

# Non-linear regional relationships between climate extremes and annual mean temperatures in model projections for 1961–2099 over Europe

P. Good<sup>1,5,\*</sup>, L. Bärring<sup>2</sup>, C. Giannakopoulos<sup>1</sup>, T. Holt<sup>3</sup>, J. Palutikof<sup>4</sup>

<sup>1</sup>National Observatory of Athens, Institute of Environmental Research and Sustainable Development, 15236 P. Penteli, Athens, Greece

<sup>2</sup>Department of Physical Geography and Ecosystems Analysis, Geobiosphere Science Centre, Lund University, Sölvegatan 12, 223 62 Lund, Sweden

<sup>3</sup>Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

<sup>4</sup>Hadley Centre, Met Office, Fitzroy Road, Exeter EX1 3PB, UK

<sup>5</sup>Present address: Hadley Centre, Met Office, Fitzroy Road, Exeter EX1 3PB, UK

**ABSTRACT:** A simple method is tested for scaling climate-extreme results from high resolution regional climate models (RCMs) to time-periods and emission scenarios for which the RCMs have not been run. The 30 yr mean relationships between indicators of extremes (IoEs) and annual mean daily maximum temperature ( $T_{xa}$ ) are investigated. Such relationships from the UK Met Office Hadley Centre RCM HadRM3P, along with temperatures from the global climate model HadCM3, are used to scale IoEs to other time periods and scenarios. This is tested for selected indicators of heat-wave and drought over Europe for the period 1961–2099. Curvature is demonstrated in the relationships between these quantities and  $T_{xa}$ . Such non-linearities are shown to have a large potential effect on how these climate extremes are likely to evolve during the century, as well as their sensitivity to emissions. A broad picture of possible changes in European heat-wave and drought severity is presented. For drought over the Mediterranean and western Europe, a very clear positive curvature in the relationship between drought length and annually averaged temperature is found. (This feature is also common in a brief study of 6 other RCMs.) It suggests a rapid increase in drought length towards the end of the century, and a strong sensitivity to the emission scenario. Extended summer dry spells are projected to become a much more regular feature of western European climates. For European heat-waves, we find a slightly earlier onset of increases in heat-wave severity and a reduced sensitivity to emission scenarios than might be expected from a more straightforward interpretation of the Hadley Centre model results. This is linked with extreme dryness occurring at high summer in all years by the end of the century, but was not evident in the 6 other RCMs studied. Based on these results, suggestions are made for choices of future RCM experiments.

**KEY WORDS:** Impacts · Precipitation · Heatwave · Warming · Pattern scaling · Drought

*Resale or republication not permitted without written consent of the publisher*

## 1. INTRODUCTION

The most dramatic impacts of global warming will be through local climate extremes (McCarthy et al. 2001), such as heat-waves or drought. The EU project MICE (Modelling the Impact of Climate Extremes, Hanson et al. in press) studied potential impacts using high-

resolution regional climate model (RCM) simulations. High spatial resolution is likely to be particularly important for the simulation of climate extremes. Further, uncertainties in future emissions and in model physics are such that large ensembles of climate model integrations would ideally be required (Houghton et al. 2001, Gregory et al. 2002, Murphy et al. 2004, Stainforth et al.

\*Email: peter.good@metoffice.gov.uk

2005). However, high-resolution models of the type used by the MICE project are computationally expensive, and are currently run for only limited time periods and scenarios. Large ensemble experiments are not yet feasible.

This type of problem has been addressed for regional projections of annual mean temperature and precipitation through pattern scaling (e.g. Santer et al. 1990, Houghton et al. 2001, Mitchell 2003). Pattern scaling uses a few experiments with complex climate models to estimate (often linear) relationships between a scalar predictor (e.g. global mean temperature) and the local quantity of interest (the predictand, e.g. annual mean temperature). These relationships are assumed to be robust such that given any value of the scalar predictor, a value of the predictand may be estimated for each location. Uncertainty in the scalar predictor can be explored using many experiments with simple models. Then, using the aforementioned relationships, estimates of uncertainty in the predictand (e.g. local annual mean temperature) may be made for each location. However, climate extremes are not expected to be well correlated with global scalars like global mean temperature (Houghton et al. 2001).

Climate extremes can evolve rather differently from mean quantities (e.g. Houghton et al. 2001, Klein Tank & Können 2003, Beniston 2004, Kjellström 2004, Schär et al. 2004, Kjellström et al. in press). However, Beniston & Diaz (2004) found a close linear relationship, for a single regional model grid cell, between summer mean daily maximum temperature and a heat-wave index. This suggests that it may be possible to predict some climate extremes using local time-averaged temperatures. However, such close linear relationships may not be found for other climate extremes and other locations. The present work tests an approach motivated by Beniston & Diaz (2004). Relationships are sought between climate extreme indices and local annual temperature. The aim is to be able to predict 30 yr mean climate extreme index values for time periods and scenarios for which the RCM has not been run. The differences from pattern scaling (using a local annually-averaged quantity rather than a global scalar, to predict climate extremes rather than annual mean quantities) reflect the advances in the resolution of climate models since pattern scaling was proposed.

The main focus is determining whether appropriate relationships exist, and whether they are non-linear, for 2 sample indices of heat-waves and drought. The results are applied in order to make climate extreme projections for time-periods for which the high-resolution RCM had not been run.

## 2. DATA

### 2.1. Climate model data

The climate model data used for this work are from the UK Hadley Centre's HadRM3P regional model and HadCM3 global coupled ocean-atmosphere model. These data are described by Hanson et al. (in press). The Hadley Centre models are described in detail by Gordon et al. (2000), Pope et al. (2000), and Jones et al. (2004a).

A subset of the results from 5 other RCMs run under the EU PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects) project (Christensen et al. in press; see <http://prudence.dmi.dk>) were also studied briefly. These RCMs took very similar boundary conditions as HadRM3P. A separate integration of the regional coupled atmosphere-ocean (RCAO) model, the Swedish Meteorological and Hydrological Institute model, was also available, with boundary conditions from ECHAM4/OPYC3 (Roeckner et al. 1999). The horizontal grid resolution for each model is either 0.44° or 0.5°. Acronyms, institutes and citations for these models are given in Table 1.

The RCM data all correspond to 2 periods: 1961–1990 and 2070–2099. Global model results are available during the period 1961–2099. Future projections used either the Special Report on Emission Scenarios (SRES) A2 or B2 scenarios (Nakicenovic et al. 2000) of future greenhouse gas concentrations. For both Hadley Centre models, results from 3 independent experiments with different initial conditions (testing sensitivity to internal natural variability) were available for 1961–1990 and for future projections under the A2 scenario. For projections under the B2 scenario, results from just 1 experiment were available.

Table 1. Regional climate models (RCMs) from which data was used (main focus is Hadley Centre model HadRM3P)

Model	Institute	Source
HIRHAM	Danish Meteorological Institute (DMI)	Christensen & Christensen (2003)
CHRM	Swiss Federal Institute of Technology (ETH)	Vidale et al. (2003)
CLM	Forschungszentrum Geesthacht GmbH (GKSS)	Stappeler et al. (2003)
HadRM3P	Hadley Centre, United Kingdom Meteorological Office (HC)	Jones et al. (2004a)
REMO	Max Planck Institute for Meteorology (MPI)	Jacob (2001)
RCAO	Swedish Meteorological and Hydrological Institute (SMHI)	Jones et al. (2004b), Räisänen et al. (2004)

## 2.2. Climate indices

Three indices are used: annual mean of daily maximum temperature (the predictor, denoted  $T_{xa}$ ) and 2 indicators of climate extremes (IoEs, the predictands). Each index is calculated once per grid cell per year.

$HT_5$  is a measure of high temperatures sustained over a 5 d period (units:  $^{\circ}\text{C}$ ).  $HT_5$  is calculated from time-series of daily maximum temperature,  $T_{max}$ . For a given year,  $HT_5$  is the highest temperature that was exceeded over 5 consecutive days during that year. If  $HT_5$  takes a value of  $37.3^{\circ}\text{C}$  for a given year, then during the ‘heat-wave’ period for that year  $T_{max}$  was at or above  $37.3^{\circ}\text{C}$  for 5 (or more) consecutive days (but there was no consecutive 5 d period with temperatures at or above  $37.4^{\circ}\text{C}$ ).  $HT_5$  is calculated by dividing the time-series into 5 d overlapping windows. For each 5 d window, the minimum value of  $T_{max}$  is found.  $HT_5$  is the highest such value in the year. Sustained high temperatures can have larger impacts than elevated temperatures on a single day. This index was chosen instead of a 5 d running mean temperature because a 5 d period with continuous high temperatures might have greater impact than a 5 d period with 2 or 3 very hot days separated by a cooler period. However, the specific choice of index probably does not strongly affect the results presented below.

MaxD is defined as the length (in days) of the longest dry spell in a given year (used by Frich et al. 2002). A dry spell is specified as a period of consecutive days with precipitation  $< 0.5 \text{ mm d}^{-1}$ . The choice of the threshold  $0.5 \text{ mm d}^{-1}$  for the HadRM3P model is based on results of Barring et al. (2006).

The high spatial resolution of HadRM3P is important for calculating IoEs, because the associated weather events (heat-waves/drought periods) may be rather small-scale. However, the spatial fields of 30 yr mean IoEs (the quantity of interest) tend to vary more smoothly in space. That is, the high resolution grid is very important for calculating IoEs, but less so for visualising them. For simplicity, all subsequent analysis takes place on the HadCM3 grid. For a given HadCM3 land cell, all HadRM3P land cells within the HadCM3 cell were located. The IoE values for these HadRM3P cells were averaged to give a single value corresponding to the HadCM3 grid cell. Results over sea are not considered. There are approximately 25 HadRM3P grid cells per HadCM3 cell. The following method could equally be applied using IoE on the high resolution grid, but this would be harder to justify in general, given the assumptions of the method (of tight, robust relationships between predictor and predictand) and also the large uncertainties in climate model results.

## 3. RELATIONSHIPS WITHIN THE RCM

The method for estimating relationships within the RCM is most easily discussed in the context of an example. Fig. 1a plots  $HT_5$  against  $T_{xa}$  for a single grid cell covering the northern Balkans. Each small cross corresponds to one year of data. There are 180 crosses, corresponding to the 3 experiments for 1961–1990 and the 3 experiments for 2070–2099 under the A2 scenario (the B2 results are reserved for testing). The 2 clusters

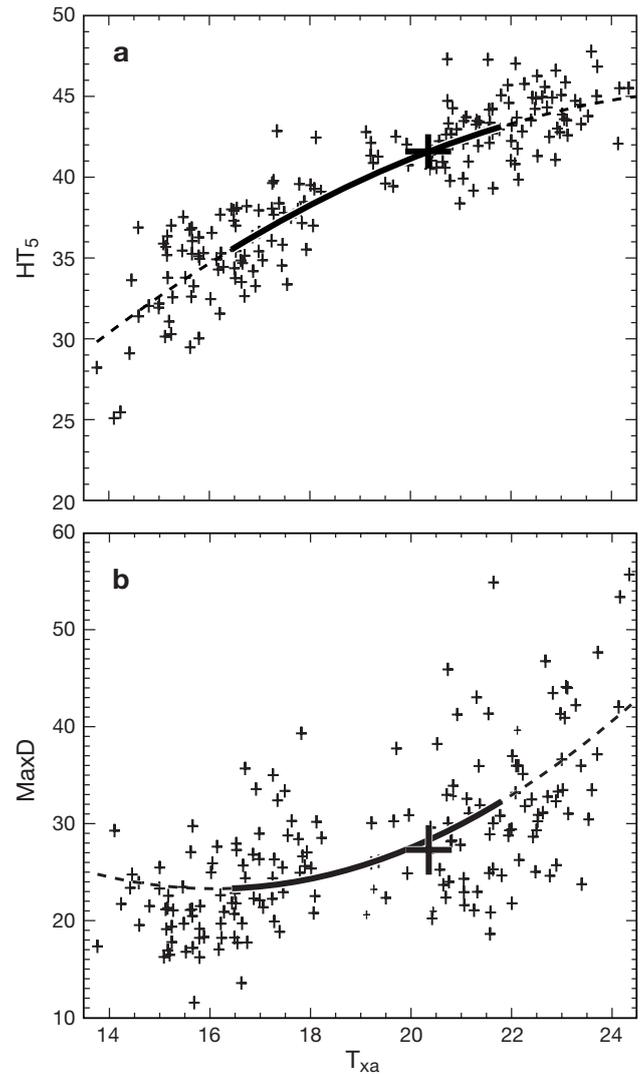


Fig. 1. Relation of IoEs (indicators of climate extremes) to annual mean of daily maximum temperature ( $T_{xa}$ ) for a grid cell over northern Balkans (see white cross in Fig. 4). (a)  $HT_5$  (length [d] of longest dry spell in a given year,  $^{\circ}\text{C}$ ) versus  $T_{xa}$  ( $^{\circ}\text{C}$ ); (b) MaxD (measure of high temperatures sustained over 5 d period, days) versus  $T_{xa}$ . +: 1 yr of data from 1 HadRM3P experiment;  $\oplus$ : 30 yr mean HadRM3P results under B2 scenario. Vertical and horizontal dimensions of large crosses are estimated error bars. Curves are quadratic fits, described in Section 3. Dashed parts of curves indicate that fits are not intended to be used for extrapolation below 1961–1990 mean or above 2070–2099 (A2) mean

of points correspond to the 2 time periods. There is a lot of scatter (associated with natural inter-annual variability), but  $HT_5$  generally increases with  $T_{xa}$ .

Our target was to find a relationship between the 30 yr mean of  $T_{xa}$  and the 30 yr mean of  $HT_5$ . That is, given any value of the 30 yr mean  $T_{xa}$ , we wanted to be able to predict the 30 yr mean of  $HT_5$ . The thick curved line represents such an estimated relationship. It was calculated by fitting a quadratic curve, as follows. First, the relationship between the 30 yr mean of  $T_{xa}$  and the 30 yr mean of  $HT_5$  is known quite precisely for 2 time periods under the A2 scenario (1961–1990 and 2070–2099). Therefore, 2 points for our estimated relationship curve were well defined. The 1961–1990 point was defined by averaging each of  $T_{xa}$  and  $HT_5$  over all 30 yr of data and all 3 experiments. The point for 2070–2099 (A2 scenario) was defined similarly. Together, these 2 points defined a linear fit (not shown). For a quadratic curve fit, constrained to pass through these 2 points, only the curvature of the fit remained to be calculated. This was estimated by least-squares regression, using all 180 yr of data. In effect, the coupled inter-annual variability of  $T_{xa}$  and  $HT_5$  was used to estimate the curvature in the 30 yr mean relationship. We assumed that the coupled inter-annual variability gives information about the coupled variability over longer timescales. This constrained fit was made, instead of fitting intercept, gradient and curvature in a 3-parameter least-squares regression, because our target is a 30 yr mean relationship between  $T_{xa}$  and  $HT_5$ , and this should pass through the 30 yr means of  $T_{xa}$  and  $HT_5$  for 1961–1990 and 2070–2099. The magnitude of the curvature estimated by this method is systematically lower than that from a simple 3-parameter least squares regression. The curve fit is dashed at large and small  $T_{xa}$ , indicating that it should not be used for extrapolation due to large potential errors. Fig. 1b shows equivalent results for the drought indicator MaxD.

A major aim of this work was to discover whether curvature is significant in any relationships between  $T_{xa}$  and IoEs. Hence, quadratic fits were tested. The model data was very clearly insufficient to justify a higher order fit, or to prefer a different second-order fit over the quadratic. Also, the inter-annual variability was such that it was impossible to test the curvature in the relationship against observations (the real climate has not changed enough yet for any curvature to be detectable). Instead, the curvature was tested using results from the B2 scenario. The large cross in each panel (Fig. 1) corresponds to the 30 yr means of the IoE ( $HT_5$  or MaxD) and  $T_{xa}$  for the B2 scenario, 2070–2099. The vertical and horizontal dimensions of the large cross indicate uncertainties in the 30 yr means. If the assumptions and curve fit are valid, the curve fit should

pass through the large cross. This is approximately true for this grid cell for both  $HT_5$  and MaxD. In particular, the curvature of the fits causes them to pass closer to the B2 points than would linear fits. This suggests that at least the sign and approximate magnitude of the curvature in the estimated relationships are reasonable. This grid cell was chosen as an example because the relationships show curvature for both  $HT_5$  and MaxD. This is not generally true; the broader European picture is examined later, with further statistical tests.

The reliability of such curve fits is dependent in part on the size of the horizontal gap between the 1961–1990 and 2070–2099 clusters (the gap is small in Fig. 1, with the clusters almost touching). Specifically, if the difference in  $T_{xa}$  between these 2 time periods is large compared to inter-annual variability, the estimated curvature will tend to be less reliable (Issue 1). Another issue is changes in inter-annual variability. In Fig. 1b, the inter-annual variability is larger for 2070–2099 than for 1961–1990. This will tend to bias the estimated curvature. However, the assumption of a quadratic form, or even the basic assumption of a robust relationship between  $T_{xa}$  and IoEs, may well constitute larger errors. Again, the model data is insufficient to test detailed aspects of the curve fit, so we simply note the potential biases. A related issue (Issue 2) is that the curve fit for MaxD is not a good fit to the set of points for 1961–1990 (left-hand cluster in Fig. 1b). That is, the coupled inter-annual variability of  $T_{xa}$  and MaxD for 1961–1990 does not lie along our estimated curve fit (MaxD increases more rapidly with  $T_{xa}$  over inter-annual timescales for this period). This is partly because the variability is larger for 2070–2099 than 1961–1990, biasing the estimated curvature. This is not necessarily a problem, because the curve fit is intended to capture the relationship between the 30 yr means of  $T_{xa}$  and MaxD. It does suggest that the magnitude of the estimated curvature may be less reliable, and caution is advised. Nonetheless, the agreement with the B2 scenario results (Fig. 1, large crosses) is encouraging (if anything, the B2 results suggest that the curvature magnitude might be underestimated). Issues 1 and 2 are quantified below for the whole of Europe. As an alternative predictor, the Europe-wide mean of  $T_{xa}$  was also tested briefly. However, the scatter in the coupled inter-annual variability of IoE and European mean  $T_{xa}$  is much larger, giving no useful information about curvature in the 30 yr mean relationship.

The negative curvature in the relationship between  $HT_5$  and  $T_{xa}$  may be explained by extreme soil dryness in HadRM3P for 2070–2099. It is known that heat-wave temperatures are amplified by enhanced soil dryness (e.g. Lenderink et al. 2003, van den Hurk

et al. 2005). The ‘heat-wave evaporation rate’ was defined as follows. For each year, the heat-wave period is defined as the 5 days in the year for which  $T_{\max}$  is continuously at or above  $HT_5$  (see definition of  $HT_5$ , Section 2.2). For some years,  $T_{\max}$  may be above  $HT_5$  for >5 consecutive days—in this case, the first 5 d of the heat-wave period are used. The daily evaporation rate is then averaged over this 5 d period (denoted the ‘heat-wave evaporation rate’). Fig. 2 plots heat-wave evaporation rate against  $T_{xa}$  (Fig. 2a), and against  $HT_5$  (Fig. 2b). Heat-wave evaporation rates are approximately linear with  $HT_5$ , confirming that heat-wave temperatures are strongly linked to evaporation rates. For 1961–1990, warmer years (larger  $T_{xa}$ ) have dryer soil by high summer, and hence more intense heat-waves (larger  $HT_5$ ). However, by the end of the century, when  $T_{xa}$  is large, heat-wave evaporation rates are close to 0 for almost every year. This means that the amplification of heat-wave temperatures by soil dryness cannot increase, even in the hottest years.

For summer drought, the situation is more complex. However, a more vigorous hydrological cycle, reductions in water supply (through reduced evaporation and westerly transport) and increased cyclonicity have all been identified as important mechanisms (e.g. Houghton et al. 2001), and it is not surprising that the balance between these and other mechanisms could lead to non-linear relationships with  $T_{xa}$ . The HadRM3P shows similar patterns of surface pressure anomalies (indicating reduced westerly transport for the Mediterranean and western Europe, and increased cyclonicity for northwest Europe) to those found by Räisänen et al. (2004) (for the RCAO model).

#### 4. PREDICTING CLIMATE EXTREMES USING $T_{XA}$ FROM HADCM3

If the RCM results show significant relationships between  $T_{xa}$  and IoEs, then it should be possible to predict IoEs for other scenarios and time periods using these relationships along with  $T_{xa}$  from the global model (HadCM3). We first apply a bias correction to  $T_{xa}$  from HadCM3, described in ‘Appendix A1’. This is because the sea-surface temperatures used to force HadRM3P experiments are bias-corrected using present-day observations (whereas HadCM3 is a coupled ocean-atmosphere model). As an alternative predictor,

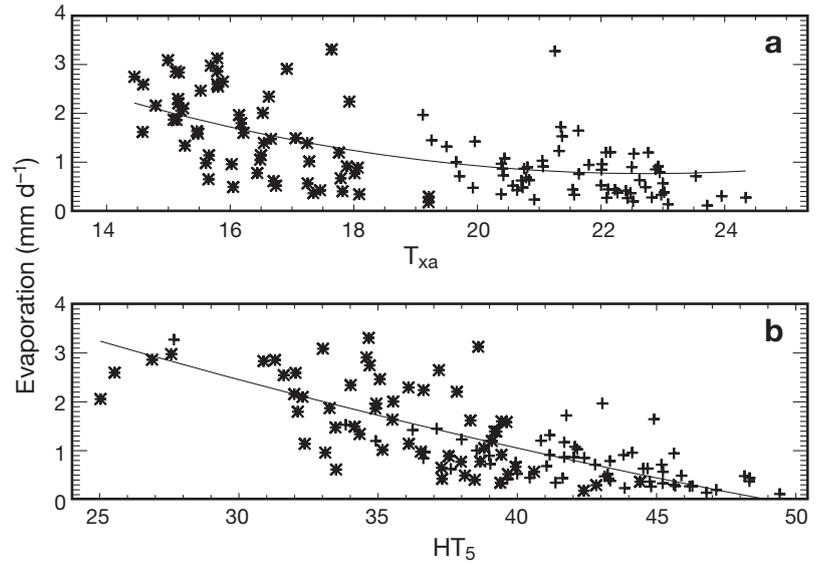


Fig. 2. Link between heat-wave evaporation rates ( $\text{mm d}^{-1}$ ) and heat-wave temperatures ( $^{\circ}\text{C}$ ) for northern Balkans. Each point represents 1 yr of data from HadRM3P. (a) Heat-wave evaporation rate against  $T_{xa}$ ; (b) heat-wave evaporation rate against  $HT_5$ . \*: 1961–1990. +: 2070–2099 (A2 scenario). Solid lines: least-squares fits of quadratic curves. See Fig. 1 for abbreviations

the June to August mean of  $T_{\max}$  was briefly tested, but the differences between HadRM3P and HadCM3 for this quantity were found to be excessive (‘Appendix A1’). Fig. 3a shows 2 projected time-series of the 30 yr mean of  $HT_5$  for the northern Balkans grid cell. Both use bias-corrected values of  $T_{xa}$  from HadCM3. These are converted to estimates of the 30 yr mean of  $HT_5$ , using relationships between  $T_{xa}$  and  $HT_5$  taken from the RCM results for 2070–2099 (A2 scenario). The solid line uses the quadratic relationship between  $T_{xa}$  and  $HT_5$  shown in Fig. 1a (thick solid curve). The dashed line corresponds to a linear relationship between  $T_{xa}$  and  $HT_5$  (as in Section 3, but with curvature set to 0). The difference between the curves illustrates the importance (or otherwise) of the curvature in the relationship between  $T_{xa}$  and  $HT_5$  (Fig. 1a). By definition, the curves cross at 1975 and 2085 (corresponding to the 1961–1990 and 2070–2099 means). However, the curves differ in the rate of change of  $HT_5$  early and late in the century.

For impacts and adaptation, the rate of change of IoEs over 30 yr timescales could be much more important than the change over a century.  $HT_5$  increases more rapidly in the early part of the century if the quadratic fit is used (mid-century, the curve from the quadratic fit leads that from the linear fit by about 9 yr). For MaxD (Fig. 3b), the curvature in the relationship with  $T_{xa}$  has a rather more dramatic result. When the quadratic fit is used, by mid-century the MaxD time-series lags that from the linear fit by about 30 yr, but towards the end of the century MaxD increases very

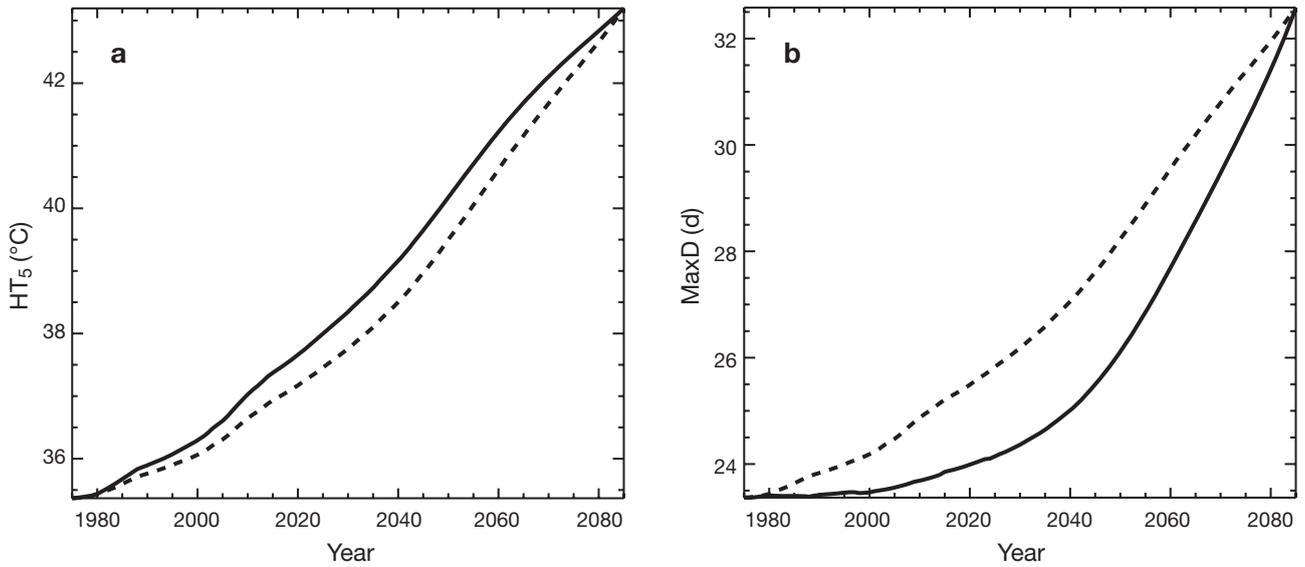


Fig. 3. Effect of estimated fit curvature on predicted evolution of each IoE over northern Balkans, from 1975 (centre of 1961–1990) to 2085 (centre of 2070–2099). (a)  $HT_5$ ; (b) MaxD. Time-series estimated using curve fits and  $T_{xa}$  from HadCM3. Dashed lines: linear fits; solid lines: quadratic fits. See Fig. 1 for abbreviations

rapidly indeed. Such behaviour could strongly affect impacts and adaptation in socio-economic and natural environment contexts. A slow initial evolution, while allowing more time for adaptation, may lead to a degree of complacency in the socio-economic sector. The subsequent very rapid increase in drought length could cause much greater impacts than would a steady increase throughout the century. This would be especially true for the natural environment, where adaptation is not pre-planned. The likely sensitivity to different forcing scenarios is also strongly affected. The various emission scenarios only differ strongly towards the end of the century, at which point  $T_{xa}$  is high. At high  $T_{xa}$ , MaxD is very much more sensitive to changes in  $T_{xa}$  than is  $HT_5$  (assuming the curvature of each fit in Fig. 1 is roughly correct). That is, for this grid cell, drought length is predicted to be much more sensitive than heat-wave intensity to future emissions. Next, the European-wide picture is examined.

### 5. METHOD PERFORMANCE ACROSS EUROPE

The method is expected to perform best when the mean change in  $T_{xa}$  over the century (under the A2 scenario) is not large compared to the inter-annual

variability (Issue 1, Section 3). This was seen visually for the northern Balkans grid cell in Fig. 1 as the horizontal gap between the 2 clusters of points (small in that case). We summarise this property for the whole of Europe in Fig. 4 by a quantity called ‘coverage’ (defined in ‘Appendix A2’). For reference, the northern Balkan cell (Fig. 4: white cross) has a coverage of about 0.8. The method might be expected to be less reliable for points away from the red zone in Fig. 4 (i.e. for North Africa and eastern Europe).

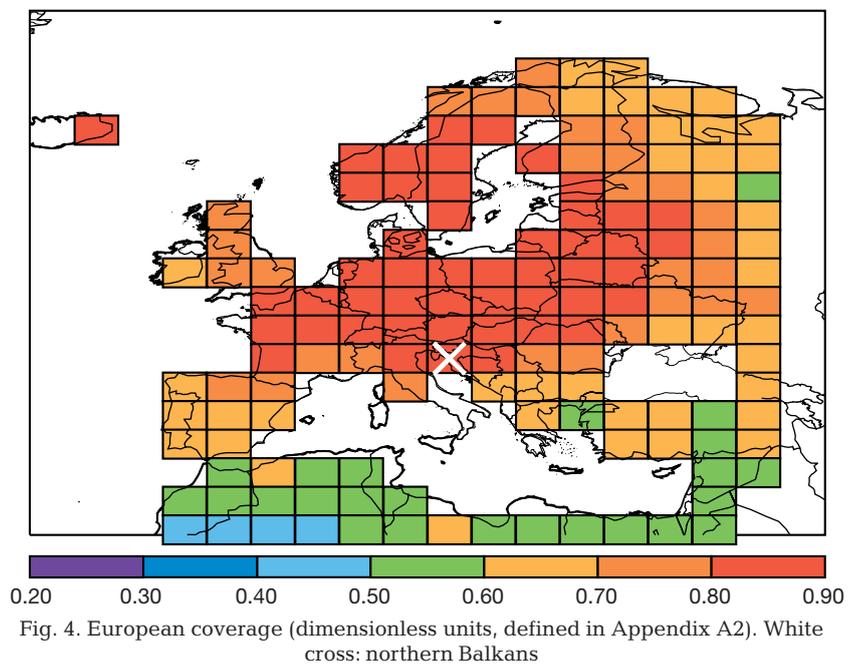


Fig. 4. European coverage (dimensionless units, defined in Appendix A2). White cross: northern Balkans

A second indicator (Issue 2, Section 3) of potential problems with a fit is where the coupled inter-annual variability of  $T_{xa}$  and IoE for either 1961–1990 or 2070–2099 does not lie along the curve fit. ‘Appendix A3’ describes a significance test for this issue based on the Mann-Kendall rank correlation test. We apply this test to both quadratic and linear fits, for each time period separately. Results are shown in Fig. 5. Blue or orange colours indicate that Issue 2 is significant, pointing to potential problems with a fit. For the linear fits, this is true for most of Europe (74% of grid cells). This indicates that a linear fit in general is probably not appropriate. In this respect, HadRM3P differs from other PRUDENCE models, for which linear fits are largely sufficient (not shown; see e.g. Beniston & Diaz 2004). For the quadratic fits, a much smaller problem area is highlighted (37% over Europe; 55% over the whole domain). This includes grid cells that are both near to the coast and close to the same latitude as the south of England. For this region, curve fits may be less

reliable. This suggests caution, but does not mean the quadratic fits will necessarily be less accurate than a linear fit in this region. For both time periods, the test statistic is positive. This indicates that the relationship between  $HT_5$  and  $T_{xa}$  arising from inter-annual variability is slightly steeper than the curve fit in both time periods. Therefore, it is not possible to satisfactorily fit a quadratic to both data clusters. This could be because the  $HT_5$ - $T_{xa}$  relationship is slightly different for inter-annual as opposed to climatological timescales. Alternatively, a higher-order curve fit might be required (not possible to investigate using the available data). For southern Europe, this test indicates no significant problem.

If the assumptions and execution of the method are valid, then it should be possible, for each grid cell to use  $T_{xa}$  from HadCM3 along with the  $T_{xa}$ - $HT_5$  relationship from HadRM3P to predict the 30 yr mean of  $HT_5$  for 2070–2099 under the B2 scenario. Since HadRM3P results for 2070–2099 under the B2 scenario

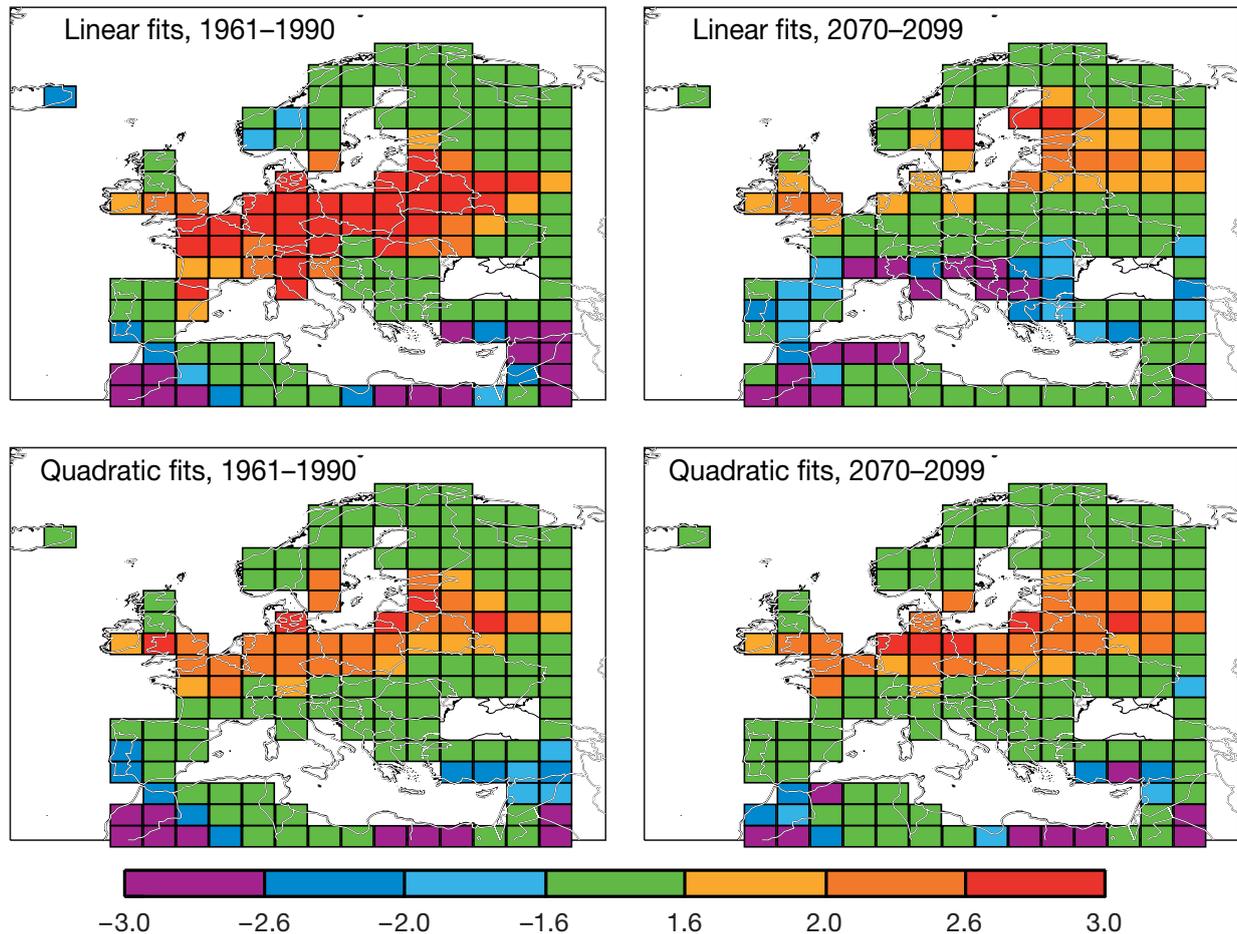


Fig. 5. Mann-Kendall test results for  $HT_5$  (see Appendix A3). Linear and quadratic fits for 1961–1999, and 2070–2099 (A2 scenario) Colour indicates the Mann-Kendall test statistic. Blue or orange: caution advised in use of curve fits. Test statistic magnitudes of 1.6, 2 and 2.6 indicate 10, 5 and 1% significance, respectively. Positive values for a particular time period indicate that over inter-annual timescales,  $HT_5$  increases faster with  $T_{xa}$  than does the curve fit. See Fig. 1 for abbreviations

are available, the method may be tested. The B2 scenario results were not used to construct the relationships between  $T_{xa}$  and  $HT_5$ , so this constitutes an independent test. The 'prediction error' is defined as the difference between the 30 yr mean of  $HT_5$  predicted using our method and the equivalent value from HadRM3P. A test of statistical significance of prediction errors is described in 'Appendix A4'. Fig. 6a shows the significance of prediction errors when linear relationships between  $HT_5$  and  $T_{xa}$  are used. Fig. 6b is as Fig. 6a, but for quadratic relationships between  $HT_5$  and  $T_{xa}$ . For cells coloured blue, the 'prediction error' is not statistically significant even at the 10% level. This means that for these cells, our method has been used to make an acceptable prediction of what the RCM would calculate for  $HT_5$  under a completely different emission scenario. This is true for most of central and southern Europe for quadratic fits, and over a smaller region for linear fits. Some regions show strongly significant errors. However, these regions mostly correspond to those highlighted as having potentially unreliable curve fits, due either to low coverage (Issue 1), or prob-

lems indicated by the Mann-Kendall tests (Issue 2). It is plausible, therefore, that the errors arose simply because the 30 yr mean relationships between  $T_{xa}$  and  $HT_5$  were not estimated accurately.

However, with the available data, we cannot rule out the possibility that it is just not sensible to define a single 30 yr mean relationship between  $HT_5$  and  $T_{xa}$  for these regions and these anthropogenic emissions scenarios. Fig. 6c shows the difference between the magnitudes of 'prediction errors' for linear and quadratic fits. With the exception of regions with low coverage, the 'prediction errors' using linear fits are either similar to or larger than those using quadratic fits. For most of central and southern Europe, errors from linear fits are >50% larger than those for quadratic fits. The negative curvature (seen over all regions with non-significant 'prediction errors') is very probably associated with extreme soil dryness by the end of the century (discussed in Section 3). Overall, quadratic fits appear to perform better than linear fits.

For the drought measure MaxD, the Mann-Kendall test of Issue 2 gives significant results across sub-

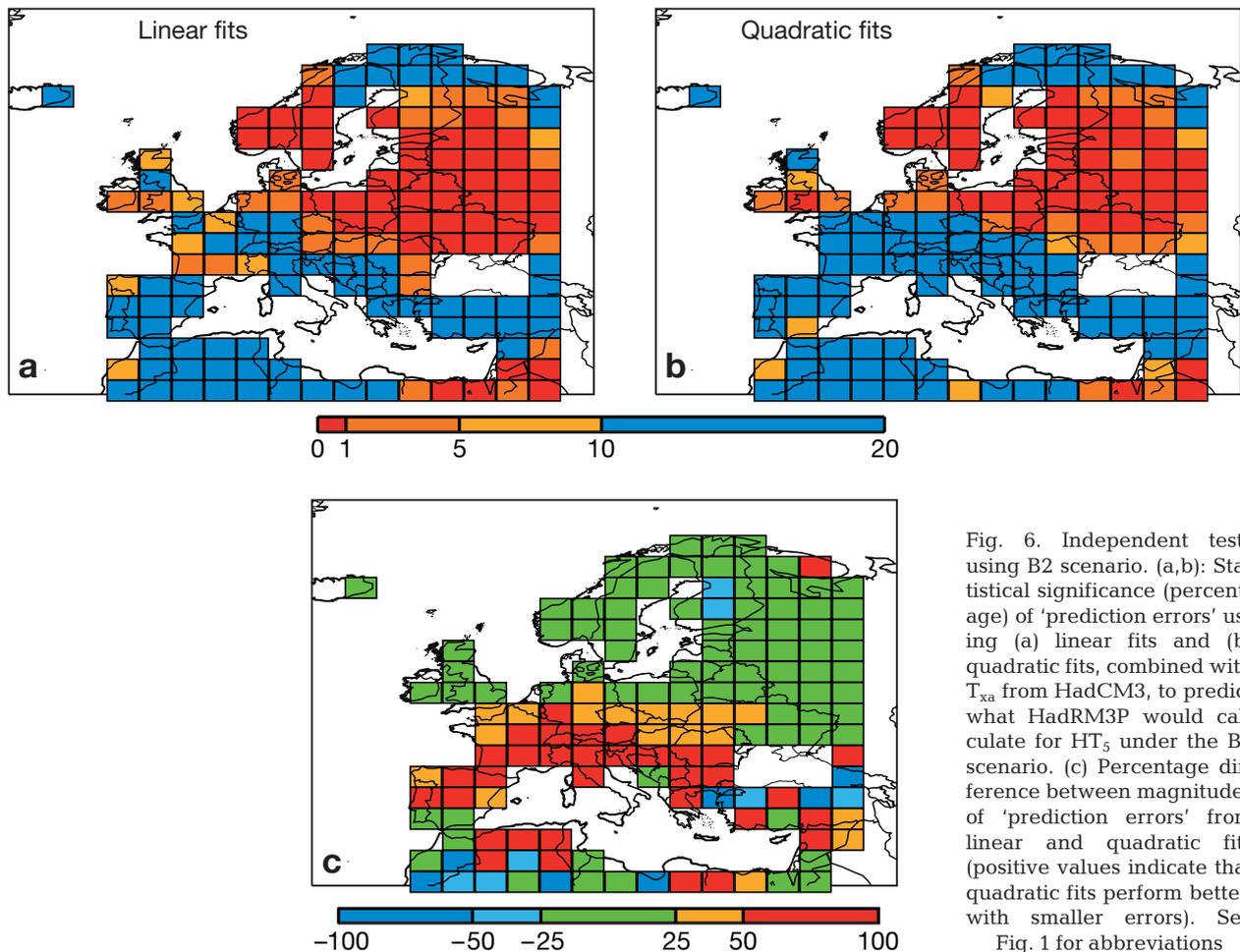


Fig. 6. Independent tests using B2 scenario. (a,b): Statistical significance (percentage) of 'prediction errors' using (a) linear fits and (b) quadratic fits, combined with  $T_{xa}$  from HadCM3, to predict what HadRM3P would calculate for  $HT_5$  under the B2 scenario. (c) Percentage difference between magnitudes of 'prediction errors' from linear and quadratic fits (positive values indicate that quadratic fits perform better, with smaller errors). See Fig. 1 for abbreviations

stantial parts of central and southern Europe, for both linear and quadratic curve fits (Fig. 7). This does not mean that the curve fits are unrepresentative of the 30 yr mean relationship between MaxD and  $T_{xa}$ , but caution is advised. It is to be expected that a measure of drought would show fewer clear relationships with  $T_{xa}$  than would a measure of heatwave severity. In contrast, the ‘prediction errors’ for the B2 scenario using quadratic fits are not significant for virtually the whole domain (Fig. 8). Prediction errors are generally less significant for MaxD than for  $HT_5$  because inter-annual variability in MaxD is larger relative to the climate change signal. However, the linear fits do show significant errors for MaxD (Fig. 8b) across the length of the European Mediterranean coast. ‘Prediction errors’ using linear fits are either similar to or larger than those using quadratic fits across almost all grid cells for which coverage is greater than 0.7 (Fig. 4). This is true except for a region covering Germany and Switzerland, and parts of more northerly Europe (for which mean MaxD changes over the century are anyway rather small).

For the cells covering France and much of Mediterranean Europe, the quadratic fits perform substantially better than the linear fits. This region is the focus of the remaining investigation. The robustness of this result for Mediterranean Europe is tested using results from the other PRUDENCE RCMs. Fig. 9 plots the estimated curvature in the relationship between MaxD and  $T_{xa}$  for each of the 6 HadCM3 grid cells extending from southwest France to west Bulgaria, for 7 different RCMs. For the RCAO model, results from integrations forced by 2 different driver global models are presented (HadAM3H and ECHAM4). The dashed line marks 0 curvature (i.e. a linear relationship between MaxD and  $T_{xa}$ ). This figure shows that positive curvature in the relationship between MaxD and  $T_{xa}$  seems to be a relatively common result in the current generation of RCMs. It is also found in the RCAO model forced by ECHAM4, indicating that it is not just an artefact of the HadAM3 boundary conditions. Further, the strong curvature found above for the HadRM3P model is not at all an extreme case. Rather, it is roughly equal to the average of curvatures found for the other models for this region.

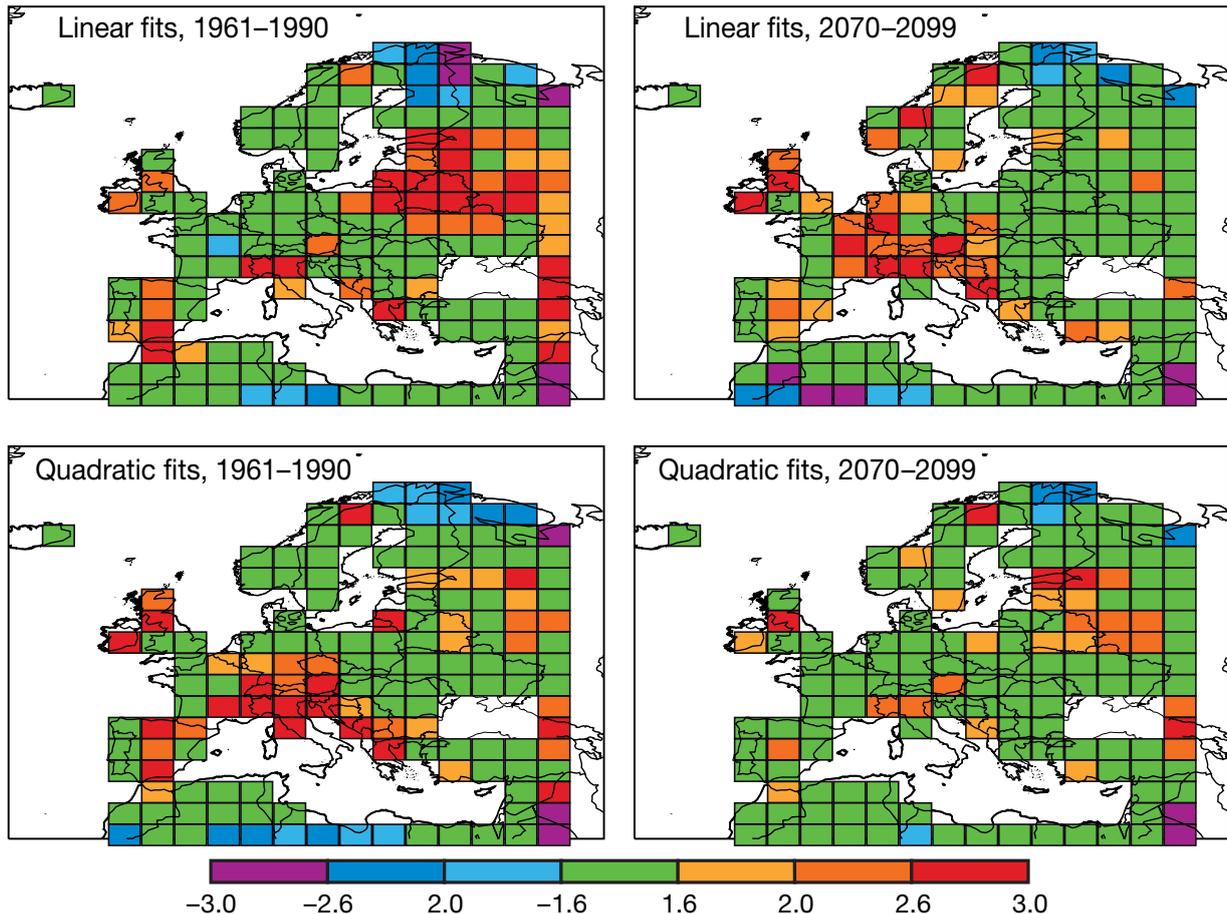


Fig. 7. As for Fig. 5, but for MaxD. See Fig. 1 for abbreviations

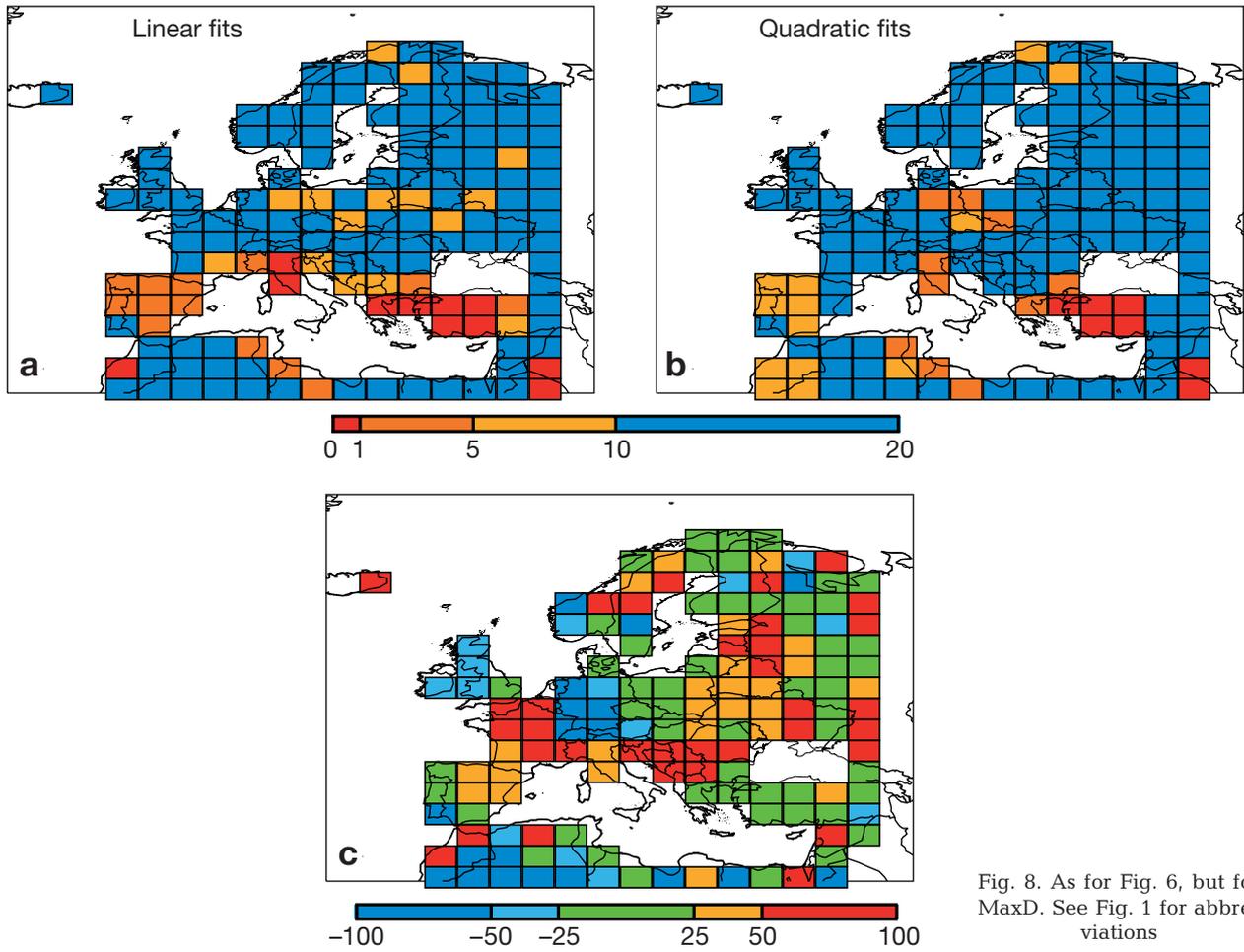


Fig. 8. As for Fig. 6, but for MaxD. See Fig. 1 for abbreviations

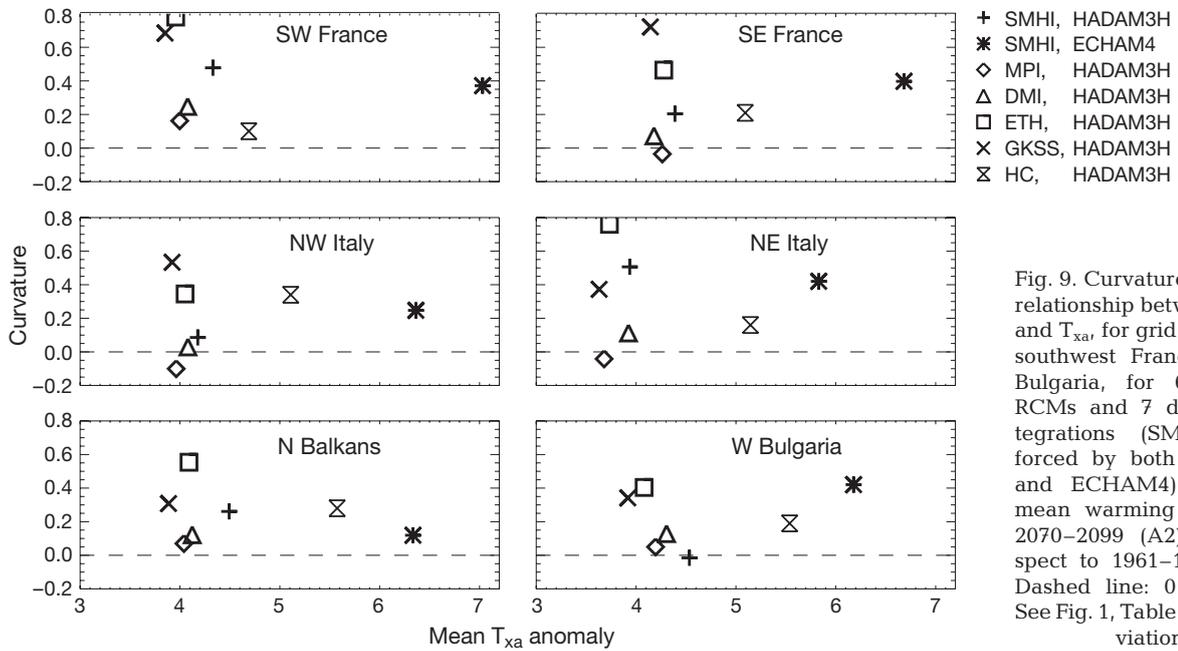


Fig. 9. Curvature (d °C<sup>-2</sup>) in relationship between MaxD and T<sub>xa</sub> for grid boxes from southwest France to west Bulgaria, for 6 different RCMs and 7 different integrations (SMHI model forced by both HadAM3P and ECHAM4). Abscissa: mean warming in T<sub>xa</sub> for 2070–2099 (A2), with respect to 1961–1990 mean. Dashed line: 0 curvature. See Fig. 1, Table 1 for abbreviations

### 6. APPLICATION: IoE PROJECTIONS THROUGHOUT THE CENTURY

Now we apply the method to make projections of  $HT_5$  and MaxD throughout the century. The quadratic curve fits are used since they generally performed better in testing. The results for  $HT_5$  are given in Fig. 10. The A2 and B2 scenarios give very similar values before 2070, so those B2 results are not shown. The results for 2070–2099 for the A2 scenario are by definition equal to the actual HadRM3P calculation. The HadRM3P results for the B2 scenario (2070–2099) are also shown. Under the A2 scenario, some dramatic increases,  $>7^{\circ}\text{C}$  by the end of the century, are predicted for large parts of Europe. This increase in heat-wave severity is relatively linear with time for most of Europe, in contrast with the slightly delayed increase in annual mean temperatures (as Fig. 3a). These results are just an example application, not to be interpreted too quantitatively, showing a scenario of how European heat-wave temperatures could evolve. Fig. 11 shows the projected evolution of MaxD. As

with  $HT_5$ , some dramatic increases are found, of  $>50\%$  in drought length under the A2 scenario. Strikingly, for cells over France and much of Mediterranean Europe, much of this increase occurs during the final 30 yr under the A2 scenario (the change from 2040–2069 to 2070–2099 is much larger than from 2010–2039 to 2040–2069). Such rapid increases in drought towards the end of the century could have large socio-economic impacts. Rapid changes over 30 yr may be much harder to adapt to than steady change over the century, especially for the natural environment. Similarly, the difference between the A2 and B2 scenarios is large, showing that changes in drought length could be strongly dependent on emission scenario. Again, these results are purely illustrative and large errors should be expected. However, they do show that curvature in the relationship between annual mean temperatures and drought length could have a large impact both on drought evolution during the century and sensitivity of drought length to emission scenarios, and should be investigated further.

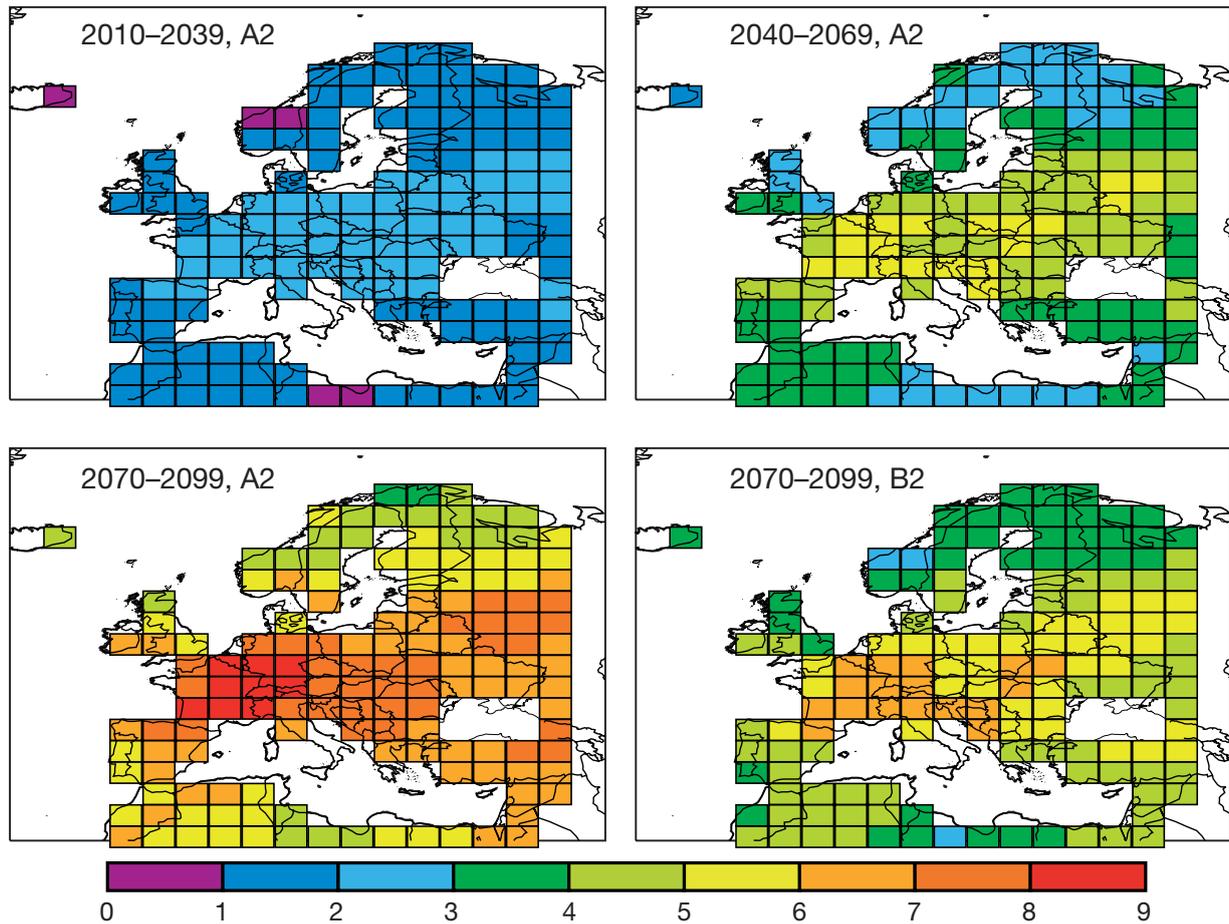


Fig. 10. Predicted evolution of  $HT_5$  ( $^{\circ}\text{C}$ ), using curve fits and  $T_{xa}$  from HadCM3 for A2 and B2 scenarios. Results illustrate a scenario of how  $HT_5$  would evolve if estimated curve fits and HadRM3P were accurate. See Fig. 1 for abbreviations

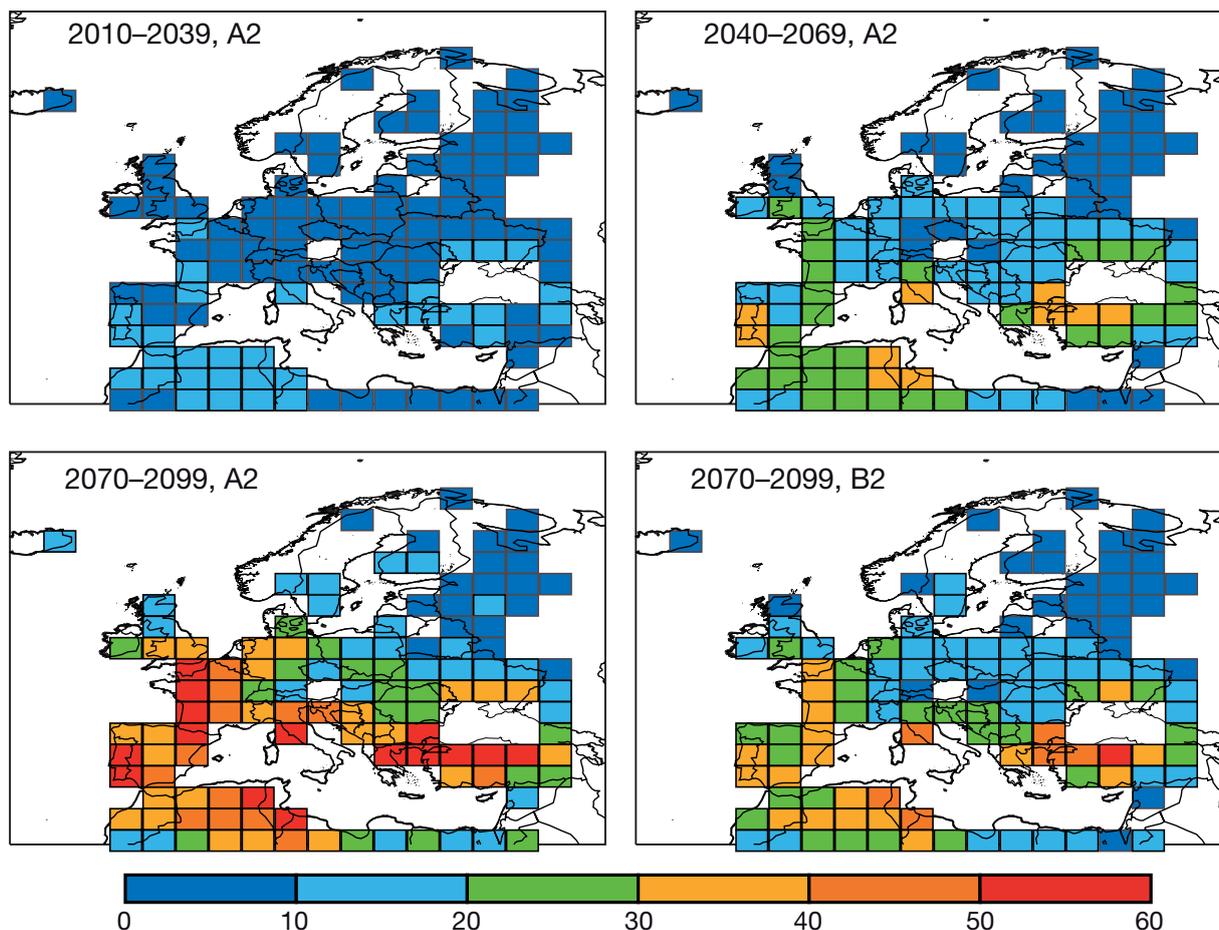


Fig. 11. As for Fig. 10, but for MaxD. Anomalies expressed as percentages of 1961–1990 mean for each grid cell. Results only plotted where the Kolmogorov-Smirnoff test (Lehmann 1975) indicates that distributions of MaxD for 1961–1990 and 2070–2099 (A2) are different at the 5% confidence level. Where the Kolmogorov-Smirnoff test failed, MaxD anomalies are <10%. See Fig. 1 for abbreviations

We further characterize the projected drought conditions using a ‘drought recurrence index’. The drought indicator MaxD identifies the length of the longest dry spell each year. As well as the length, the timing of these dry periods is important for their socio-economic impact. Details of this timing could be strongly dependent on the particular measure of drought used. However, a broad characterisation is made using 2 quantities. First, the date of the centre of the longest dry spell. Second, a quantity indicating whether the dry spell is a recurrent feature of a given time of year:

$$\text{Drought recurrence index} = \frac{\overline{\text{MaxD}}}{\sigma_{\text{DC}}} \quad (1)$$

where  $\overline{\text{MaxD}}$  is the 30 yr mean of MaxD, and  $\sigma_{\text{DC}}$  is the 30 yr standard deviation of the date of the centre-point of the dry spell. If this quantity is greater than about unity, it indicates that the dry spell mean length is similar to or longer than inter-annual fluctuations in the timing of the spell, suggesting a recurrent feature

covering a relatively predictable part of the year. Fig. 12 plots the drought recurrence index. For 1961–1990 there is no well-defined recurrent dry period over much of Europe, except for the eastern Mediterranean. However, by the end of the century under the A2 scenario, much of Mediterranean and western Europe has a relatively well-defined dry period. These dry periods occur during summer, centred on July or August. Together with the projected increases in MaxD, this indicates a rather dramatic change in summer water availability over large parts of Europe.

## 7. DISCUSSION

We have demonstrated the potential for a relatively simple method for scaling climate extreme results from high resolution RCM runs to other time periods and other forcing scenarios. The method uses RCM results

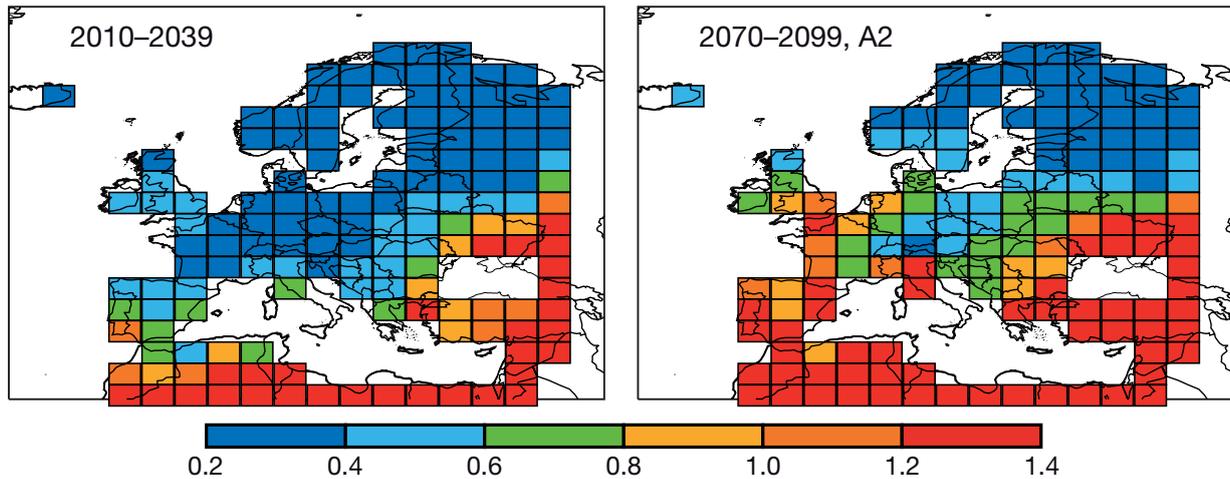


Fig. 12. Drought recurrence index (Eq. 1)

to estimate 30 yr mean relationships between IoEs and  $T_{xa}$ . Values of  $T_{xa}$  from a lower-resolution model are used to scale IoEs to conditions for which the RCM has not been run. Tests highlighting potential problem areas were suggested. The proposed method was shown to be acceptable in reproducing RCM calculations of drought and heat-wave indices over large parts of Europe for an anthropogenic emission scenario not run by the RCM. For some regions, this scaling produced large errors for the heat-wave index. This was mostly in the highlighted problem areas, but further model data would be needed for more rigorous testing of the basic assumptions of the method. Clear curvature in the estimated relationships appeared across substantial parts of Europe, even though the errors in the magnitude of the estimated curvature may be large. However, for drought in particular, such non-linearities could have a large impact on both drought evolution during the century and sensitivity of regional climate extremes to greenhouse forcing.

For heat-waves over much of Europe, the gradient of the 30 yr mean relationship between heat-wave temperature and  $T_{xa}$  in the Hadley Centre HadRM3P RCM decreases with increasing  $T_{xa}$ . That is, for a given rise in annual mean temperature, a bigger heat-wave response is expected at the start of the century than at its end. This suggests that increases in heat-wave temperatures might be felt earlier in the century than would be expected from a linear scaling. However, the sensitivity of heat-wave severity to different anthropogenic emission scenarios is relatively low in HadRM3P. The curvature in the relationship between heat-wave severity and  $T_{xa}$  is easily explainable in terms of soil dryness, and might be expected to be a general result for sufficiently large warming. Such curvature was not evident in a brief study of results from 6 other RCMs, or in the result of Beniston & Diaz

(2004). This may be linked to the fact that these other models mostly show rather less warming than HadRM3P. Also, in all but 1 case, only 60 yr of results were available (as opposed to 180 yr for HadRM3P), making the determination of curvature rather imprecise for these models.

For the chosen measure of summer drought length, the relationship with  $T_{xa}$  was found to have strong positive curvature over southern and western Europe. That is, the gradient of the relationship between drought length and  $T_{xa}$  increases with increasing  $T_{xa}$ . Large errors are expected in the magnitude of the estimated curvature. However, over Mediterranean Europe, non-linear relationships were required to scale, acceptably, HadRM3P drought results to the B2 scenario. Further, positive curvature in the relationship between drought length and  $T_{xa}$  was found to be a reasonably systematic feature over Mediterranean Europe in several different RCMs forced by 2 different global models. These results suggest that, for parts of Europe, rather sudden increases in drought length could occur towards the end of the century, while the magnitude of this increase is strongly dependent on the anthropogenic emission scenario. Under the A2 scenario, much of Mediterranean and western Europe is predicted to have a substantially more robust, persistent summer dry spell by the end of the century.

These results support the view (e.g. Houghton et al. 2001) that linear pattern-scaling techniques should be applied with caution to climate extremes. A quadratic form relating IoEs and  $T_{xa}$  was used for simplicity. To suggest that the quadratic form should be abandoned in favour of a different family of curves, more data, and possibly a detailed investigation into the physical processes, would be required. A second point is that the curvature of these fits could be biased because least-squares regression assumes a uniform variance about

the curve. For the drought indicator (MaxD), and for some grid cells, the variance is clearly larger at the end of the century. Hence, substantial errors should be expected from these curve fits, and further model experiments are required to constrain the results. However, the results still give useful information on the behaviour of climate extremes.

Conclusions based on climate model projections are necessarily limited by uncertainties, particularly in model physics. However, these results are at least 'features to look out for'. The method could potentially be extended to the high-resolution RCM grid, avoiding the upscaling step, but this would add further uncertainty. More reliable curve-fitting methods would also be possible. However, more climate model data is probably required first, to test and constrain the results further.

The method demonstrated here could be applied to other pairs of high- and low-resolution models. Alternatively, results produced using statistical downscaling methods could be substituted for high resolution model calculations. For the Hadley Centre models, the  $T_{xa}$  was found to be a useful parameter to link the global and regional models. For other models, different link parameters may work better. As global models are improved in the future, it should be possible to use less highly-averaged parameters, such as summer mean  $T_{xa}$ . This will improve the reliability of the method. These results might also suggest a different strategy for future RCM experiments. Much more accurate estimation and assessment of the curvature in these relationships, and hence of the evolution and sensitivity to emissions of climate extremes, would be possible if mid-century RCM results were available. For this particular application, results from 2 experiments for 3 different time periods (e.g. 1961–1990, 2030–2060 and 2070–2099) would be much more useful than 3 experiments for 2 time periods. Further investigation using existing model results to explore mechanisms of change and other extreme indices should indicate whether such a change in resource allocation is important.

#### LITERATURE CITED

- Barring L, Holt T, Linderson ML, Radziejewski M, Bindi M, Palutikof JP (2006) Defining dry/wet spells for point observations, observed area averages, and regional climate model gridboxes in Europe. *Clim Res* 31:35–49
- Beniston M (2004) The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations. *Geophys Res Lett* 31: 2022–2026, doi:10.1029/2003GL018879
- Beniston M, Diaz HF (2004) The 2003 heat wave as an example of summers in a greenhouse climate? Observations and climate model simulations for Basel, Switzerland. *Global Planet Change* 44:1–9
- Christensen JH, Christensen OB (2003) Climate modeling: severe summertime flooding in Europe. *Nature* 421: 805–806
- Christensen JH, Carter T, Tummukainen M (in press) Evaluating the performance of regional climate models: the PRUDENCE Project. *Clim Change (Spec Issue)*
- Frich P, Alexander LV, Della-Marta P, Gleason B, Haylock M, Klein Tank AMG, Peterson T (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. *Clim Res* 19:193–212
- Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn* 16:147–168
- Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA (2002) An observationally based estimate of the climate sensitivity. *J Clim* 15:3117–3121
- Hanson CE, Palutikof JP, Livermore MTJ, Barring L and 13 others (in press) Modelling the impact of climate extremes: an overview of the MICE Project. *Clim Change (Spec Issue)*
- Hesterberg T, Monaghan S, Moore DS, Clipson A, Epstein R (2003) Bootstrap methods and permutation tests. WH Freeman, New York
- Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (2001) Climate change 2001: the scientific basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
- Jacob D (2001) A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. *Meteorol Atmos Phys* 77:61–73
- Jones R, Noguer M, Hassell D, Hudson D, Wilson S, Jenkins G, Mitchell J (2004a) Generating high resolution climate change scenarios using PRECIS. Available at [http://precis.metoffice.com/docs/PRECIS\\_Handbook.pdf](http://precis.metoffice.com/docs/PRECIS_Handbook.pdf)
- Jones CG, Willén U, Ullerstig A, Hansson U (2004b) The Rossby Centre regional atmospheric climate model part 1. model climatology and performance for the present climate over Europe. *Ambio* 33:199–210
- Kendall MG (1975) Rank correlation methods. Charles Griffin, London
- Kjellström E (2004) Recent and future signatures of climate change in Europe. *Ambio* 33:193–198
- Kjellström E, Barring L, Jacob D, Jones R, Lenderink G, Schär C (in press) Variability in daily maximum and minimum temperatures: recent and future changes over Europe. *Clim Change (Spec Issue)*
- Klein Tank AMG, Können GP (2003) Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. *J Clim* 16:3665–3680
- Lehmann EL (1975) Nonparametrics: statistical methods based on ranks. Holden-Day, San Francisco
- Lenderink G, van den Hurk B, Meijgaard E, van Ulden A, Cuijpers H (2003) Simulation of present-day climate in RACMO2: first results and model developments. KNMI Technical Report 252, Utrecht
- Mann HB (1945) Non-parametric tests against trend. *Econometrica* 13:245–259
- McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS (2001) Climate Change 2001: impacts, adaptation and vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
- Mitchell TD (2003) Pattern scaling: an examination of the

- accuracy of the technique for describing future climates. *Clim Change* 60:217–242
- Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430:768–772
- Nakicenovic N, Alcamo J, Davis G, de Vries B and 24 others (2000) IPCC special report on emissions scenarios. Cambridge University Press, Cambridge
- Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Clim Dyn* 16:123–146
- Räisänen J, Hansson U, Ullerstig A, Döscher R and 5 others (2004) European climate in the late 21st century: regional simulations with two driving global models and two forcing scenarios. *Clim Dyn* 22:13–31
- Roeckner E, Bengtsson L, Feichter J, Lelieveld J, Rodhe H (1999) Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. *J Clim* 12:3004–3032
- Santer BD, Wigley TML, Schlesinger ME, Mitchell JFB (1990) Developing climate scenarios from equilibrium GCM results. Report No. 47, Max Planck Institute for Meteorology, Hamburg
- Schär C, Vidale PL, Luthi D, Frei C, Haberli C, Liniger MA, Appenzeller C (2004) The role of increasing temperature variability in European summer heatwaves. *Nature* 427:332–336
- Stainforth DA, Aina T, Christensen C, Collins M and 12 others (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433:403–406
- Stappeler J, Doms G, Schättler U, Bitzer HW, Gassmann A, Damrath U, Gregoric G (2003) Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorol Atmos Phys* 82:75–96
- van den Hurk B, Hirschi M, Schär C, Lenderink G and 7 others (2005) Soil control on runoff response to climate change signal in regional climate model simulations. *J Clim* 18(17):3536–3551
- Vidale PL, Lüthi C, Frei C, Seneviratne S, Schär C (2003) Predictability and uncertainty in a regional climate model. *J Geophys Res Atmos*, 108:4586 doi:10.1029/2002JD002810

#### Appendix 1. Bias correction for $T_{xa}$ from HadCM3

Bias-corrected values of  $T_{xa}$  were derived from HadCM3 as follows. For a given year Y1 in HadCM3 results, the anomaly in  $T_{xa}$  with respect to the 1961–1990 mean was ascertained. This anomaly was added to the 1961–1990 mean from HadRM3P, giving a bias-corrected estimate of  $T_{xa}$  for the year Y1. These HadCM3-derived  $T_{xa}$  values were compared with HadRM3P results for the A2 scenario, 2070–2099. For grid cells over Europe, the anomalies with respect to 1961–1990 differ by up to 20%, with a root mean square difference of 10% and a mean close to 0. For most grid cells, the difference between the models in the A2 scenario is less than the uncertainty (from natural variability) in a 30 yr mean of  $T_{xa}$ . Therefore, it seems reasonable to use HadCM3 results with a bias correction to estimate  $T_{xa}$ , and hence the IoEs, for periods and scenarios where HadRM3P has not been run. However, there is a clear large-scale geographical pattern (an approximate meridional feature, with HadCM3 generally warmer over southern Europe and cooler over northern Europe) to the model differences. This pattern may represent a systematic discrepancy, and a further correction was applied as follows. The ratio between HadRM3P and HadCM3  $T_{xa}$  anomalies for 2070–2099 (A2 scenario) was calculated at each gridpoint. Then, HadCM3  $T_{xa}$  anomalies for other years and other scenarios were all scaled using these ratios. Since only 30 yr of B2 scenario results were available to test this correction, no significant advantage or disadvantage of such a correction could be found. Whether appropriate or not, this correction does not change the conclusions regarding the significance of ‘prediction errors’ and has little effect on the projected evolution of IoEs. As an alternative predictor, the June to August mean of  $T_{max}$  was briefly tested. However, the differences between HadRM3P and HadCM3 anomalies in this quantity reached 50% over parts of Europe, so it was clearly not useful for our purposes.

#### Appendix 2. Definition of ‘coverage’

The method is expected to work best when the mean difference in  $T_{xa}$  between the 2 data periods (1961–1990 and 2070–2099, A2 scenario in this case) is not large compared to the inter-annual variability. This is quantified as:

$$\text{coverage} = 100 \frac{2\sigma_f + 2\sigma_r}{T_{xaf} - T_{xar}} \quad (\text{A1})$$

where  $T_{xaf}$  and  $\sigma_f$  are, respectively, mean and SD of  $T_{xa}$  for 2070–2099, and  $T_{xar}$  and  $\sigma_r$  are mean and SD for 1961–1990

#### Appendix 3. Mann-Kendall test of Issue 2

Signs of Issue 2 (Section 3) were tested as follows. First, for each year of data, the curve fit was subtracted from the value of IoE for that year, giving a set of 180 residuals (for 180 yr of data). This is like a de-trending procedure, with  $T_{xa}$  (instead of time) as the independent variable. If the curve fit is good, then the residuals should scatter evenly about 0 and show no correlation with  $T_{xa}$ .

Taking the residuals for 1961–1990 only, the statistical significance of any correlation between the residuals and  $T_{xa}$  was tested using the non-parametric Mann-Kendall rank correlation test (Mann 1945, Kendall 1975). The same procedure was repeated for the residuals of 2070–2099. If either of the 2 anomaly clusters shows significant correlations with  $T_{xa}$ , this means that the coupled inter-annual variability for that cluster does not lie along the curve fit. This does not necessarily mean the curve fit is unrepresentative of the 30 yr mean relationship of IoE with  $T_{xa}$ , but caution is advised.

**Appendix 4.** Significance of 'prediction errors'

Statistical significance of 'prediction errors' (defined in Section 5) was estimated as follows. Confidence intervals (CIs) for this 'prediction error' were estimated for each grid cell using bootstrap re-sampling with replacement (with 1000 repetitions) and a *t*-test (bootstrap methods described by e.g. Hesterberg et al. 2003). At the start of each sampling step, the 90 yr of data from 1961–1990 were re-sampled randomly, as in the standard bootstrap method. The same was done independently for both the 90 yr of data from 2070–2099 (A2 scenario) and for the 30 yr under the B2 scenario. Both the curve fits and the 30 yr mean of the

HadRM3P B2 results, and hence the 'prediction error', were all recalculated based on this resampled data. Standard application of the bootstrap method then allowed estimation of CIs. The sizes of the estimated CIs were dominated here by the uncertainty in mean of the HadRM3P B2 scenario results, because only 30 yr of B2 results were available, compared to 180 yr for the curve fits. Of course, this assumes that a quadratic least-squares fit is correct: the real curve fit uncertainty is much larger. The significant *p* value for a given 'prediction error' may be derived directly from these CIs.

*Submitted: April 6, 2005; Accepted: February 25, 2006*

*Proofs received from author(s): June 12, 2006*