Tellus (2006), 58A, 461–472
Printed in Singapore. All rights reserved

TELLUS

Probabilistic forecasts of near-term climate change based on a resampling ensemble technique

By J. RÄISÄNEN* and L. RUOKOLAINEN, *Department of Physical Sciences, Division of Atmospheric Sciences, P.O. Box 64, FIN-00014 University of Helsinki, Finland*

(Manuscript received 3 January 2006; in final form 8 March 2006)

ABSTRACT

Probabilistic forecasts of near-term climate change are derived by using a multimodel ensemble of climate change simulations and a simple resampling technique that increases the number of realizations for the possible combination of anthropogenic climate change and internal climate variability. The technique is based on the assumption that the probability distribution of local climate changes is only a function of the all-model mean global average warming. Although this is unlikely to be exactly true, cross-verification indicates that the resulting biases are more than compensated by the increased sample size provided by the method. The resulting forecasts for southern Finland suggest a 95% probability of annual mean warming and an 80% probability of increased annual mean precipitation from the years 1971–2000 to 2011–2020 under the SRES A1B emissions scenario. The choice of the emissions scenario is unimportant for such short-term forecasts but becomes gradually more important towards the late 21st century. The simulations also suggest that the probability of near-term warming that is large enough to be discernible from internal variability is largest in the tropics where internal temperature variability is small, not in the Arctic where the average model-simulated warming is largest.

1. Introduction

Projections of anthropogenic climate change often have a long-time perspective. One of the best-known results from the Third Assessment Report of the Intergovernmental Panel on Climate Change (Houghton et al., 2001) is the prediction that the global mean temperature in the year 2100 would be 1.4°–5.8°C higher than in the year 1990. A centennial timescale has also been common in regional studies, such as in the PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects) project (Christensen et al., 2002) in which several regional climate models were used to simulate European climate changes from 1961–1990 to 2071–2100.

A focus on a relatively distant future is partly motivated by the needs of climate policy. To estimate the urgency of emission reductions, information on the possibly severe long-term consequences of unmitigated greenhouse gas emissions is needed. However, there is also an important technical viewpoint. The larger the greenhouse-induced climate changes grow, the easier it is to distinguish them from natural variability. Conversely, natural variability is sometimes considered to make short-term (say, 10–20 yr) climate projections too uncertain to be of any practical value. Yet, when considering adaptation to climate changes, century-scale climate projections are directly relevant only for sectors such as forestry and water authorities that have a very long planning horizon. For many parts of the society, shorter-term projections would be more useful, among other things for giving people a realistic idea of the climate changes expected in their own lifetimes.

Here, we explore prospects of near-term climate change in probabilistic terms, using a multimodel ensemble of recent coupled atmosphere-ocean general circulation model (AOGCM) simulations. The climate change projections obtained from the individual models differ both because the models are different and because each model simulation has its own realization of unforced natural variability (e.g. Räisänen, 2001a). Thus, to the extent that differences between the models cover the uncertainty in modelling climate processes and the amplitude of the simulated variability is realistic, the multimodel ensemble gives a plausible sample of the climate changes that may occur under a given forcing scenario. However, the limited number of models included in the ensemble (in our case, at best 21) makes the sample rather small for estimating, in particular, the tails of the underlying probability distribution. A simple resampling methodology is, therefore, used to enlarge the samples. Objective

*Corresponding author.
e-mail: jouni.raisanen@helsinki.fi
DOI: 10.1111/j.1600-0870.2006.00189.x

Table 1. The models used in this study. Models marked with a star (∗) did not provide data for the B1 and/or A2 scenario(s) and are, therefore, excluded in Section 7. For more details, see http://www-pcmdi.llnl.gov

Model	Institution
CCSM3	National Center for Atmospheric Research, USA
CGCM3.1 (T47)	Canadian Centre for Climate Modelling and Analysis
*CGCM3.1 (T63)	Same as previous
CNRM-CM3	Météo-France
CSIRO-MK3.0	CSIRO Atmospheric Research, Australia
ECHAM5/MPI-OM	Max Planck Institute (MPI) for Meteorology, Germany
ECHO-G	University of Bonn and Model & Data Group, Germany; Korean Meteorological Agency
*FGOALS-g1.0	Chinese Academy of Sciences
GFDL-CM2.0	Geophysical Fluid Dynamics Laboratory, USA
GFDL-CM2.1	Same as previous
*GISS-AOM	Goddard Institute for Space Studies, USA
*GISS-EH	Same as previous
GISS-ER	Same as previous
INM-CM3.0	Institute for Numerical Mathematics, Russia
IPSL-CM4	Institut Pierre Simon Laplace, France
*MIROC3.2 (hires)	Center for Climate System Research, National Institute for Enviromental Studies and Frontier Research Center for Global Change, Japan
MIROC3 (medres)	Same as previous
MRI-CGCM2.3.2	Meteorological Research Institute, Japan
PCM	National Center for Atmospheric Research, USA
UKMO-HadCM3	Hadley Centre for Climate Prediction and Research / Met Office, UK
*UKMO-HadGEM	Same as previous

probabilistic cross-verification indicates that the increased sample size outweighs biases eventually caused by the resampling itself.

In the next two sections of this paper, we describe the model simulations used for this study and the resampling technique used to increase the sample size. In Section 4, the performance of the technique is compared in a cross-verification mode with the standard technique of only using each simulation once. Section 5 evaluates the success of the technique in hindcasting the climate changes that occurred from 1961–1990 to 1991–2000. The next section gives probabilistic forecasts of climate change for the early 21st century and later periods, based on the SRES A1B emissions scenario. Most of the examples are for southern Finland but selected results are also shown for other areas to highlight, in particular, the implications of the geographically varying amplitude of internal climate variability. In Section 7, the sensitivity of the findings to the choice of the emissions scenario is studied. Finally, the results are summarized and some issues related to their interpretation are discussed in Section 8.

2. Data set

We use a new set of AOGCM simulations performed for the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR4) (Table 1). First, the data set includes '20th Century Climate in Couple Climate Models' (20C3M) simulations covering the 20th century and forced by a mixture of anthropogenic and (in most models) natural forcing factors. The details of the forcing vary with model, but all models include at least the increase in major anthropogenic greenhouse gases and some representation of anthropogenic aerosols. Second, there are 21st century simulations with anthropogenic greenhouse gas and aerosol forcing based on the SRES A1B, B1 and A2 emissions scenarios (Nakićenović and Swart, 2000). Altogether, there are 21 models with data for both the 20C3M and A1B simulations. Of these, 15 also have data for both the B1 and A2 scenarios (Table 1). Although parallel runs started from slightly different initial conditions are available for some models, only one model run per one forcing scenario is used in this study. Third, the data set also includes long (at least 200 yr for almost all models) control simulations with constant external conditions. These simulations are used in Sections 5 and 6 to estimate the probability distribution of climate changes in the absence of external forcing.

Throughout this study, we assume that all models give equally plausible simulations of climate change and thus deserve the same weight in the calculations. Alternatives to this simple method have been proposed (e.g. Giorgi and Mearns, 2002; 2003) but are not considered here. Note that weighting between models is less of an issue for near-term than long-term climate changes because, for weak forcing, the uncertainty in climate changes is dominated by internal variability rather than by model

differences (Räisänen, 2001a). Similarly, while observations do not allow us to exclude the possibility that the sensitivity of the real climate system to increasing greenhouse gas forcing would be outside the range of current model results (Allen and Ingram, 2002; Andreae et al., 2005), this is less of a problem for short- than for long-term forecasts.

3. Methods

We focus on decadal means of temperature and precipitation in the early 21st century. Although 30 yr means are standard in the definition of climatic means, we prefer the decadal scale here because it is probably better compatible with the perception of climate by lay people. However, 30 yr means for the period 1971–2000 are used for the definition of the baseline climate from which the changes are calculated.

An example of the simulated near-term climate changes under the A1B forcing scenario—the annual mean temperature and precipitation changes from 1971–2000 to 2011–2020 in a grid box in southern Finland (60°N, 25°E)—is given by the 21 black markers in Fig. 1. The markers are strongly concentrated in the upper right quadrant representing increased temperature and precipitation. Thus, at this time already, the greenhouse-gas-dominated forcing has a large impact on the distribution of the model results. Yet, 21 simulations is not a large sample for estimating the underlying distribution that would be obtained if an infinite number of model simulations from the same hypothetical population were analysed. A sample of this size gives essentially no information below the 5th and above the 95th percentile of the distribution, and substantial sampling errors occur even in the middle of the distribution. For example, if a given value of climate change is exceeded with a probability of 50% in the underlying distribution, the root-mean-square error when estimating the probability of this value from a sample of 21 is $[1/2 \times (1 - 1/2) / 21]^{1/2} \approx 11\%$.

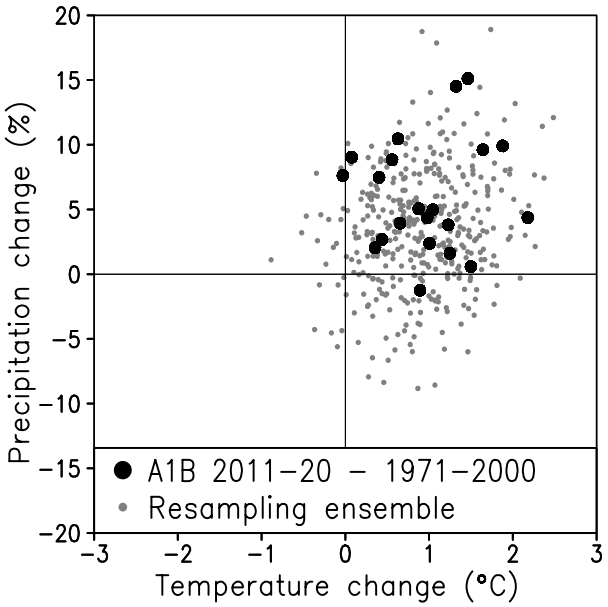

Fig. 1. Changes from 1971–2000 to 2011–2020 in annual mean temperature (°C, horizontal axis) and precipitation (%) in southern Finland (60°N, 25°E). Black markers: as simulated by 21 models under the SRES A1B emissions scenario. Grey markers: as inferred from the resampling ensemble.

To increase the sample size, a simple resampling method is used. This method is based on the finding that, apart from the random noise associated with internal variability, model-simulated climate changes tend to scale quasi-linearly with the change in global mean temperature (e.g. Santer et al., 1990; Mitchell et al., 1999; Huntingford and Cox, 2000; Mitchell, 2003; Harvey, 2004), at least for forcing scenarios dominated by increased greenhouse gas concentrations.

As an example, Fig. 2 shows the 21-model mean changes in annual mean temperature for two pairs of periods (from 1971–2000 to 2011–2020 and from 2051–2080 to 2086–2095) that share under the A1B scenario the same difference (0.62°C) in 21-model mean global mean temperature. The overall patterns of change are very similar in the two cases. Although some differences are visible (in particular, the warming over the Arctic Ocean is larger in the former than in the latter case), these are much smaller than the differences between the individual model simulations for either of the two pairs of periods. It is, therefore, reasonable to assume that temperature changes (which are always a combination of a model-dependent signal and internal variability) should follow approximately the same probability distribution in the two cases.

Thus, we assume that the probability distribution of climate changes is determined simply by the multimodel mean change in the global mean temperature. Under this assumption, we can form a resampling ensemble for the climate change from period *P1* (e.g. 1971–2000) to *P2* (e.g. 2011–2020) that includes all the pairs (*P3*, *P4*) (e.g. 2051–2080, 2086–2095) for which the multimodel mean global warming is the same as from *P1* to *P2*. In practice, we subsample *P4* with 5 yr interval and select for each *P4* the 30 yr period *P3*, which gives the closest match to the target multimodel mean global warming. *P3* and *P4* are not allowed to overlap, and the multimodel global mean warming is required to match the target value within 0.03°C.

From the time-series obtained from the 20C3M and A1B simulations (years 1901–2098), our algorithm finds 20 suitable pairs (*P3*, *P4*) for the target pair of periods (1971–2000, 2011–2020), the first being (1906–1935, 1991–2000), the second (1932–1961, 1996–2005), the second last (2047–2076, 2081–2090) and the last (2051–2080, 2086–2095). The number of available pairs for cases with *P2* later in the 21st century is smaller because the larger target global means warming in these cases cannot be achieved with *P4* much before *P2*. For example, for

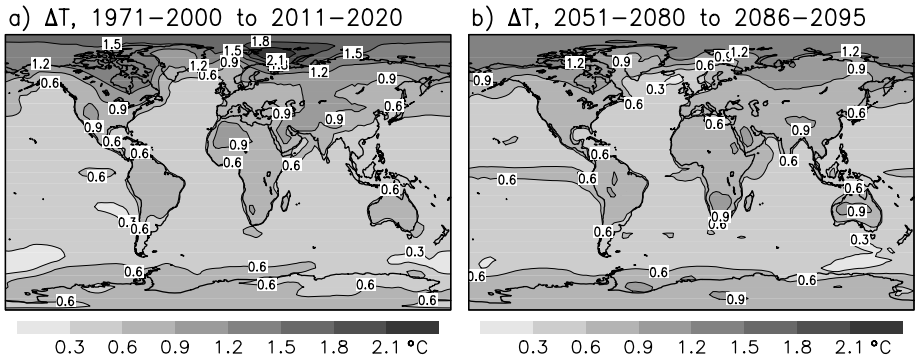

Fig. 2. 21-model mean changes in annual mean temperature under the A1B scenario (a) from 1971–2000 to 2011–2020 and (b) from 2051–2080 to 2086–2095.

$(P1, P2) = (1971–2000, 2041–2050)$ 13 pairs are found, whereas only 7 pairs are found for (1971–2000, 2071–2080).

In Fig. 1, the resulting $20 \times 21 = 420$ realizations of climate change for $(P1, P2) = (1971–2000, 2011–2020)$ are shown by the small grey markers. As expected, the increase in ensemble size reveals a wider range of possible temperature and precipitation changes than the simulated changes for the target pair of periods only. The probability of warming as inferred from the resampling ensemble is practically the same as inferred directly from the changes simulated from 1971–2000 to 2011–2020 (95%). The probability of precipitation increase is smaller for the resampling ensemble (80% versus 95%), but the difference is not statistically significant.

Because of the subsampling, our realizations of climate change are only partly independent. However, cross-verification with methods described in Section 4 indicated the subsampling to be a better alternative than the use of independent periods. Subsampling with 1 yr interval worked even better than the 5 yr interval, but the difference was small, obviously because the subsequent realizations of climate change become highly correlated in this case. Thus, the 5 yr interval was preferred to reduce the computations.

Our resampling method has similarities with the pattern-scaling technique first introduced by Santer et al. (1990). However, while pattern-scaling methods are generally designed for obtaining a good estimate of the noise-free climate change signal (e.g. Mitchell, 2003), our method aims to produce a realistic distribution for the combination of the signal and the noise caused by internal variability.

The resampling provides a computationally cheap alternative for exploring the uncertainty associated with *internal climate variability*, which can also be studied by running climate models several times with identical forcing but different initial conditions. This should not be mixed with the aims of the perturbed-parameter technique, which is currently applied on a massive scale in the ClimatePrediction.net project (Stainforth et al., 2005) and is hoped to cover *model-related uncertainty* more completely than is possible with traditional multimodel ensembles like the one used in this study.

4. Cross-verification

In comparison with the standard method of only using each model simulation once, the larger sample size allowed by the resampling reduces random errors in the derived probability estimates. On the other hand, the resampling may introduce systematic errors, since its basic assumption is unlikely to be exactly valid. Even for the same global mean warming, the patterns of climate change may vary for at least two reasons. First, the spatial patterns of forcing change with time, because the geographical distribution of aerosol forcing and its importance in relation to greenhouse gas forcing change with time under the SRES scenarios. Second, the regional distribution of the feedbacks that regulate the magnitude and distribution of climate changes may change as the global warming proceeds. For example, the ice-albedo feedback may become gradually less important with successively decreasing ice cover (Mitchell et al., 1999), which might in fact contribute to the smaller warming over the Arctic Ocean in Fig. 2(b) than 2(a). It is, therefore, necessary to ask whether the increased sample size outweighs the systematic errors eventually caused by the resampling.

To answer our question, we use cross-verification. The climate changes in one of the 21 models are chosen as a pseudo-truth, against which the probability distributions derived from the other model simulations either with the standard method or with the resampling ensemble are verified. This is then repeated for all choices of the verifying model and the verification statistics is averaged over all cases.

A basic verification statistic for probability forecasts is the Brier score B (Brier, 1950; Wilks, 1995). Consider a binary event E. For example, E might be defined to occur if the mean temperature in a given decade and location is at least 1°C above the normal. As evaluated over N cases

$$B = \sum_{i=1}^{N} m_i (p_i - o_i)^2 \Big/ \sum_{i=1}^{N} m_i, \qquad (1)$$

where $o_i = 1$ (0) if E occurs (does not occur) in case i and p_i is the corresponding forecast probability. m_i is the weight given to case i; here we weight grid boxes according to their area.

The better the forecast, the lower the Brier score. For a perfect deterministic forecast system, which gives a probability of 1 (0) always when E occurs (does not occur), $B = 0$.

The Brier score can be evaluated for any threshold value ξ used in the definition of E. Integrating B over ξ gives the continuous ranked probability score *CRPS* (Stanski et al., 1989; Herbach, 2000; Candille and Talagrand, 2005)

$$CRPS = \int_{-\infty}^{\infty} B(\xi)\mu(\xi)\, d\xi, \tag{2}$$

where $\mu(\xi)$ is a weighting function. Here, we use unity weighting ($\mu(\xi) = 1$) for temperature changes in the whole real axis. For precipitation changes, unity weighting is used from -100% to 200% but zero weighting above 200%, to avoid *CRPS* from being unduly affected by the huge per cent changes that occasionally occur in models in desert areas where absolute precipitation approaches zero.

We evaluated *CRPS* in cross-verification mode and averaged the resulting score over all choices of the verifying model and over the global domain. Examples of the results for the A1B scenario are shown in Table 2. The *CRPS* scores for the resampling ensemble are consistently lower than those for the standard method, although the improvement is smaller in the later periods (2041–2050 and 2071–2080) than in the early 21st century (2011–2020). As a sensitivity test we also repeated these calculations for absolute precipitation changes, which are well defined even in dry areas, and found almost exactly the same ratios of *CRPS* scores as for the relative changes (not shown).

The decreases in *CRPS* achieved by the resampling may appear small. However, most of the *CRPS* is attributable to the inherent variation of climate changes between models. Denoting the score that would be obtained if the probability distribution of the data were known exactly (e.g. from an infinite ensemble) as $CRPS_\infty$, the expected value of cross-verified *CRPS* when using a M-member forecast ensemble is

$$CRPS_M = \left(1 + \frac{1}{M}\right) CRPS_\infty, \tag{3}$$

when the M forecast values are independent from each other (Appendix). In our cross-verification, the forecast ensemble size for the standard method was $M = 20$ and from (3) $CRPS_\infty$ is only 4.8% below $CRPS_{20}$. From this perspective the decreases in *CRPS* achieved by the resampling are substantial, particularly for the period 2011–2020. The smaller relative decreases in the later periods probably relate to the facts that (1) the number of resampled realizations per model was smaller in the later periods and that (2) the directly model-related differences in climate change grow in magnitude with increasing forcing (Räisänen, 2001a). Because of (2), realizations obtained from a single model, which only sample the uncertainty due to internal variability, are less independent of each other for a stronger forcing.

Table 2. Cross-verification CRPS (see text) for the standard method, and the ratio between the CRPS scores for the resampling ensemble method and the standard method. Ratios below one indicate improvement. Results are shown for annual mean temperature and precipitation changes in the years 2011–2020, 2041–2050 and 2071–2080 under the A1B emissions scenario, relative to the baseline period 1971–2000. The ratios for seasonal and monthly climate changes are slightly lower (not shown)

	Temperature		Precipitation	
	Standard	Ratio	Standard	Ratio
2011–2020	0.188°C	0.963	4.34%	0.964
2041–2050	0.285°C	0.980	5.60%	0.973
2071–2080	0.417°C	0.987	7.04%	0.982

5. Verification against recent observed climate changes

Cross-verification only tests the sampling properties of a probabilistic forecast method. If the sample from which the forecast is derived is biased, for example if the models systematically over- or underestimate the magnitude of natural variability or if they misrepresent climate response to anthropogenic forcing, the real skill of the forecasts may be compromised. As a brief check of our method against reality, we compared the observed temperature and precipitation changes from 1961–1990 to 1991–2000 as presented by the Tyndall Centre TYN SC 2.0 data set (Mitchell et al., 2004) with the model-based probability distributions. In addition to the distributions derived from the forced 20C3M and A1B simulations, the corresponding distributions of 10 yr minus 30 yr climate changes from the unforced control simulations were also used in this exercise.

Almost all land areas covered by the TYN SC 2.0 data set experienced annual mean warming from 1961–1990 to 1991–2000 (Fig. 3a). This warming was consistent with the probability distributions estimated from the forced simulations (Fig. 3b): the fractional area with observed temperature change outside the 5–95% range of the model-based distribution was only 8%, of which 3% in the cold and 5% in the warm end of the distribution. Similar results were obtained in the individual 3 month seasons, with the total fraction of observed temperature changes outside the model-based 5–95% range varying from 6% in the boreal autumn to 14% in winter. Thus, the temperature forecasts were reliable in a probabilistic sense: the observed changes fell in the tails of the distributions approximately as often as on the average expected from pure chance. For precipitation changes, however, slight symptoms of unreliability were found (maps not shown). The fractional area with observed precipitation change outside the 5–95% range of the model-based distribution varied from 11% (summer) to 17% (winter and the annual mean

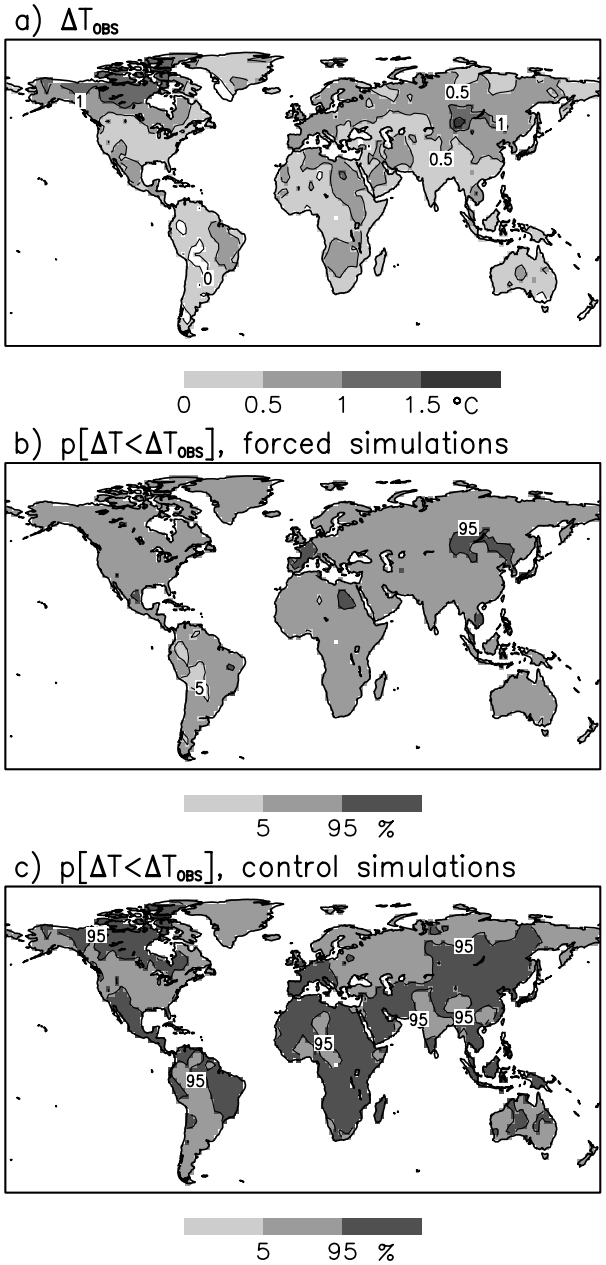

Fig. 3. (a) Observed (TYN SC 2.0, covering land at 60°S–90°N) annual mean temperature changes from 1961–1990 to 1991–2000, and grid boxes in which the observed changes fell in the bottom (lightest shading) or the top (darkest shading) 5% of the distribution derived from (b) the forced simulations and (c) the unforced control simulations.

change). This bears a suggestion that the models might underestimate precipitation variability on interdecadal timescales, but other factors such as observational errors and temporal inhomogeneities might also play a role.

The observed changes were also compared with the distribution of 10 yr minus 30 yr climate changes from the unforced control simulations. The observed annual mean temperature change was in the top 5% of this distribution in 58% of the studied area, but nowhere in the bottom 5% of the distribution (Fig. 3c). This is consistent with the findings of Karoly and Wu (2005), who showed that the recent warming has been statistically significant in a much larger part of the world than can be explained by internal climate variability. In marked contrast, the fraction of observed precipitation changes falling in the top and bottom 5% of the distribution was practically the same for the forced and the unforced simulations. This primarily reflects the much lower signal-to-noise ratio of precipitation than temperature changes but could also be affected by errors in the models and/or observations.

6. Probabilistic estimates of climate change under the A1B emissions scenario

Probabilistic forecasts of temperature and precipitation change for a grid box in southern Finland (60°N, 25°E) from the 30 yr period 1971–2000 to the decade 2011–2020, as derived with the resampling ensemble method, are shown in Fig. 4. The calculation suggests a 95% probability of annual mean warming and an 80% probability of increasing annual precipitation (rightmost whiskers in Figs. 4a and b), with a 5–95% uncertainty range of 0.0°–1.9°C (−4 to 11%) for temperature (precipitation) change. Uncertainty ranges for the changes in seasonal and particularly monthly 10 yr means are wider than those for the annual means, due to increasing variability with decreasing temporal averaging. However, for temperature the width of the distribution differs only slightly between the summer (June–July–August = JJA) mean and the annual mean.

For both temperature and precipitation, the best-estimate increases as inferred from the median of the distribution are largest in winter and smallest in summer, in agreement with earlier studies of simulated greenhouse gas induced climate change in northern Europe (e.g. Räisänen, 2001b). However, the uncertainty in temperature change is also much larger in winter than in summer, reflecting the seasonal contrast in the amplitude of temperature variability in this part of the world. The estimated probability of warming, therefore, varies only slightly with the time of the year, being approximately 90% for seasonal and 80–85% for monthly mean changes. For precipitation, the probability of increase is distinctly larger in winter than in summer, but a substantial probability of reduced precipitation remains in all seasons and calendar months.

As time proceeds, the best-estimate temperature and precipitation changes increase in magnitude, as shown for the annual means in Fig. 5. At the same time, the estimated probability distributions grow wider. This is because the uncertainty associated with model differences, which is initially small compared with internal variability (Räisänen, 2001a), increases with increasing magnitude of the greenhouse gas forcing. A small chance of precipitation decrease is found even in the decade 2081–2090, but

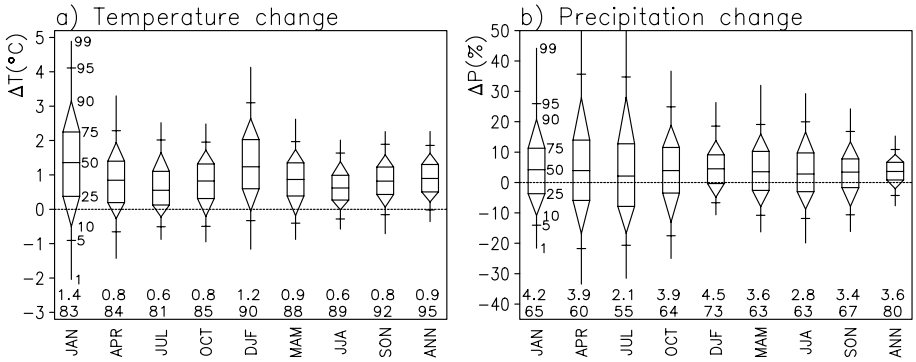

Fig. 4. Probabilistic forecasts of (a) temperature and (b) precipitation change in southern Finland from 1971–2000 to 2011–2020. The box plots show the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentiles of the distribution for four calendar months (January, April, July and October), the four 3 month seasons (DJF, MAM, JJA and SON) and for the annual 10 yr means. The 99th percentiles of precipitation change in April and July exceed 50%. The two rows of numbers in the bottom give the medians of the estimated probability distributions (in °C for temperature change and% for precipitation change) and the probability of increase (in per cent).

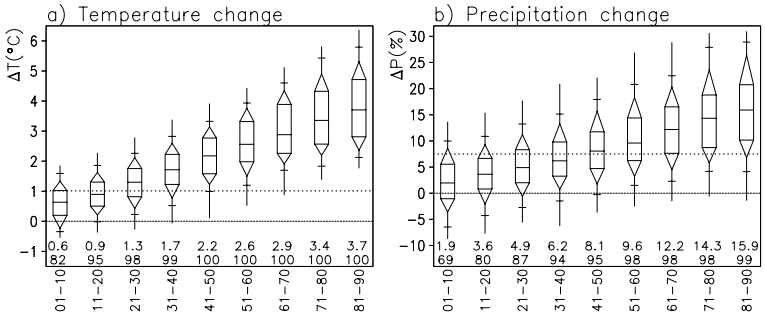

Fig. 5. As Fig 4 but for annual mean (a) temperature and (b) precipitation changes relative to 1971–2000, decade by decade from 2001–2010 to 2081–2090. The dotted lines at 1.0°C and 7.5% indicate the 95th percentiles of 10 yr minus 30 yr temperature and precipitation changes in the unforced control simulations.

the calculated probability of warming reaches unity already in the decade 2051–2060 (and exceeds 99.5% one decade earlier). A larger ensemble of models might of course indicate a non-zero probability of cooling even in the late 21st century.

Figure 5 also includes the 95th percentiles of the probability distribution of 10 yr mean minus 30 yr mean temperature and precipitation changes as estimated from the unforced control simulations. This threshold (approximately 1.0°C for temperature and 7.5% for precipitation change) gives a lower limit of 'detectable' climate change, which can be considered unusual in the context of internal variability. The first decade when the probability of detectable warming (detectable precipitation increase) relative to 1971–2000 exceeds 50% is 2021–2030 (2041–2050).

A global view of the probabilistic results for the annual mean climate change from 1971–2000 to 2011–2020 is given in Fig. 6. As expected from earlier work, the median estimate of the warming (Fig. 6a) is largest over the Arctic Ocean. The estimated probability of warming exceeds 90% everywhere except for the high-latitude Southern Ocean and the northern North Atlantic (Fig. 6c). In wide areas in low latitudes, this probability already approaches unity in the decade 2011–2020, despite the relatively modest median warming there. This reflects the small internal temperature variability in low latitudes, which is illustrated in Fig. 6e with the 95th percentile of the control run 10 yr mean minus 30 yr mean temperature differences.

The probability of exceeding the mentioned 95% threshold of unusual warming in 2011–2020 is actually very high (95–100%) in large parts of the tropics (Fig. 6g). Over much of the tropical Indian Ocean and the western tropical Pacific, this probability already exceeds 95% in the decade 2001–2010 (not shown). The corresponding probabilities in higher latitudes are typically 50–75% in the decade 2011–2020, with even lower values over the Southern Ocean and in areas surrounding the northern North Pacific and North Atlantic. Thus, although the simulated warming is larger over the Arctic Ocean and many other high-latitude areas than in the tropics, this difference is more than compensated by the difference in the magnitude of internal variability. In other words, according to these model experiments high latitudes are not the best place to look for the first signs of anthropogenic climate change.

A similar sequence of figures for precipitation change is given in the right column of Fig. 6. The median changes from 1971–2000 to 2011–2020 (Fig. 6b) are modest in magnitude but display the familiar pattern of generally increasing precipitation in high latitudes and in the tropics and decreasing precipitation in many areas in the subtropics and the lower mid-latitudes

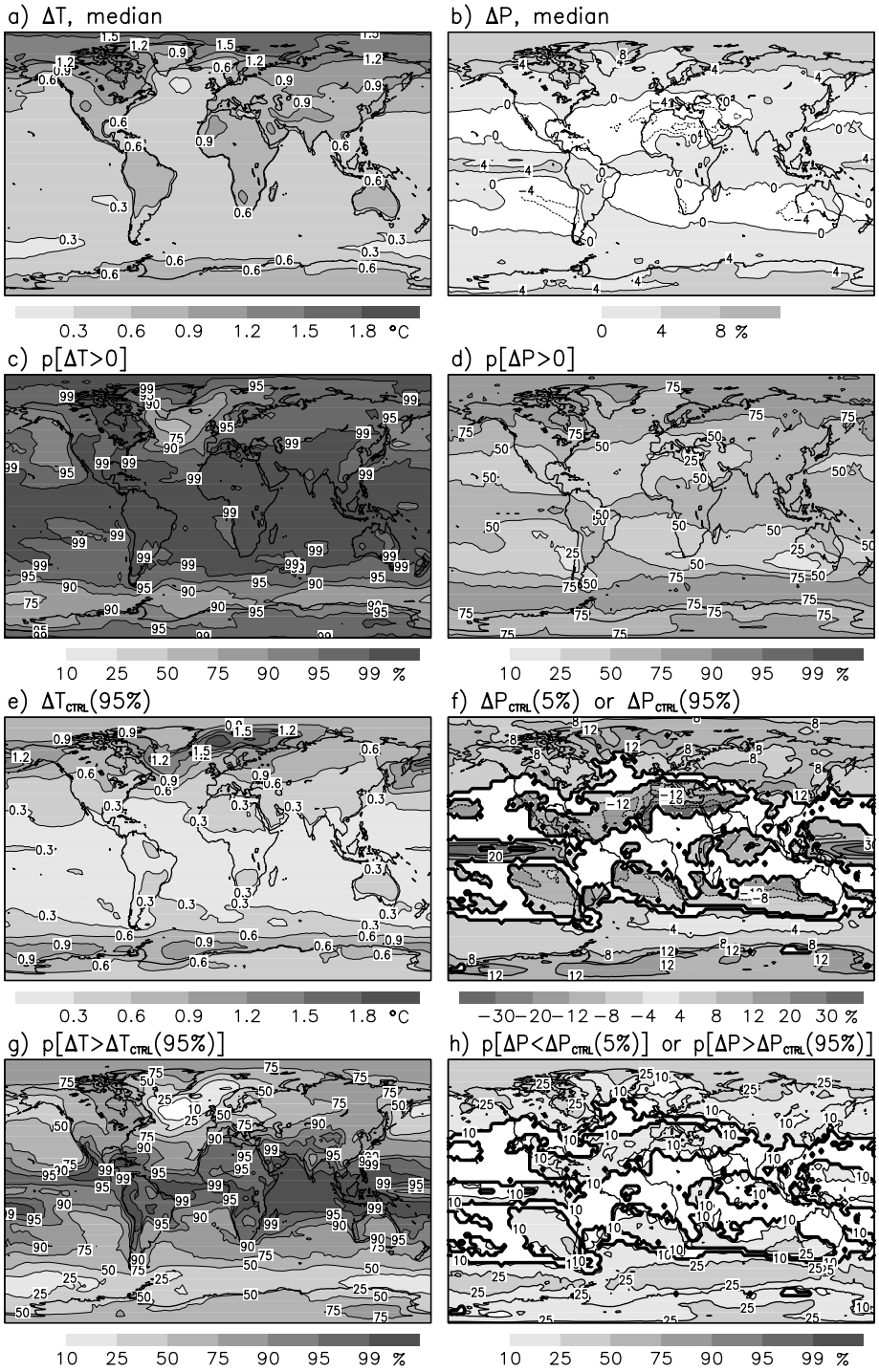

Fig. 6. Probabilistic forecasts of annual mean temperature change (left) and precipitation change (right) from 1971–2000 to 2011–2020. (a)-(b) Median of the forecast distribution; (c)–(d) probability of increase; (e)-(f) threshold of unusual 10 yr minus 30 yr change inferred from the control simulations (see text); (g)-(h) probability of exceeding the threshold of unusual change. In (f) and (h), areas with substantial intermodel disagreement on the sign of the precipitation change are left blank (see text).

(Cubasch et al., 2001). The remaining panels all illustrate the fact that the signal-to-noise ratio is lower for precipitation than temperature changes. The probability of increasing precipitation is mostly in the range 25–75%, although values of 75–90% are found in some high-latitude areas where precipitation has a relatively direct link to temperature via the increased moisture transport capacity of a warmer atmosphere (Manabe and Wetherald, 1987) (Fig. 6d). The threshold of unusual change in

Fig. 6f is defined as the 95th (5th) percentile of the control run distribution of changes where at least 16 of the 21 models agree on precipitation increase (decrease) from 1901–2000 to 2001–2098, and is left undefined elsewhere. This threshold exceeds, in absolute value, the median precipitation change everywhere. The probability that the (positive or negative, depending on the criteria defined above) precipitation change from 1971–2000 to 2011–2020 would exceed this threshold is, therefore, invariably below 50% (Fig. 6h). In most areas this probability is in the range 10–25%.

7. Sensitivity of the forecasts to emissions scenario

The results presented this far have been based on the SRES A1B emissions scenario. To test the sensitivity of our forecasts to the assumed magnitude of greenhouse gas and aerosol emissions, similar calculations were made for the other two SRES scenarios available in the IPCC AR4 database (B1 and A2). As measured by the accumulated anthropogenic CO_2 emissions during the 21st century, A2 is the second highest (after A1FI) and B1 the lowest of the six illustrative SRES scenarios used by Houghton et al. (2001). However, the differences between the scenarios are much smaller in the early than in the late 21st century, particularly so for the CO_2 concentration that responds slowly to changes in emissions (Houghton et al., 2001, p. 14).

Probabilistic forecasts of annual mean temperature and precipitation change in southern Finland (60°N, 25°E), as derived for the B1, A1B and A2 scenarios, are compared in Figs. 7(a) and (b). In this case, we only use those 15 models for which simulations are available for all the three scenarios (Table 1). The results for the A1B scenario differ, therefore, slightly from those given in Fig. 5. The following conclusions arise:

(1) In the late 21st century, exemplified in the figure by the decade 2071–2080, the median changes are largest for the A2 and smallest for the B1 scenario. However, there is a considerable overlap between the distributions derived for the three scenarios. For the B1 scenario with lower emissions, the probability of exceeding the median warming (3.2°C) of the A1B scenario is 25%, whereas the same number for the A2 scenario is 63%. The corresponding numbers for the changes in precipitation are 29% and 62%.

(2) In the early 21st century (2011–2020), differences between the scenarios are negligible compared with the uncertainties associated with internal variability and model differences.

(3) In the mid-21st century (2041–2050) as well, the inter-scenario differences are relatively small. At this time, the median changes for the A1B scenario are slightly higher than those for either of B1 and A2, consistent with the projections of global mean warming by Cubasch et al. (2001). The lower global warming under the A2 than the A1B scenario at this time is due to larger aerosol emissions under the A2 scenario; in terms of the greenhouse gas emissions the two scenarios are similar in the first half of the 21st century.

For comparison, the same analysis is repeated in Figs. 7(c) and (d) for a grid box in equatorial Africa (0°N, 25°E). At this location, internally generated temperature variability is weak and the impact of model differences on simulated temperature changes is also much smaller than in Finland. Consequently, the probability distributions derived for the three scenarios are in

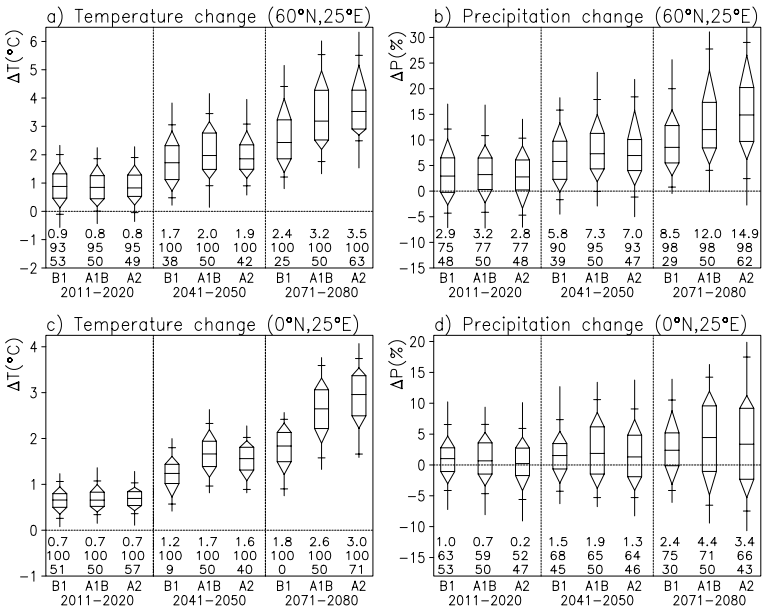

Fig. 7. Probabilistic forecasts of annual mean temperature (left) and precipitation change (right) in southern Finland (top) and Zaire (bottom) under three SRES emissions scenarios. Results are given for the decades 2011–2020, 2041–2050 and 2071–2080, with 1971–2000 as the baseline of the calculation. At the bottom of each panel, the three rows of numbers show the median estimate of the change (in °C for temperature and in% for precipitation), the probability of increase (in%), and the probability of exceeding the median for the A1B scenario (for A1B, this is by definition 50%). See Fig. 4 for the explanation of the box plots.

the decades 2041–2050 and 2071–2080 more distinct from each other in Fig. 7(c) than 7(a). In particular, the distribution for the B1 scenario is in 2071–2080 completely below the median changes for A1B and A2. For precipitation changes, however, the overlap between the scenarios is at least as large as in southern Finland (Fig. 7d). For both temperature and precipitation, the differences between the three scenarios are even at this location negligible in the decade 2011–2020.

Some authors (e.g. New and Hulme, 2000; Giorgi and Mearns, 2003) have derived probability distributions of climate change by combining model results for two or more emissions scenarios. Doing this for the late 21st century requires a non-trivial decision on the relative likelihood of different scenarios. On the other hand, Fig. 7 indicates that such multiscenario forecasts for the early 21st century would be insensitive to how the scenarios are weighted. A multiscenario approach with uniform weighting between scenarios might, therefore, be used to increase the sample size for probabilistic estimates of near-term climate change, without losing the simple interpretation of single-scenario distributions. The only practical complication in the construction of such multiscenario ensembles is the fact that all models have not been run for all scenarios.

8. Summary and discussion

In this paper, we have studied the prospects of near-term decadal-scale climate change in probabilistic terms, using a method that takes into account both the anthropogenic greenhouse gas and aerosol forcing and internal climate variability. Our results are based on a newly completed multimodel ensemble of climate change simulations, together with a resampling technique that helps to increase the number of realizations from which the probability estimates are derived. Below we first list our main results, before discussing some of the issues related to their interpretation.

(1) The resampling technique increases the skill of the resulting probabilistic forecasts in cross-verification, with the largest improvement for near-term forecasts. This indicates that, at least in most areas of the world, the reduced sampling uncertainty in the derived probabilities outweighs the biases eventually caused by the resampling.

(2) The temperature changes that occurred from 1961–1990 to 1991–2000 were in good agreement with our probabilistic hindcasts. The observed changes only fell outside our 5–95% probability range approximately as often as on the average expected from pure chance. By contrast, the observed warming was in wide areas unusually large in comparison with the distribution derived from the unforced control simulations.

(3) The hindcast for precipitation changes from 1961–1990 to 1991–2000 was somewhat worse than the hindcast for temperature, with a slightly too large fraction of the observed changes in the tails of the estimated probability distribution.

(4) In most parts of the world, there is a very high probability of annual mean warming relative to current baselines already in the first few decades of the ongoing century. For southern Finland, for example, our method indicates a 95% probability of a higher mean temperature in 2011–2020 than in 1971–2000 under the SRES A1B emissions scenario.

(5) Precipitation forecasts are in relative terms more uncertain than those for temperature. In southern Finland, the estimated probability of increased precipitation from 1971–2000 to 2011–2020 is 80%. Also, more so than for temperature, it is important to note that the probability estimates in this study pertain to changes in area mean precipitation on the GCM grid box scale (about 10^5 km^2). Because internal climate variability and features of local geography may induce a lot of variation in precipitation changes on scales smaller than that (e.g. Hellström et al., 2001), our results are not directly applicable to precipitation on the truly local scale.

(6) Because reduced temporal averaging leads to larger internal variability, forecasts for seasonal and particularly monthly 10 yr means of climate are more uncertain than those for the annual means.

(7) Internal temperature variability is much smaller in the tropics than in high latitudes. Therefore, although the average model-simulated warming is largest over the Arctic Ocean, the probability of near-term warming and the probability that this warming will be unusual in terms of internal variability both appear to be highest in the tropics.

(8) The best (median) estimates of anthropogenic climate change increase with time. However, the uncertainty ranges also grow wider, as model differences in the response to anthropogenic forcing become increasingly important with increasing magnitude of the forcing.

(9) Probabilistic climate change forecasts for the first decades of this century are almost identical between the SRES A1B, B1 and A2 scenarios. Later on, the scenario-related uncertainty increases but substantial overlap between the probability distributions derived for these three scenarios remains even in the late 21st century.

Our probabilistic forecasts are conditional on the assumptions that (1) the uncertainty in the climate system response to anthropogenic greenhouse gas and aerosol forcing is captured by the multimodel ensemble, and (2) that the magnitude of decadal-scale climate variability is similar between the models and the real world. The forecasts also (3) neglect the uncertainty in the link between greenhouse gas emissions and concentrations. The extent to which (4) uncertainty in aerosol forcing is represented by differences in the treatment of aerosols in different models is also unclear.

Although (1) and (3) are both important when considering climate change on the centennial timescale (Andreae et al., 2005), their impact on shorter-term forecasts is expected to be smaller. The importance of the aerosol issue (4) is more difficult to

estimate; in any case we would expect it to be of concern primarily in those parts of the world where major changes in aerosol emissions are expected in the near future. The magnitude of simulated variability (2) is, however, definitely an important question for the near-term forecasts whose uncertainty arises to a large extent from natural variability.

Whether or not simulated interdecadal variability is realistic is unfortunately not easy to assess, because the observational record is short and affected by anthropogenic forcing. In principle, a negative bias in the simulated variability might be expected because the models exclude in the 21st century simulations the effects of solar variability and volcanic eruptions (even though these factors are present in many of the 20th century simulations that form the first part of the time-series used in our analysis). However, although solar variability and volcanoes have been found to affect the interdecadal variations of the global mean temperature (e.g. Stott et al., 2000) and global land-area mean precipitation (Lambert et al, 2004), their relative contribution to climate variability on small horizontal scales is likely to be more modest, simply because internal variability increases strongly with decreasing scale. Thus, the key issue is probably the magnitude of internal variability.

In an earlier study, Räisänen and Alexandersson (2003) adjusted simulated interdecadal variability based on a comparison of simulated and observed variability on interannual timescales. Here, we have excluded this adjustment partly for simplicity but also because the basic assumption underlying this adjustment (similarity of variability biases on different timescales) remains to be verified. In our future research, we plan to assess the validity of this adjustment technique in a cross-verification framework and study the sensitivity of the probabilistic forecasts to such adjustments and other choices in the methodology.

9. Acknowledgments

We acknowledge the international modelling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. We also thank two anonymous reviewers for their helpful comments. This research has been supported by the Academy of Finland (decision 106979).

10. Appendix. Dependence of cross-verification statistics on ensemble size

Let E be a binary event with probability P. The forecast probability of E can be written as $p = P + p$'. Similarly, the observed occurrence of E in a single forecast case ($o = 0$ or $o = 1$) can be formally decomposed as $o = P + o'$. When the forecast data and the verifying observations come from the same statistical population, which is always the case in cross-verification, the expected values of both o' and p' are zero. The expected value of the Brier score becomes

$$\varepsilon(B) = \varepsilon(p - o)^2 = \varepsilon(p' - o')^2 = \varepsilon(o'^2) + \varepsilon(p'^2) - 2\varepsilon(p'o'). \tag{A1}$$

The two possible values of o', $-P$ and $1 - P$, have probabilities of $1 - P$ and P, respectively. Hence,

$$\varepsilon(o'^2) = (1 - P)P^2 + P(1 - P)^2 = P(1 - P). \tag{A2}$$

On the other hand, if p is estimated from a forecast ensemble of M independent simulations, p follows the binomial distribution with mean P and variance

$$\varepsilon((p - P)^2) = \varepsilon(p'^2) = P(1 - P)/M. \tag{A3}$$

Finally, the deviations p' and o' can be assumed to be uncorrelated, so that the last term in (A1) is zero. Thus

$$\varepsilon(B) = P(1 - P)\left(1 + \frac{1}{M}\right) = B_\infty\left(1 + \frac{1}{M}\right), \tag{A4}$$

where $B_\infty = P(1 - P)$. Because (A4) holds regardless of the threshold used in defining E, a similar equation also holds for the continuous ranked probability score *CRPS*.

References

Allen, M. R. and Ingram, W. J. 2002. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232.

Andreae, M. O., Jones, C. D. and Cox, P. M. 2005. Strong present-day aerosol cooling implies a hot future. *Nature* **435**, 1187–1190.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **78**, 1–3.

Candille, G. and Talagrand, O. 2005. Evaluation of probabilistic prediction systems for a scalar variable. *Q. J. R. Meteorol. Soc.* **131**, 2131–2150.

Christensen, J. H., Carter, T., and Giorgi, F. 2002. PRUDENCE employs new methods to assess European climate change. *EOS* **83**, 147.

Cubasch, U., Meehl, G. A., Boer, G. J., Stouffer, R. J., Dix, M., and co-authors. 2001. Projections of future climate change. In: *Climate change 2001. The Scientific Basis* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson). Cambridge University Press, Cambridge. 525–582.

Giorgi, F. and Mearns, L. O. 2002. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. *J. Climate* **15**, 1141–1158.

Giorgi, F. and Mearns, L. O. 2003. Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. *Geophys. Res. Lett.* **30**, 1629. (doi:10.1029./2003.GL0171.30).

Harvey, L. D. D. 2004. Characterizing the annual-mean climatic effect of anthropogenic CO_2 and aerosol emissions in eight coupled atmosphere-ocean GCMs. *Climate Dyn.* **23**, 569–599.

Hellström, C., Chen, D., Achberger, C. and Räisänen, J. 2001. A comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Res.* **19**, 45–55.

Hersbach, H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. and Forecasting* **15**, 559–570.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., and co-authors. 2001. *Climate Change 2001. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*. Cambridge University Press, Cambridge. 881 p.

Huntingford, C. and Cox, P. M. 2000. An analogue model to derive additional climate change scenarios from existing GCM simulations. *Climate Dyn.* **16**, 575–586.

Karoly, D. J. and Wu, Q. 2005. Detection of regional surface temperature trends. *J. Climate* **18**, 4337–4343.

Lambert, F. H., Stott, P. A., Allen, M. R. and Palmer, M. A. 2004. Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.* **31**, L10203.

Manabe, S. and Wetherald, R. T. 1987. Large-scale changes of soil wetness induced by an increase in atmospheric carbon dioxide. *J. Atmos. Sci.* **44**, 1211–1235.

Mitchell, T. D. 2003. Pattern scaling: an examination of the accuracy of the technique for describing future climate. *Climatic Change* **60**, 217–242.

Mitchell, J. F. B., Johns, T. C., Eagles, M., Ingram, W. J. and Davis, R. A. 1999. Towards the construction of climate change scenarios. *Climatic Change* **41**, 547–581.

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., and New, M. 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). Tyndall Centre Working Paper 55, 30 p.

Nakićenović, N. and Swart, R. (eds.) 2000. *Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge. 599p.

New, M. and Hulme, M. 2000. Representing uncertainty in climate change scenarios: a Monte-Carlo approach. *Integrated Assessment* **1**, 203–213.

Räisänen, J. 2001a. CO_2-induced climate change in CMIP2 experiments. Quantification of agreement and role of internal variability. *J. Climate* **14**, 2088–2104.

Räisänen, J. 2001b. Hiilidioksidin lisääntymisen vaikutus Pohjois-Euroopan ilmastoon globaaleissa ilmastomalleissa. *Terra* **113**, 139–151. (The impact of increasing carbon dioxide on the climate of northern Europe in global climate models; in Finnish with English abstract and figure and table captions).

Räisänen, J. and Alexandersson, H. 2003. A probabilistic view on recent and near future climate change in Sweden. *Tellus* **55A**, 113–125.

Santer, B. D., Wigley, T. M. L., Schlesinger, M. E. and Mitchell, J. F. B. 1990. Developing climate scenarios from equilibrium GCM results. Report 47, Max-Planck Institut für Meteorologie, Hamburg, Germany, 29 p.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Fauli, N., and co-authors. 2005. Uncertainty in the predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403–406.

Stanski, H. R., Wilson, L. J. and Burrows, W. R. 1989. Survey of common verification methods in meteorology. Research report 89–5, Atmospheric Environment Service Forecast Research Division, Canada.

Stott, P. A., Tett, S. F. B., Jones, G. J., Allen, M. R., Mitchell, J. F. B. and co-authors. 2000. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137.

Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego. 467 p.