WP5-recommended common diagnostics for PRUDENCE: time-slice comparisons of temperature, wind speed and precipitation

Christopher A. T. Ferro*, Abdel Hannachi,
and David B. Stephenson
Department of Meteorology, The University of Reading

December 5, 2002

1 Aims

This note presents the recommendations for preliminary, exploratory analyses of PRUDENCE model runs. The aim of the methods described below is to summarise important features of the distributions of meteorological variables so that comparisons may be made between different runs of the same model and between runs of different models. The focus is on summarising the overall shape of the body and the tails of distributions, which admits the investigation of whether or not the extremes are changing in ways that may be explained by overall changes in the distribution, such as a shift in location. The methods are based on calculating a few sample percentiles, which are easy to compute, provide a good description of the whole distribution, and are statistically robust. See Lanzante (1996) for many examples illustrating the benefits of using robust statistics. No attempt is made at this stage to assess the significance of distributional changes, nor to characterise the temporal structure within model runs (Stephenson *et al.*, 2002). Recommendations for the examination of these issues will follow in future communications.

In Section 2 we list the common diagnostics used to summarise distributions of temperature, wind speed, and precipitation variables. Summaries for other variables may be defined similarly. Example analyses in which we apply

*Please send any questions or comments to c.a.t.ferro@reading.ac.uk

our methods to single series of measurements of each variable are presented in Section 3. The data used in these examples are available on the web at www.met.rdg.ac.uk/∼sws02caf/prudence.html for use as a testing ground for computer code. A brief example of the methods applied to an entire field using output from two runs of the regional climate model HadRM3 is given in Section 4. We conclude with a summary of the recommendations in Section 5.

2 Summarising the distributions

Let X represent a variable to be measured at a single location in the control run, and let X' represent the same variable in the perturbed run. Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics from the control run, that is the observations sorted into ascending order, and let $X'_{(1)} < \cdots < X'_{(n)}$ be the order statistics for the perturbed run. For a percentage, p, the p-percentiles are defined to be

$$X_p = X_{([(p/100)n+0.5])} \quad \text{and} \quad X'_p = X'_{([(p/100)n+0.5])}, \tag{1}$$

where $[z]$ denotes the integer part of z. Next we define three summary statistics of the distributions of X and X': the median (denoted m_X and m'_X) which measures central *location*, the inter-quartile range (s_X and s'_X) which measures *scale*, and the quartile skewness statistic (a_X and a'_X) which measures *asymmetry*. These measures are defined in terms of just three, central percentiles:

$$\begin{aligned} m_X &= X_{50}, \\ s_X &= X_{75} - X_{25}, \\ a_X &= (X_{75} - 2X_{50} + X_{25})/s_X, \end{aligned} \tag{2}$$

and similarly for m'_X, s'_X and a'_X. For reference, a Gaussian (normal) random variable with mean μ and standard deviation σ would have $m_X = \mu$, $s_X = 1.35\sigma$, and $a_X = 0$.

A selection of percentiles (Equation 1) and the summary statistics (Equation 2) form the basis of the common diagnostics. Slightly different sets of diagnostics are used for the temperature, wind speed and precipitation fields, which are discussed in the following paragraphs; a list of the common diagnostics appears in Table 1.

Temperature For each run, compute nine percentiles, X_p and X'_p with $p = 1, 5, 10, 25, 50, 75, 90, 95, 99$, and the three summary statistics.

Wind speed As for temperature, but omit the three small percentiles, with $p = 1$, 5, 10, since low wind speeds tend to have no major impact.

Precipitation It is important to distinguish between wet and dry days: a day is defined to be wet if more than 1mm of precipitation is recorded. Summarise the distribution of the precipitation recorded on wet days by the six percentiles and three summary statistics used for wind speeds. In addition, record the proportions, P_w and P'_w, of wet days, and the precipitation means, $\bar{P}$ and $\bar{P}'$, over all days (both wet and dry).

	Temperature	Wind Speed	Wet Days
Small Percentiles	T_1	–	–
	T_5	–	–
	T_{10}	–	–
Central Percentiles	T_{25}	W_{25}	P_{25}
	T_{50}	W_{50}	P_{50}
	T_{75}	W_{75}	P_{75}
Large Percentiles	T_{90}	W_{90}	P_{90}
	T_{95}	W_{95}	P_{95}
	T_{99}	W_{99}	P_{99}
Location	m_T	m_W	m_P
Scale	s_T	s_W	s_P
Asymmetry	a_T	a_W	a_P

Table 1: Common diagnostics recommended for PRUDENCE. Two, additional, precipitation statistics are computed: the proportion of wet days, P_w, and the mean precipitation, $\bar{P}$.

The values computed for the common diagnostics should be plotted for each experiment to give an idea of the distribution of each variable over the field. The differences between the values from the two experiments should also be plotted in order to give an impression of how each distribution has changed. In the next section, we show how to investigate whether or not these changes can be explained in terms of shifts in location, scale, and asymmetry.

3 Explaining changes in distributions

3.1 Temperature

The change in the distribution of temperature may be visualised by plotting

$$\Delta T_p = T'_p - T_p \tag{3}$$

for each value of p. The changes in the percentiles after adjusting for the change in location may also be examined by plotting

$$\left(T'_p - m'_T\right) - \left(T_p - m_T\right) = T'_p - \left\{m'_T + \left(T_p - m_T\right)\right\}. \tag{4}$$

Note that this quantity might also be written as $\Delta T_p - \Delta m_T$, where $\Delta m_T = m'_T - m_T$. Finally, the changes in the percentiles after adjusting for the changes in both location and scale are described by

$$s'_T\left(\frac{T'_p - m'_T}{s'_T} - \frac{T_p - m_T}{s_T}\right) = T'_p - \left\{m'_T + s'_T\left(\frac{T_p - m_T}{s_T}\right)\right\}, \tag{5}$$

which may also be written $\Delta T_p - \Delta m_T - \Delta s_T(T_p - m_T)/s_T$, where $\Delta s_T = s'_T - s_T$.

We illustrate the technique with data taken from two decades of the daily Central England Temperature series. The first decade comprises DJF (1 December to 28 February) for 1960/61 to 1969/70; the second decade comprises DJF for 1970/71 to 1979/80. The data (in degrees Celsius) are stored in the file `t.data`. The values of the common diagnostics are exhibited in Table 2 and Figure 1 shows the quantities (3) – (5).

Each of the percentiles increases from the first time period to the second, indicating a general warming. Moreover, the changes are larger in the lower tail, at colder temperatures, which explains the reduction in scale. Figure 1 shows that the changes in the body of the distribution are well described by the change in location and scale. However, the changes in the tails of the distribution appear to be of a different nature; they might be characterised by a change in the steepness, or kurtosis, of the distribution.

See Section 4 for an illustration using HadRM3 gridded temperature data.

3.2 Wind speed

It is inappropriate to talk about a shift in location of non-negative variables such as wind speed and precipitation since this can result in a range encompassing negative values. Instead, we consider changes in scale and asymmetry by shifting the location and scale after transforming to a logarithmic scale.

Decade	Temperature T 1960s	T' 1970s	Wind Speed W 1980s	W' 1990s	Wet Days P 1960s	P' 1970s
	-4.5	-3.4	–	–	–	–
Small Percentiles	-2.3	-1.0	–	–	–	–
	-1.0	0.3	–	–	–	–
	1.0	2.5	2.6	2.5	2.3	2.1
Central Percentiles	3.6	4.7	3.8	3.5	4.1	3.6
	6.1	6.8	5.4	5.0	6.9	6.2
	7.8	8.4	7.1	6.6	9.9	10.9
Large Percentiles	8.8	9.3	8.6	7.6	13.0	14.2
	10.0	10.7	10.6	9.5	19.0	18.4
Location	3.60	4.70	3.80	3.50	4.10	3.60
Scale	5.10	4.30	2.80	2.50	4.60	4.10
Asymmetry	-0.02	-0.02	0.14	0.20	0.22	0.27

Table 2: Estimates of the common diagnostics in Table 1 for the simulated data. The additional precipitation statistics are $P_w = 0.30$ and $P'_w = 0.34$, $\bar{P} = 1.67$ and $\bar{P}' = 1.78$.

As long as the lowest percentile considered ($p = 25$ for these variables) is non-zero, taking logarithms does not pose a problem.

For wind speed, let $\tilde{W}_p = \log W_p$ and denote the location, scale and asymmetry measures of the transformed variable by $\tilde{m}_W$, $\tilde{s}_W$ and $\tilde{a}_W$. Define similar transformed statistics for W'_p. We plot three quantities for each value of p: the change in percentile,

$$\exp\left(\tilde{W}'_p\right) - \exp\left(\tilde{W}_p\right) = W'_p - W_p; \tag{6}$$

the change in percentile adjusted for the change in scale,

$$\exp\left(\tilde{W}'_p\right) - \exp\left\{\tilde{m}'_W + \left(\tilde{W}_p - \tilde{m}_W\right)\right\} = W'_p - m'_W\left(\frac{W_p}{m_W}\right); \tag{7}$$

and the change in percentile adjusted for the change in both scale and asymmetry,

$$\exp\left(\tilde{W}'_p\right) - \exp\left\{\tilde{m}'_W + \tilde{s}'_W\left(\frac{\tilde{W}_p - \tilde{m}_W}{\tilde{s}_W}\right)\right\} = W'_p - m'_W\left(\frac{W_p}{m_W}\right)^{\tilde{s}'_W/\tilde{s}_W}. \tag{8}$$

We illustrate the technique with two decades of daily, mean, 10m wind-speeds recorded at Bourges in France. The first decade comprises DJF

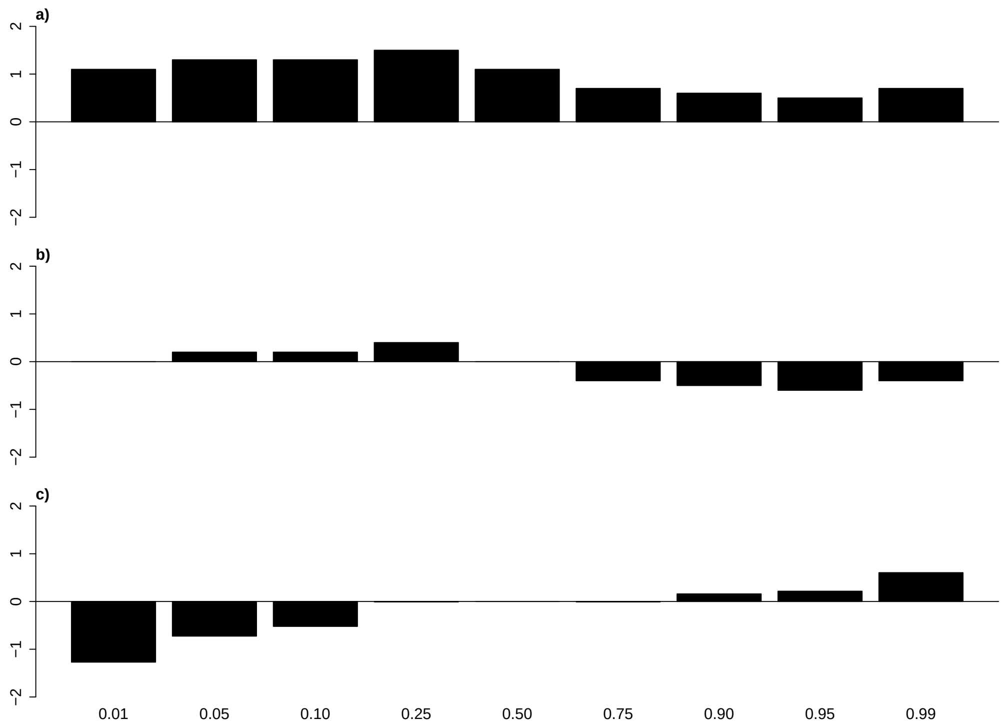

Figure 1: a) Differences between temperature percentiles, $T'_p - T_p$, b) after adjusting for location, $T'_p - \{m'_T + (T_p - m_T)\}$, and c) after adjusting for location and scale, $T'_p - \{m'_T + s'_T(T_p - m_T)/s_T\}$.

for 1980/81 to 1989/90; the second decade comprises DJF for 1990/91 to 1999/00. The data (in metres per second) are stored in the file `w.data`. The values of the common diagnostics are exhibited in Table 2 and Figure 2 shows the quantities (6) – (8).

Wind speeds decrease in general, with greater changes evident at higher percentiles. Figure 2 shows that the change is reasonably well explained just by the reduction in scale.

3.3 Precipitation

For wet days, we must ensure that the range of the variables remains as $(1, \infty)$. This is accomplished by modifying the logarithmic transformation to $\tilde{P}_p = \log(P_p - 1)$. Again, we plot three quantities for each value of p: the

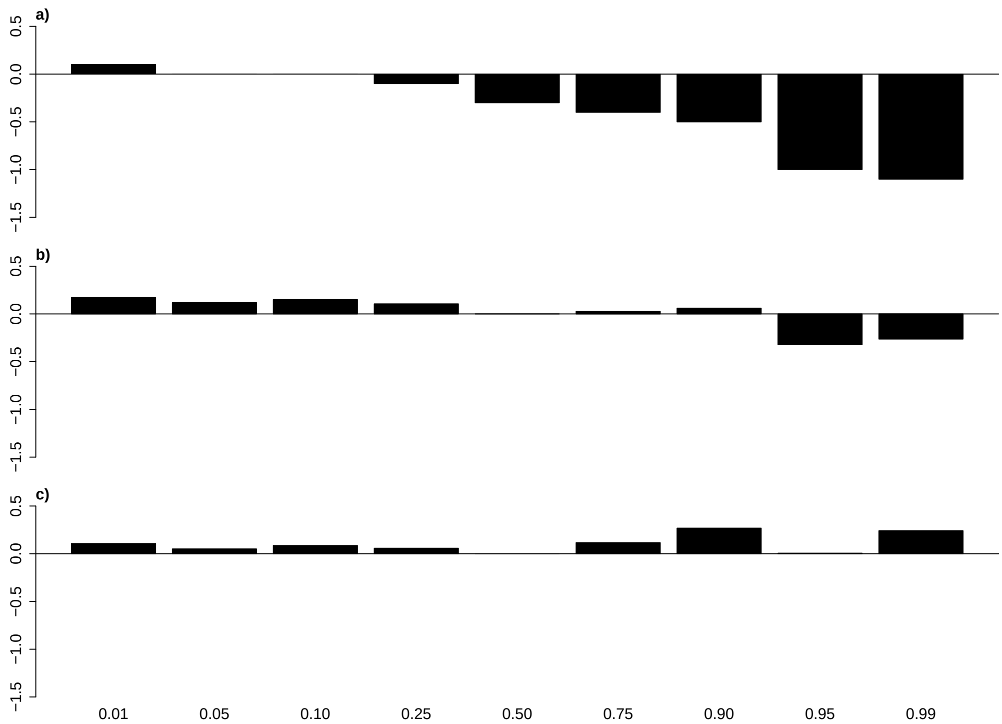

Figure 2: a) Differences between wind speed percentiles, $W_p' - W_p$, b) after adjusting for location, $W_p' - m_W'(W_p/m_W)$, and c) after adjusting for location and scale, $W_p' - m_W'(W_p/m_W)^{\tilde{s}_W'/\tilde{s}_W}$.

change in percentile,

$$\left(1 + e^{\tilde{P}_p'}\right) - \left(1 + e^{\tilde{P}_p}\right) = \left(P_p' - 1\right) - \left(P_p - 1\right); \tag{9}$$

the change in percentile adjusted for the change in scale,

$$\left(1 + e^{\tilde{P}_p'}\right) - \left\{1 + e^{\tilde{m}_P' + (\tilde{P}_p - \tilde{m}_P)}\right\} = \left(P_p' - 1\right) - \left(m_P' - 1\right)\left(\frac{P_p - 1}{m_P - 1}\right); \tag{10}$$

and the change in percentile adjusted for the change in both scale and asymmetry,

$$\left(1 + e^{\tilde{P}_p'}\right) - \left\{1 + e^{\tilde{m}_P' + \tilde{s}_P'(\tilde{P}_p - \tilde{m}_P)/\tilde{s}_P}\right\} = \left(P_p' - 1\right) - \left(m_P' - 1\right)\left(\frac{P_p - 1}{m_P - 1}\right)^{\tilde{s}_P'/\tilde{s}_P}. \tag{11}$$

We illustrate the technique with two decades of the Oxford precipitation series from the European Climate Assessment datasets (Klein Tank *et al.*,

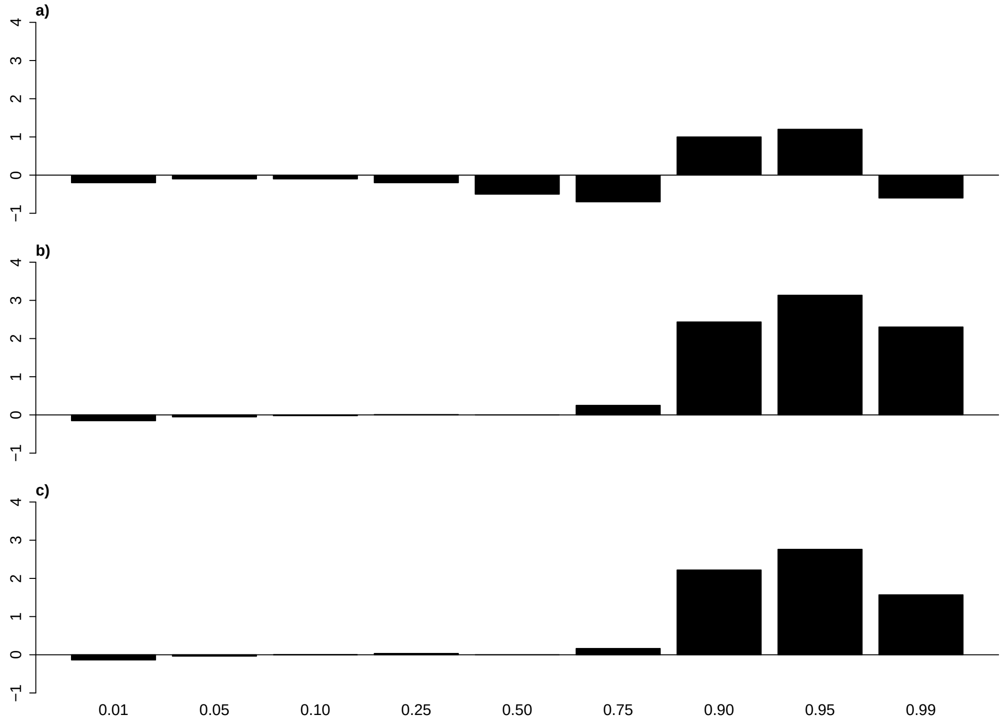

Figure 3: a) Differences between wet-day precipitation percentiles, $(P'_p - 1) - (P_p - 1)$, b) after adjusting for scale, $(P'_p - 1) - (m'_P - 1)(P_p - 1)/(m_P - 1)$, and c) after adjusting for scale and skewness, $(P'_p - 1) - (m'_P - 1)\{(P_p - 1)/(m_P - 1)\}^{\tilde{s}'_P/\tilde{s}_P}$.

2002). The two decades cover the same periods as the temperature data in Section 3.1. The data (in millimetres) are stored in the file `p.data`. The values of the common diagnostics are exhibited in Table 2 and Figure 3 shows the quantities (9) – (11).

The proportion of wet days increases slightly, from 0.30 to 0.34, which contributes to the increase in overall mean rainfall from 1.67mm to 1.78mm. The distribution of wet-day precipitation behaves somewhat differently, however. The body of the distribution moves towards smaller values, a change that is well described by the reduction in scale, but the upper tail exhibits a general increase towards larger values. These contrasting features explain the increase in asymmetry, towards a more positively skewed distribution.

4 Field example

Figures 4 – 6 show plots of some of the quantities mentioned earlier in this note for a control run and a perturbed ($2 \times CO_2$) run of the Hadley Centre regional climate model HadRM3 in winter (DJF). The plots are created by performing the previously documented calculations at each grid point.

Figure 5 shows the spatial distribution of the warming in median temperatures, with a positive trend over land from west to east. There is also a reduction in scale almost everywhere. The changes in asymmetry have lower spatial coherence and are possibly statistically insignificant.

The change in the 10th percentile (cold temperatures) is evident in Figure 6. Again, there is a warming everywhere with an increasing trend west to east. Some change remains in Eastern Europe after adjusting for the change in location; more is accounted for by the additional adjustment for scale; the remainder might be explained by changes in other features, such as shape.

5 Summary of recommendations

The approach is summarised by the following checklist for each variable X:

1. Compute the percentiles, X_p, for each value of p and for each experiment.

2. Compute the summary statistics, m_X, s_X, and a_X for each experiment.

3. Plot the changes in the summary statistics.

4. For each value of p, plot the change in percentile, as well as the change after adjusting for location and scale, or scale and asymmetry.

These are exploratory tools, and just a few of the plots may be sufficient to summarise the changes in distribution and highlight the most interesting features: information that will be invaluable for PRUDENCE.

Finally, note that it is misleading to attempt to quantify the fraction of the change in a particular percentile attributable to a change in, for example, location or scale. The reason for this may be understood by looking at the 99th percentile in Figure 3. After adjusting for scale, the change in this percentile switches sign and increases in magnitude: the change in percentile is not guaranteed to decrease.

References

Klein Tank, A. M. G. and co-authors (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.* **22** (12) 1441–1453.

Lanzante, J. R. (1996) Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.* **16** (11) 1197–1226.

Stephenson, D. B., Hannachi, A. and Ferro, C. A. T. (2002) Point process indices for the assessment of changes in extreme events due to global warming. In preparation.

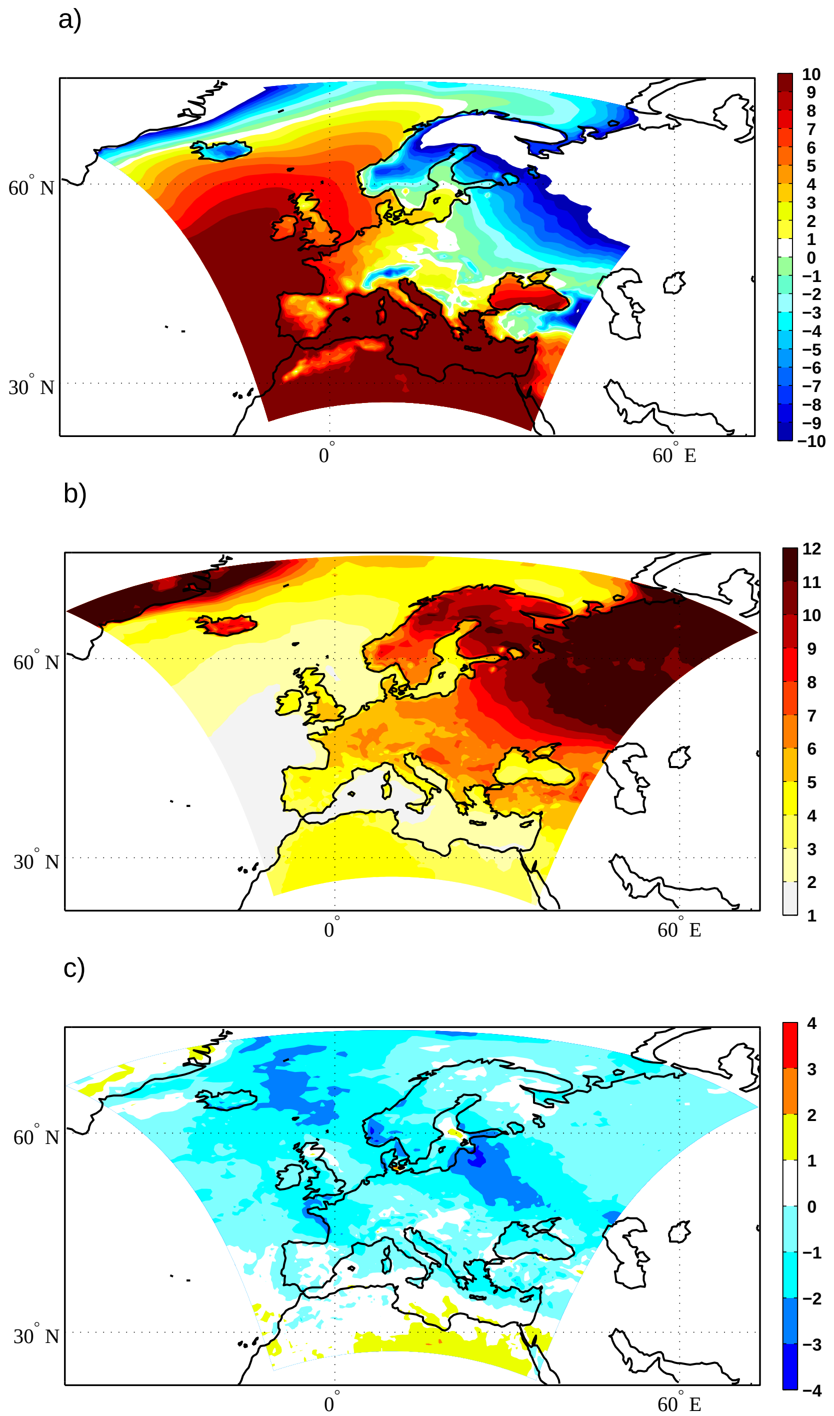

Figure 4: Summary statistics for 2m temperature from the control run: a) m_T, b) s_T, and c) a_T.

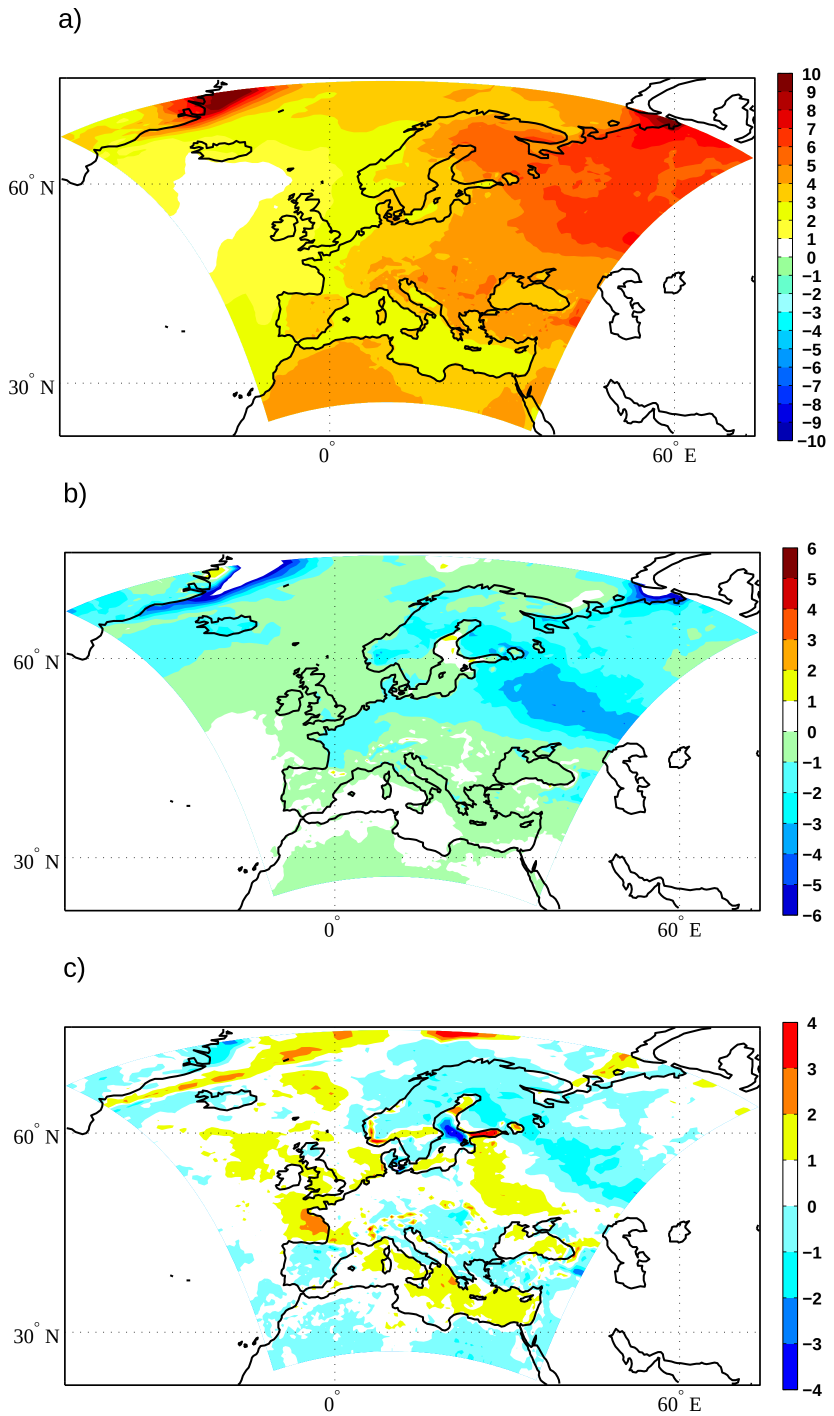

Figure 5: Changes in summary statistics for 2m temperature: a) $\Delta m_T = m'_T - m_T$, b) $\Delta s_T = s'_T - s_T$, and c) $10\Delta a_T = 10(a'_T - a_T)$.

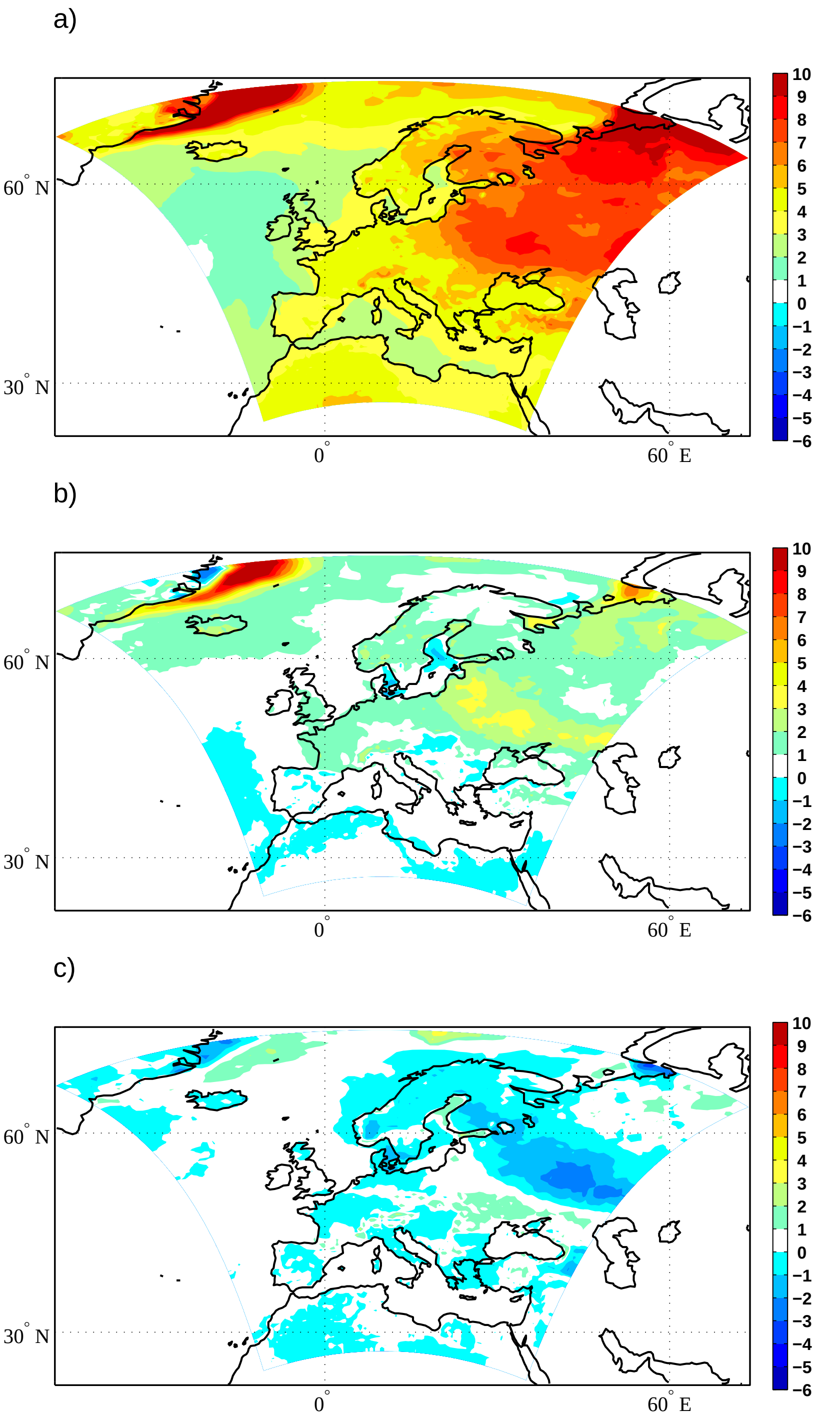

Figure 6: Changes in the 10th percentile of 2m temperature: a) $T'_{10} - T_{10}$, b) $T'_{10} - \{m'_T + (T_{10} - m_T)\}$, and c) $T'_{10} - \{m'_T + s'_T(T_{10} - m_T)/s_T\}$.