCSTC%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3c2217:RTCSTC%3e2.0.CO;2)
- Nurmi P (2003) Recommendations on the verification of local weather forecasts. ECMWF Techn Memo 430:19. <https://doi.org/10.21957/y1z1thg5l>
- Rahmstorf S, Cazenave A, Church JA, Hansen JE, Keeling RF, Parker DE, Somerville RCJ (2007) Recent climate observations compared to projections. *Science* 316:709–709. <https://doi.org/10.1126/science.1136843>
- Räisänen J (2007) How reliable are climate models? *Tellus* 59A:2–29. <https://doi.org/10.1111/j.1600-0870.2006.00211.x>
- Räisänen J, Ruokolainen L (2006) Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus* 58A:461–472. <https://doi.org/10.1111/j.1600-0870.2006.00189.x>
- Rohde RA, Hausfather Z (2020) The Berkeley earth land/ocean temperature record. *Earth Syst Sci Data* 12:3469–3479. <https://doi.org/10.5194/essd-12-3469-2020>
- Schmidt G, Bader AD, Donner LJ, Elsaesser GS, Golaz J-C, Hannay C, Molod A, Neale RB, Saha S (2017) Practice and philosophy of climate model tuning across six US modeling centers. *Geosci Model Dev* 10:3207–3223. <https://doi.org/10.5194/gmd-10-3207-2017>
- Schneider U, Becker A, Finger P, Rustemeier E, Ziese M (2020) GPCC full data monthly product version 2020 at 0.5°: monthly land-surface precipitation from rain-gauges built on GTS-based and historical data. https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_050
- Shiogama H, Stone DA, Nagashima T, Nozawa T, Emori S (2013) On the linear additivity of climate forcing-response relationships at global and continental scales. *Int J Climatol* 33:2542–2550. <https://doi.org/10.1002/joc.3607>
- Stolpe MB, Cowtan K, Medhaug I, Knutti R (2020) Pacific variability reconciles observed and modelled global mean temperature increase since 1950. *Clim Dyn* 56:613–634. <https://doi.org/10.1007/s00382-020-05493-y>
- Stouffer RJ, Manabe S (2017) Assessing temperature pattern projections made in 1989. *Nat Clim Change* 7:163–165. <https://doi.org/10.1038/nclimate3224>
- Stouffer R, Manabe S, Bryan K (1989) Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO₂. *Nature* 342:660–662. <https://doi.org/10.1038/342660a0>
- Suckling E (2018) Seasonal-to-decadal climate forecasting. In: Troccoli A (ed) *Weather and climate services for the energy industry*. Palgrave Macmillan, pp 123–137. https://doi.org/10.1007/978-3-319-68418-5_9
- Tebaldi C, Arblaster J (2014) Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Clim Change* 122:459–471. <https://doi.org/10.1007/s10584-013-1032-9>
- van Oldenborgh GJ, Doblas Reyes FJ, Drijfhout SS, Hawkins E (2013) Reliability of regional climate model trends. *Environ Res Lett* 8:014055. <https://doi.org/10.1088/1748-9326/8/1/014055>
- Zhang H-M, Lawrimore JH, Huang B, Menne MJ, Yin X, Sánchez-Lugo A, Gleason BE, Vose R, Arndt D, Rennie JJ, Williams CN (2019) Updated temperature data give a sharper view of climate trends. *Eos*. <https://doi.org/10.1029/2019EO128229>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.