WHAT CAN PRESENT CLIMATE MODELS TELL US ABOUT CLIMATE CHANGE?

RASMUS E. BENESTAD

The Norwegian Meteorological Institute, P.O. Box 43, N-0313 Oslo, Norway
E-mail: rasmus.benestad@met.no

Abstract. Climate models are evaluated in terms of their ability to describe the past climatic changes. Past climatic trends are inferred from fitting a truncated Taylor series to the observational record and an ensemble of downscaled results from climate models. Analytical expressions are derived for the warming rates associated with long-term temperature trends using simple calculus. Different trend models are compared, and a third-order polynomial gives the best description of the past winter warming over southwestern Scandinavia. The coefficients from the regression analysis are used in an objective comparison of the past climatic evolution in the models and the observations. Comparisons between the temperature trends from observations and the output from an ensemble of various climate models suggest that single climate model scenarios do not provide a reliable description of the climatic evolution. Ensembles of state-of-the-art climate models, on the other hand, capture the main features of the past climatic evolution. However, there has been an interval with pronounced local winter warming over Scandinavia in the past, which is not reproduced by the majority of climate models. It is difficult to say whether this accelerated warming event was part of natural decadal variations or induced by external factors. The climate models may not yet be able to predict similar local episodes for the future if they are related to events unaccounted for, such as solar activity or volcanism.

1. Introduction

Atmospheric greenhouse gas concentrations, such as CO_2, have increased since the industrial revolution as a result of an accelerated fossil fuel consumption (Houghton et al., 2001). The Swedish scientist Svante Arrhenius (Arrhenius, 1896) proposed more than 100 years ago that the CO_2 may cause a warming of Earth's surface. Observations do indeed suggest warmer global mean surface temperatures, melting of sea-ice, accumulation of heat in the upper oceans, and rising sea levels (Houghton et al., 2001). Since the prospect of a climate change is of major concern, climate models have been used to make scenarios for the future. Hence it is essential to ask what the climate models can tell us.

The nature of the climate system is chaotic (Lorenz, 1967) and its exact state cannot be predicted years ahead. However, a long-term systematic change in the boundary conditions may influence the climate statistics and the resulting long-term climatic trends may still be predictable (Palmer, 1996). The question to ask then is whether the climate models are capable of predicting these long-term trends. It is difficult to use single model scenarios for predicting the climatic evolution

Climatic Change **59:** 311–331, 2003.
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*

because different scenarios give different descriptions of local climatic trends (Houghton et al., 2001), and it is not known a priori which model is closest to the truth. An alternative approach is to use an ensemble of climate scenarios from a number of different climate models (Räisänen and Palmer, 2001; Benestad, 2002). An evaluation of the models' ability to predict the past climatic evolution gives an indication of the credibility of the climate scenarios. The model temperature reconstructions are not affected by the observed temperatures and the only factor linking the simulations to the observations is the prescribed greenhouse gas emissions (and a constant atmosphere-ocean flux correction that imposes a realistic present-day climate). The model simulations therefore provide 'predictions' of the past that are equally independent of the observations as the predictions for the future.

It is often the climatic *changes* that are of primary interest for societies, and climate models need to be evaluated by comparing the reconstructed warming for the past with the corresponding observed warming. A good agreement gives increased confidence in the future climate projections whereas large discrepancies point to serious model shortcomings. In other words, the minimum requirement for credible climate change scenarios is that the climate models are able to reproduce the long-term climate evolution in the past.

Long-term trends in time series are often inferred from a linear best-fit obtained through regression analysis. A linear approach is useful for studying trends over centennial time scales, but may not be the best way to estimate trends for shorter periods such as 2000–2050. The Nordic region experienced a rapid warming before 1930 and after the 1960s, interspersed with an interval of cooling, and these slow variations are not captured by a linear trend over the entire observational record (1890–2001). The time scale of these warming and cooling intervals is of the order 20–50 years, which is similar to the time horizon of the climate change scenarios.

In contrast to a linear model that assumes a constant rate of temperature change, higher-order polynomial trend approximations to the 1890–2001 temperature record may account for the warm winters during 1930s–1940s, the cool 1960s–1970s, and the latest enhanced warming. For example, cubic (third-order polynomial) fits give a good description of the observed trends seen in the temperature record. The motivation for adopting polynomial models to describe the long-term trends is that we want to make use of all the available information about the past evolution. There is little point in trying to fit a linear model that we a priori know does not (for our purpose) give an adequately good description of the past climatic trends. There have in the past been a few studies which have used trend models with higher order than one (linear). Polsky et al. (2000), for instance, used a second order polynomial to describe the temperature trend in the northeastern part of the U.S.A.

The use of higher-order polynomial models (which mathematically are similar to Taylor approximations) allows for 'higher fidelity' in terms of changes in the warming rate. The polynomial model approach is analogous to principal component analysis (PCA) and the geophysical variant of PCA known as empirical orthogonal functions (EOFs) which describe spatial structures: by including higher orders, the spatial description becomes increasingly more detailed. Another analogy is the reconstruction of a signal in terms of a Fourier series with various degrees of truncation. The truncation point which determines the order of the polynomial is chosen depending on the use of the data and the expectation about model skill. The higher-order polynomial models give a more detailed description of the trend within the calibration interval, but is inappropriate for predictions (extrapolations) outside this interval (this is also generally true for Fourier transforms). The test of similarity (model validation) between trends derived from the models and observations becomes more stringent with the inclusion of more 'details' (higher orders). Since the purpose of the polynomial trend-fitting is to give an optimal description of the trend in dependent variables, the issue about over-fit is not important in this context.

2. Method and Data

A polynomial fit to the long-term trend (y) may be found using multiple regression against time (x) according to:

$$\hat{y} = c_0 + c_1 x + c_2 x^2 + c_3 x^3 \dots \qquad (1)$$

A cubic model is described by Equation (1) where the series is truncated after the x^3 term. For records of spatial mean temperature, the regression analysis may take into account the varying data quality by using weights based on the number of observations in the estimate of the spatial mean (weights $=$ no.obs^2/max(no.obs)^2). The warming rate may be taken as the first derivative of the cubic-fit (Equation (1)), which means that an analytical expression can be found for the long-term warming rate:

$$\frac{d\hat{y}}{dx} = c_1 + 2c_2 x + 3c_3 x^2 \dots \qquad (2)$$

A climate change scenario can therefore be described in terms of the three coefficients c_1, c_2 and c_3 for a given region if a cubic model is adopted, as these give an approximate description of the long-term warming, provided that the cubic-fit gives a good representation of the long-term trend. The coefficients can be estimated using an ordinary regression package. In order to get the most stable estimates (smallest errors) x should be in the range $[-1, \dots, 1]$ (where the x intervals, Δx, are determined by N, the number of years) rather than the actual years (i.e., [1880, …, 1999]).[1] If this standard is adopted for x, then the regression coefficients together with two references years for the beginning and the end of the time period will give an unambiguous description of the long-term evolution. For comparison studies, it is important to compare trends fitted to series over exactly the same time interval. In the following analysis, the same trend models were used for the climate model results as were used to describe the trend in the observational records.

A 95% confidence interval for these trend fits can be estimated from the standard errors estimates obtained for each of the regression coefficients,[2] c_1, $c_2, \ldots, c_n$. It is assumed here that the error in these estimates can be treated as being independent of each other.[3] Hence, in order to compute the confidence interval for all coefficients taken together (total error η_i), the uncertainty interval is estimated by taking the square-root of the sum of the contribution (squared) from each term separately (setting the error of the other terms to zero):

$$\eta_i^{\pm} = \sqrt{\begin{aligned}&[c_0 + c_1x_i + c_2x_i^2 + c_3x_i^3 - (c_0 \pm s_0) - c_1x_i - c_2x_i^2 - c_3x_i^3]^2 + \\ &[c_0 + c_1x_i + c_2x_i^2 + c_3x_i^3 - c_0^2 - (c_1 \pm s_1)x_i - c_2x_i^2 - c_3x_i^3]^2 + \\ &[c_0 + c_1x_i + c_2x_i^2 + c_3x_i^3 - c_0^2 - c_1x_i - (c_2 \pm s_2)x_i^2 - c_3x_i^3]^2 + \\ &[c_0 + c_1x_i + c_2x_i^2 + c_3x_i^3 - c_0^2 - c_1x_i - c_2x_i^2 - (c_3 \pm s_3)x_i^3]^2\end{aligned}} \quad (3)$$

The confidence intervals for the rate of change were estimated in a similar fashion as in Equation (3), but using the expression in Equation (2) instead Equation (1). Due to the non-linear character of the polynomial models, the expression for the confidence intervals gets complicated. The analysis was carried out in the *R* data analysis language (Ellner, 2001; Gentleman and Ihaka, 2000).[4]

The results presented in this paper are derived from monthly mean temperature observations taken from Nordklim Dataset (Tuomenvirta et al., 2001) that covers the period 1890–1999, as well as gridded (HadCRUT) global land and sea temperature data from the Hadley Centre and the Climate Research Unit at the University of East Anglia (Jones et al., 1998). The northern hemispheric mean data are presented here instead of the global mean because the quality of the former is believed to be higher, however, the same analysis for the global mean gave similar conclusions. The multi-model ensemble includes 14 different climate scenarios based on models described in the Intergovernmental Panel of Climate Change (IPCC) reports (Table I).

The information about the long-term temporal evolution contained in the coefficients can be used in an objective and quantitative measure of the ensemble's ability to describe the observations. In this context, we are only interested in the rate of change, and only the three coefficients $c_1, \ldots, c_3$ will be compared here. These can be represented as points in a 3-dimensional Eucledian space ($\mathcal{R}^3$). The theoretical spread of the coefficients from a hypothetical infinite model ensemble is assumed to have the form of an ellipsoid, which for our finite sample, can be found by taking the ensemble mean values as the centre point, with the 1.96 times the ensemble standard deviation as the semi-axes (the ellipsoid then contains the 95% confidence region). In this kind of frame work, it is possible to check whether the coefficients for the observed fit is within this confidence volume. This confidence

Table I

A summary of the climate model outputs used in the analysis. The second column show the scenarios analysed from the various models, which include integrations accounting for direct effects of aerosols only (GSA) as well as direct-plus-indirect effects (GSDIO). The four integrations from the HadCM2 model, 3 from the CCCma, and 2 from ECHAM3 are different members of ensemble integrations with the same model physics and external forcing. The greenhouse scenarios are the IPCC IS92a

Climate model	Scenarios	Reference
ECHAM4/OPYC3	1GSA + 1 GSDIO	(Roeckner et al., 1992; Oberhuber, 1993)
HadCM3	1 GSA	(Gordon et al., 2000)
HadCM2	4 GSA	(Johns et al., 1997)
ECHAM3	2 GSA	(Cubash et al., 1997)
CCCma	3 GSA	(Flato et al., 2000)
CSIRO	1 GSA	(Gordon and O'Farrell, 1997)
GFDL-15a	1 GSA	(Manabe and Stouffer, 1996)

region is based on the ensemble spread, whereas the confidence intervals derived from Equation (3) describe the uncertainties related to the individual trend-fits. For small samples, the estimate of the standard deviation is often unstable, and the coefficients may be compared directly instead of checking whether the observations are inside or outside the ellipsoid. Points with coordinates smaller than the semi-axes can only be outside the ellipsoid if two or more coordinates are similar to that of the semi-axes. By sub-sampling the ensemble values, it is furthermore possible to get an approximate estimate for the ensemble size required to span the observed trends.

Although global mean and hemispheric mean values serve as useful indicators for the climatic state, they often have no direct implication for societies. Local climate scenarios, however, are of greater interest as they are directly related to the local impacts on societies and eco-systems. A focus on local and region scales puts greater demands on the climate simulations, since geographical factors not adequately resolved by the climate models become increasingly important. Small-scale processes, represented by crude sub-grid model parameterisation in the models, influence the simulation of the local climate. Furthermore, the ratio of long-term trends to natural variations (signal-to-noise) diminishes as the spatial scale increases. In fact, the climate models do generally not give a good representation of local and regional[5] climates (Grotch and MacCracken, 1991). One way to overcome this problem is through downscaling analysis where well-established relationships between large-scale anomalies and local climate variations are used to infer local climate change.

The downscaled scenarios shown here are the same as in Benestad (2002) and the downscaling methodology incorporates a step-wise screening based on cross-validation and canonical correlation analysis (CCA), using common principal components (Flury, 1988; Sengupta and Boyle, 1998; Barnett, 1999) as basis functions. The downscaling methodology is documented and tested in Benestad (2001b). The predictor was the anomalous 2-meter temperature field, which is highly correlated with the local climate series ($r \sim 0.9$ in January and $r \sim 0.7$ in July). All the scenarios were based on predictors covering the 'Nordic' region, 20° W–40° E, 50° N–75° N, except for member 1 of the HadCM2 ensemble which due to an accidental omission covered 0° W–30° E, 55° N–75° N. It is important to keep in mind that the downscaled scenarios may be sensitive to the choice of predictor area, and the size of the optimal predictor region may vary with the season (Benestad, 2001b).

3. Results

The polynomial trend fitting was applied to the Oslo temperature (Figure 1) as well as other series. For the winter (December–February) Oslo temperature, the results suggests that the long-term evolution can be best described as a third-order polynomial (Figure 1a). The cubic-fit furthermore gives the lowest probability for a 'coincidentally good' fit (i.e., the p-value from the ANOVA (Wilks, 1995)) and a high estimate for the variance accounted for by the regression, R^2. Similarly, the cubic-fit gives the best description (in terms of R^2 and p-values) of the winter temperatures in Tromsø (Norway), Kjøremsgrendi (Norway), Gotland (Sweden, in the Baltic Sea), Stockholm, Torshavn (Faeroes in the North Sea), and Jyväskylä (Finland), but not for other places like Helsinki (not shown). For Copenhagen, there was no trend-model that was clearly superior.

An alternative way to study trends is to low-pass-filter the data. Although different low-pass-filtering may give slightly different results and may introduce artifacts such as the Slusky-Yule effect (spurious oscillations), a filter-based approach often gives more stable results than a trend-fit. The disadvantage with the low-pass-filtering is primarily the cut-off at the ends. Figure 1 shows a number of low-pass-filtered curves where different window widths have been used. The cut-off near the ends depends on the filter window width, and shorter windows give smaller cut-off. The trade-off with smaller window width is a stronger sensitivity to the interannual variations ('noise'). The longer windows, only give valid results for a small part of the mid-section of the interval. Although the polynomial trend-fitting is expected to give less stable results near the end points, Figure 1 shows that there is a good agreement between the cubic trend and the filtered curves (10-year window). The advantages with the trend-fit approach, in addition to having no cut-off near the end, is that the tendencies can be calculated directly from the coefficients, that confidence intervals can be computed, the analysis can employ

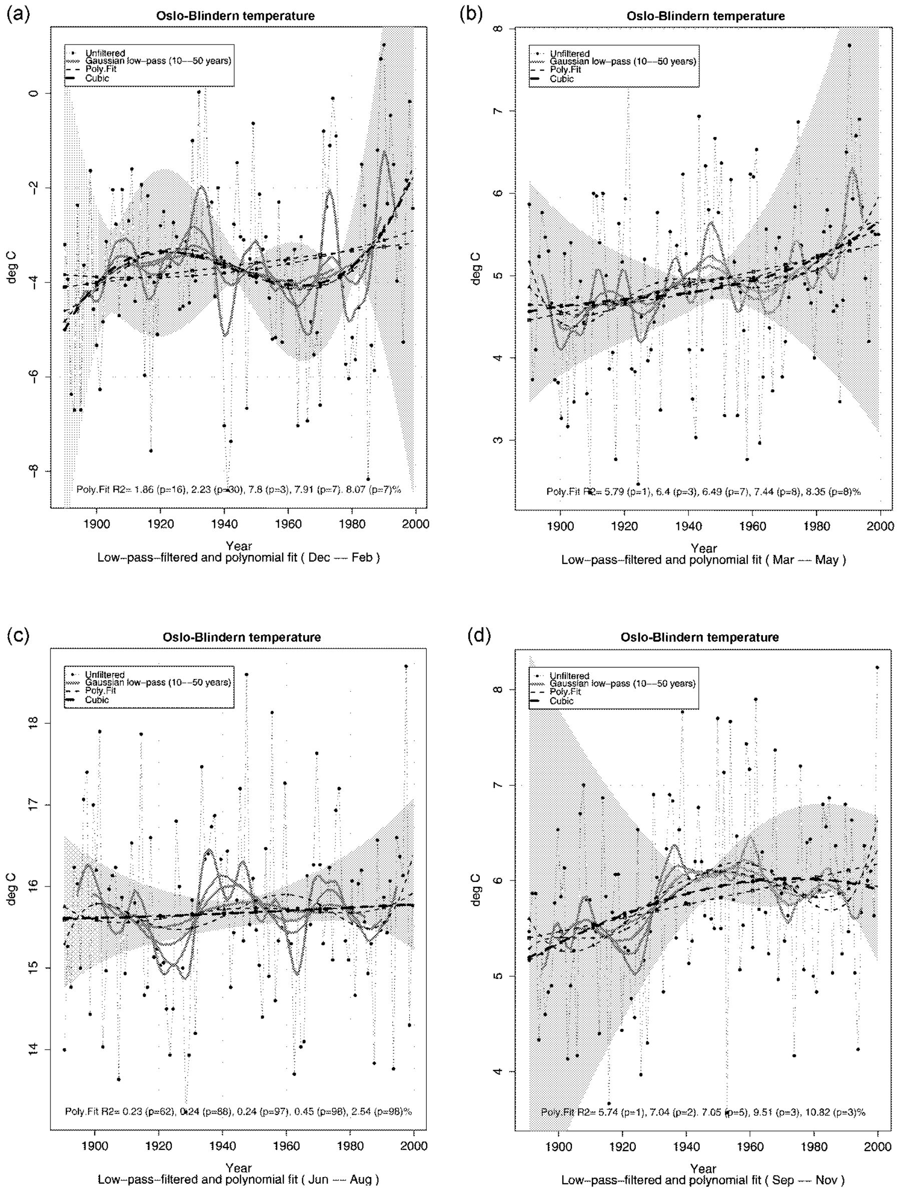

Figure 1. The winter (a), spring (b), summer (c) and autumn (d) temperature in Oslo shown with 1–3 order polynomial trend fits and 10–50 year low-pass filtered values. The text in grey near the bottom of the figures gives the R^2-estimates from the polynomial fit for orders 1–5 and the corresponding p-values in parentheses (both in %). The grey shaded region is the 95% confidence area of the cubic-fit.

weighting for the data quality, and that the temporal evolution can be described in terms of a few parameters (for a cubic model: 3 coefficients, and two reference years → 5 parameters in total). Furthermore, the trend-fits can more easily deal with missing data, whereas small gaps in the series lead to large holes in the filtered versions.

In spring (March–May), the warming rate has been approximately constant, and hence the best model for the long-term evolution is the linear-fit (lowest p-value, 1, and an R^2-value which is slightly lower than the second and third order polynomials). For the Stockholm spring temperature, the quadratic model gave the best description of the past evolution with an acceleration of the warming at the end of the period. A recent cooling can be seen in the Reykjavik, Stykkisholmur (Iceland), and the Torshavn spring-time temperature records (not shown), for which the R^2-estimates and the p-values from the regression analysis suggest that fourth and fifth order polynomials give the best description (not shown).

There has been no warming in Oslo during the summer (June–August), and none of the models yield coefficients that are statistically different from zero (Figure 1c). In Torshavn, Jyväskylä and Tromsø (Norway), on the other hand, the records suggests there has been a long-term summer warming that is best described by a fourth–fifth order polynomial.

The autumn (September–November) temperatures in Oslo decreased since 1960 until very recently (extremely high value for 2000, Figure 1d). This decrease cannot be captured by a linear trend-fit, but the higher-order polynomials give a better description of the temporal evolution (the R^2-estimates are highest for the higher orders and the p-value stays below 5%). Similar evolution can be seen in the records from Jyväskylä, Copenhagen and Tromsø (not shown), and in these cases the best description of the long-term trends are obtained from fourth–fifth order polynomial models. The long-term Helsinki autumn temperature evolution (not shown) is probably best described in terms of a quadratic equation.

A comparison was made between the residuals from the various trend models (Figure 2, showing the results for the Oslo temperature) to check whether there were important features that were not captured. All the residuals had a distribution that was close to being Gaussian, and the difference between the different models was small with greatest scatter in the tails of the distribution. Small differences are expected between the various trend models when the trends are weak. Figure 2a indicates that residuals from the cubic-fit lie closer to the straight (for a perfect Gaussian distribution) line than other models, and that the residuals from the linear and quadratic models are slightly higher than expected for values near ±1.5 standard deviations. These results further suggest that the cubic model gives a better description of the long-term evolution of the winter temperature in Oslo. For the spring, summer and autumn results, there is less of a systematic difference. The time series of the residuals (not shown) look like white noise processes[6] and do not appear to contain any structure.

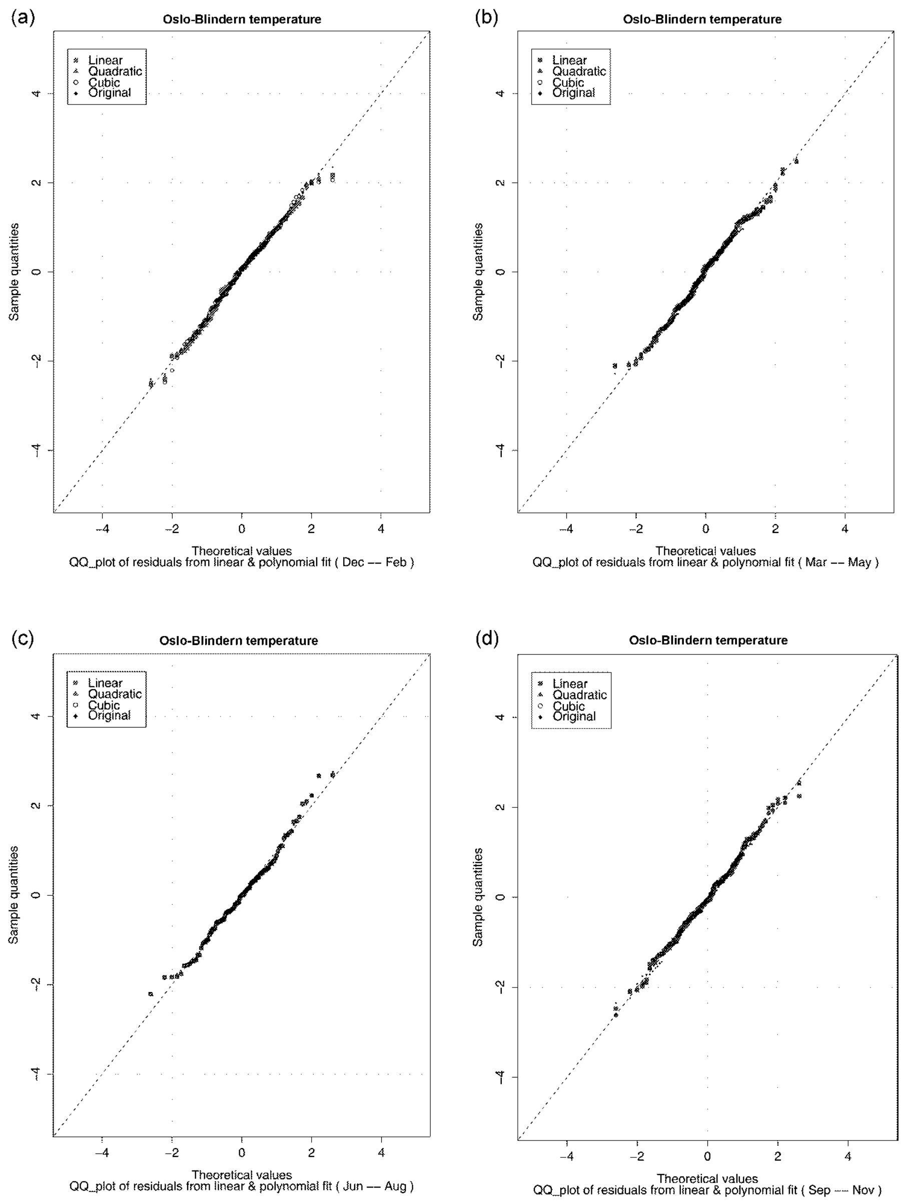

Figure 2. Normal qq-plots of the regression residuals obtained for the linear, quadratic, and cubic models as well as for the original data. The data were standardised prior to the analysis. The data shown are for the winter (a), spring (b), summer (c) and autumn (d) temperature.

Figure 3 shows how the derived tendencies (rate of change) change with time for the Oslo temperature. These results were obtained using Equation (2). The cubic model clearly describes much faster winter-time warming than the lower-order trend models in the early and late parts of the period.

The reconstructions of the past northern hemispheric and the 58° N–70° N, 5° E–40° E annual mean temperature are shown in Figure 4 together with corresponding estimates based on observations. The trend analysis shown in Figure 4 used a weighted regression based on the number of valid observations used for estimating the spatial mean values. Super-imposed on this figure are the cubic-fits according to Equation (1). The comparison between the estimated annual mean warming rates from the climate models and the observations shows that the observed past warming tends to be within the multi-model ensemble spread. The climate model simulations in Figure 4a are similar to the observations, although most of the climate models indicate slightly warmer conditions after the 1990s. Figure 4b, however, shows that the rapid warming in the early 20th century is more pronounced when the temperature averaged over a smaller area, and this warming is not readily seen in the regional mean values derived directly from the climate model outputs.

An interesting observation from Figure 4 is that the interannual variability in the GCM fields often is more pronounced than in the actual observations (the NCEP re-analysis). Eight of the 14 models give stronger variations than seen in the observations, and the model-to-observed variance ratios in the annual mean temperature vary from 0.3–1.6. The strongest interannual variations were predicted by the Canadian models (CCCma: ratio of 1.6) whereas the ECHAM3 (ratios of 0.3 and 0.9 for the northern hemispheric mean and 58° N–70° N, 5° E–40° E mean respectively) and the GFDL (0.6 for the 58° N–70° N, 5° E–40° E mean temperature) models described the weakest interannual variations.

Figure 5 presents a series of monthly mean temperature from Oslo for 4 different seasons. The downscaled scenarios have variations of similar magnitude as the observations, without the need of inflation adjustments (von Storch, 1999). The high variance of the downscaled scenarios can be explained in terms of the high correlations between the predictor fields and the predictands, similar spatial structures in climate model results and observations, and the stronger variance seen in many of the climate model outputs. The inter-annual variations are weaker in the spring, summer and autumn, both in the observations and the local scenarios. The long-term trend may be difficult to define if prominent inter-annual variations are present.

The individual scenarios give different descriptions of the past warming, and it is clear that one single scenario cannot give a reliable description of the climatic evolution. The observations tend to lie within the range of model values, suggesting that the multi-model ensemble gives a good reconstruction of the past climatic evolution. The same analysis for other locations (Norway: Bergen, Tromsø and Værnes; Sweden: Stockholm, Gothenburg and Stensele; Finland: Helsinki, Turku,

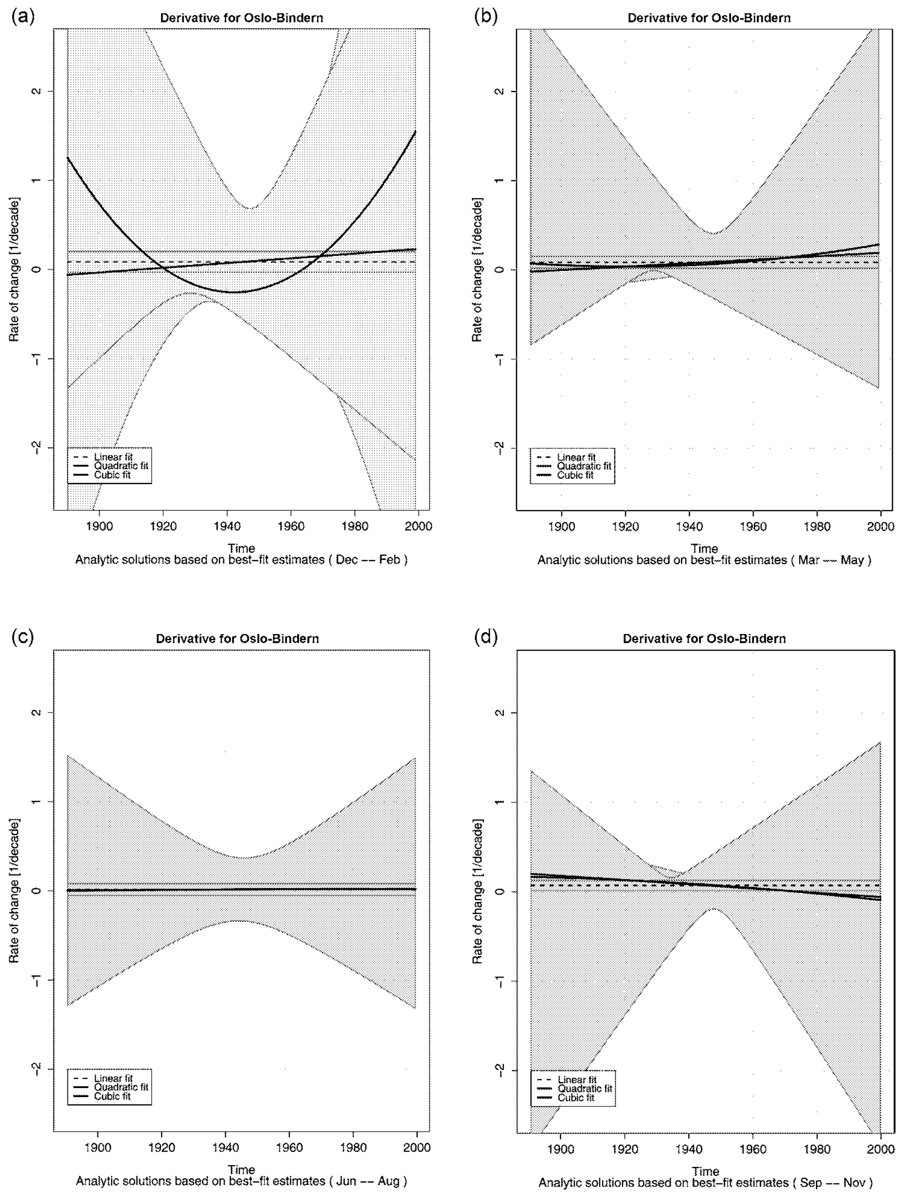

Figure 3. The temporal evolution of the warming rate estimated by taking first time derivative described by Equation (2). The different panels show the warming trends in Oslo for winter (a), spring (b), summer (c) and autumn (d). The grey shaded region is the 95% confidence area.

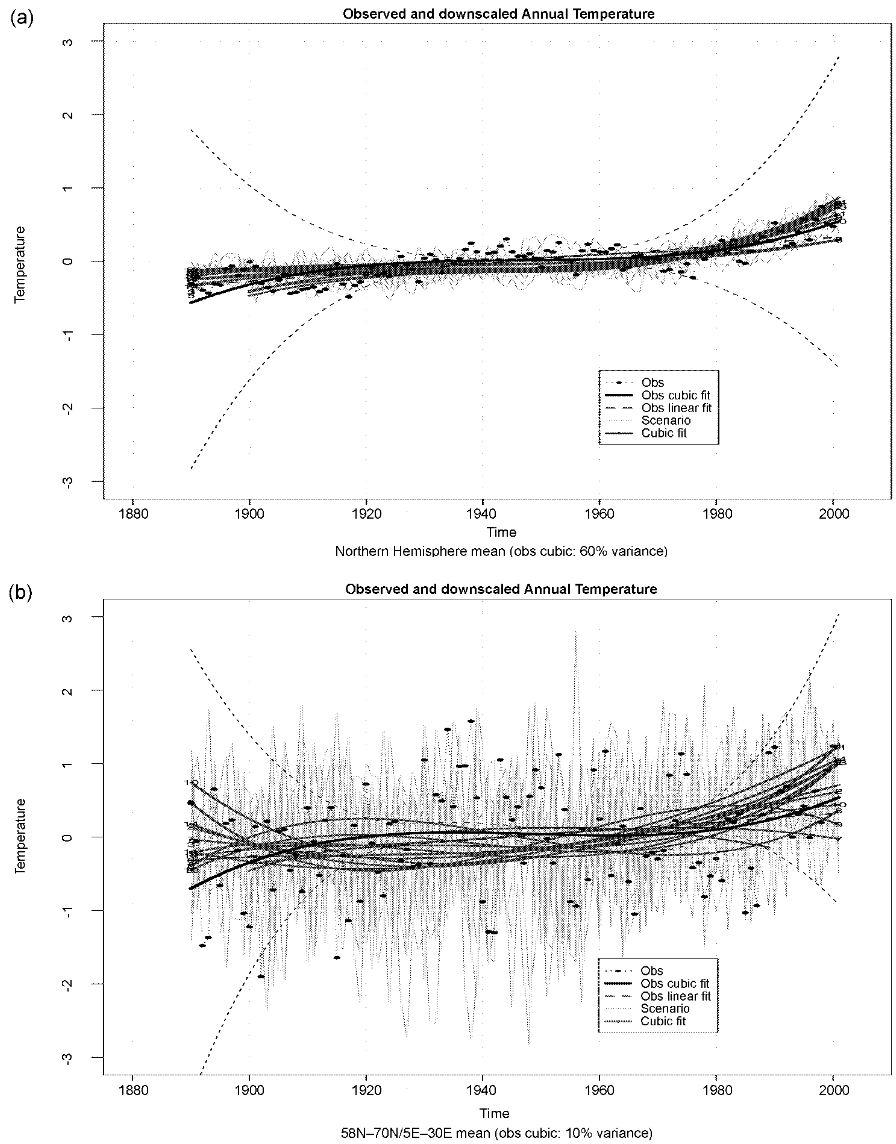

Figure 4. The spatial mean temperature for the northern hemisphere (a) and the 58° N–70° N, 5° E–40° E region (b), estimated from monthly combined mean observations University of East Anglia and the Hadley Centre, and climate scenarios (Table I). The cubic-fits are shown as thick black (observations) and grey (models) curves. A linear fit is also shown for the observations (thin dashed). The numbers mark the different model scenarios: 1 ECHAM4 GSDIO, 2 GFDL-19 GSA, 3–5 CCCma GSA ensemble, 6–7 ECHAM3 GSA ensemble, 8 HadCM3 GSA, 9-12 HadCM2 GSA ensemble, 13 CSIRO GSA, and 14 ECHAM GSA. The regression used weights based on the number of valid observations used for the estimation of the mean value.

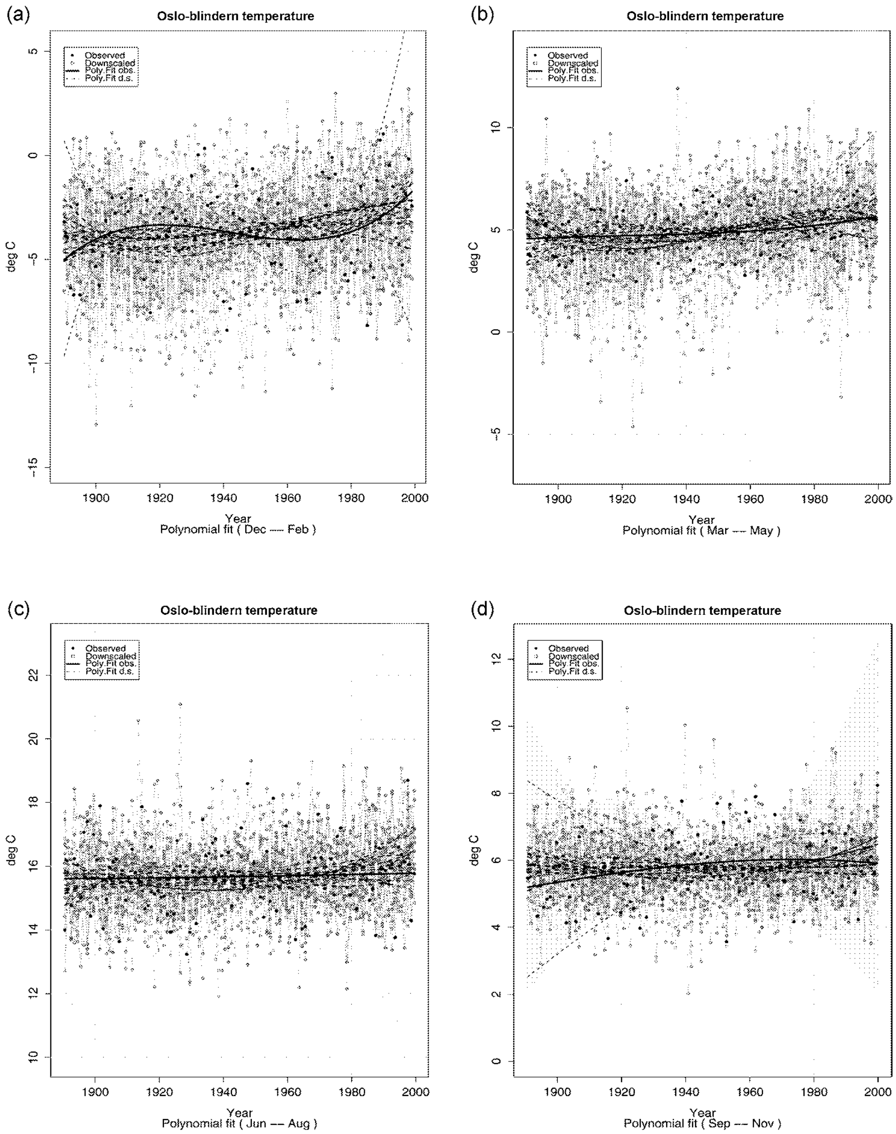

Figure 5. Empirically downscaled Oslo monthly mean temperature scenarios for January (a), April (b), July (c), and October (d). The thick, dashed, dark grey curves represent the model scenarios whereas the thick black line shows the trend according to the observations. The light grey region shows the combined confidence region from the individual climate models whereas the thin, black, dashed lines mark the confidence interval derived from the observations.

Kuopio and Sodankylä; Denmark: Copenhagen and Vestervig – not shown) gives similar results, and regional averages (58° N–70° N/5° E–30° E) lie somewhere between the downscaled and the hemispheric mean scenarios.

Figure 6 presents warming rate estimates derived from the cubic-fits shown in Figure 5, based on Equation (2). Each curve represents a climate 'scenario' for the past. A close agreement between the linear trend from the observations and the models (not shown) may suggests that models generally give a credible reproduction of the warming rate over the entire period (1890–1999). However, the details revealed by the cubic-fits suggest that the rapid winter warming in Oslo before 1930 is captured only by one of the models (Figure 6a). Most of the trends derived from the climate model results are outside the 95% confidence interval for the observations (thin, black, dashed lines) during the 1940s, when the models describe a general warming while the observations suggest an interval with cooling. The observed tendency, however, falls within the confidence intervals of the multi-model ensemble during the 1940s.

The ratio of the variance in the residuals from the trend-analysis of the ensemble to the residual of the observations is 0.43–2.47, suggesting that range of short-term variance in the climate model ensemble spans the observed power. The models also indicate multi-decadal variations, which are as prominent as seen in the observations. Several models indicate a brief interval of cooling at the beginning of the record, which also can be seen in Figure 4b. For spring, the agreement between the observed and downscaled long-term trends is remarkable for the cubic-fit, while linear trend estimates point to a slightly stronger warming in the models (not shown). The observed autumn temperature has dropped slightly since 1980, whereas most of the models predict an accelerated warming. The observed tendencies nevertheless lies within the multi-model ensemble range.

The coefficients $c_1, c_2, \ldots$ obtained from the regression can be used to assess the climate models in terms of reproducing past climate changes. Figure 7 illustrates the concept of the climate model ensemble confidence region in two dimensions (c_1, c_2 and c_1, c_3). In two dimensions, the 95% confidence region is an ellipse whose semi-axes are based on the respective ($1.96\times$) standard deviation of the climate model ensemble coefficients. For small ensembles, an exact estimation of the confidence region is difficult. Furthermore, coefficients exceeding the semi-axes always imply that the observations are outside the confidence region. The hatching in Figure 7a marks the regions where the values of c_1 and c_2 are less than the semi-axes but where the observations nevertheless fall outside the confidence region. Figure 7a suggests that the trend coefficients for the winter temperature in Oslo is just outside the confidence region, but that the error bars associated with these estimates extend well into this ellipse. A similar analysis for c_1 and c_3 is shown in Figure 7b which suggests that the one ensemble member (HadCM2 member 3) that describes a similar trend as the observations is an outlier and is outside the multi-model ensemble confidence region.

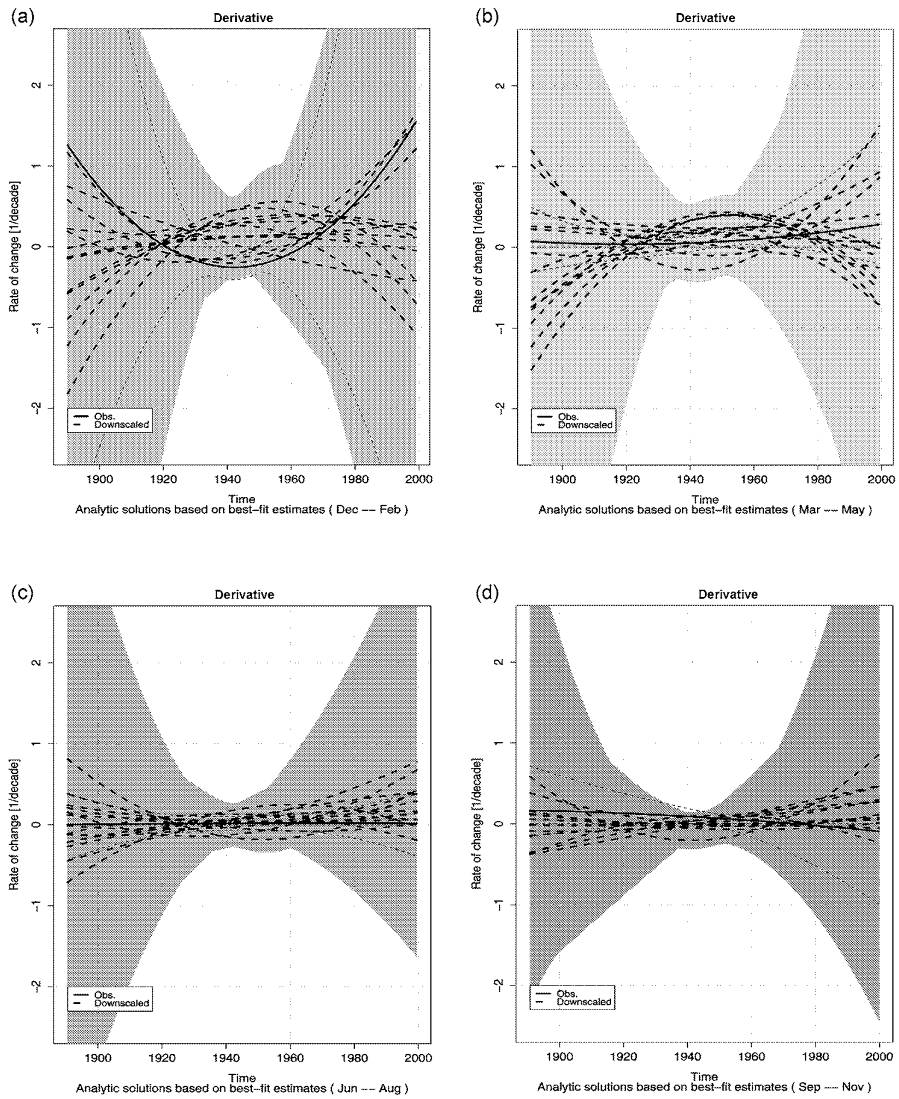

Figure 6. The warming rate derived from the cubic-fit trends shown in Figure 5. The thick, dashed, dark grey curves represent the model scenarios whereas the thick black line shows the trend according to the observations. The light grey region shows the combined confidence region from the individual climate models whereas the thin, black, dashed lines mark the confidence interval derived from the observations.

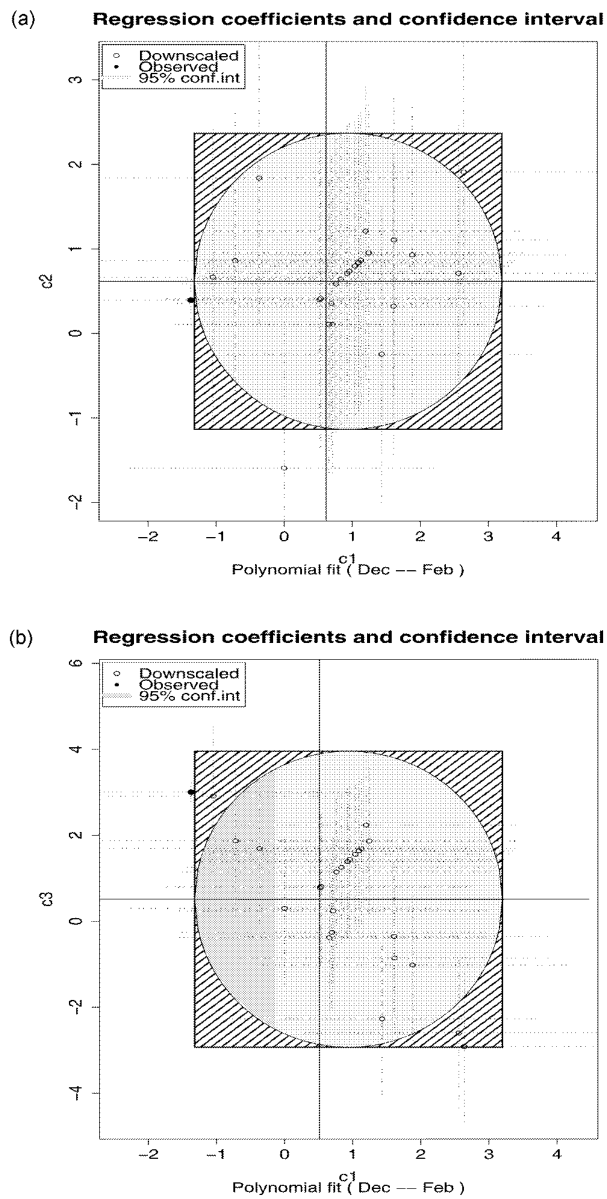

Figure 7. The confidence region for two coefficients c_1 and c_2 for the linear and quadratic terms respectively (ellipse, shown as a circle on these stretched figures) (a) and for c_1 and c_3 representing the linear and cubic terms (b). The polynomial coefficients are derived from the climate model ensemble (open circles), and from observations (filled circle). The dashed lines indicate the error-bars associated with these estimates. The hatched regions show the region where coefficients are smaller than the corresponding semi-axes, but where the two estimates taken together fall outside the confidence region.

A simple comparison between the coefficients $c_1, c_2, \ldots$ derived from observations and corresponding confidence interval from the multi-model ensemble will also give a good indication of whether the climate models can reproduce past trends or not. Box-plots, such as those shown in Figure 8, can be used to make such simple comparisons. The fact that the coefficients for the observed trend tend to be within the ensemble range or near the confidence limits for the climate model ensemble suggests that the climate model ensemble size in general is not too small, with the possible exception for the winter warming trend.

4. Discussion and Conclusion

The use of polynomials to describe the temporal evolution has been examined and the advantage over a linear trend approximation and low-pass-filtering has been demonstrated. Linear trend models do not describe slow variations in the warming rate and do not capture features such as the rapid warming at the beginning of the 20th century, the cooling between the 1940s and 1960s, and the most recent period of accelerated warming. If climate models are used to make scenarios for the next 30–50 years, it is important to know whether they may describe such variations in the climatic trends. It is argued here that low-pass-filtering is not as useful as the polynomial fit approach in terms of trend studies. The regressional approach, allows a simple estimation of confidence intervals, compresses the trend information down to 5 numbers, and facilitates an objective comparison between the observed trends and the trends derived from climate models.

Trend-fits based on regression to polynomials may give unstable results and may be strongly affected by pronounced interannual variability or outliers near the ends of the records. It is therefore important to use this method with caution. It is important to stress that polynomial trends models are inappropriate for extrapolation outside the calibration interval. It is often useful to compare trends derived using different models or with low-pass-filtered records.

The solutions tend to exhibit a 'tightening' near the middle of the interval since the solutions are constructed for the interval $x \in [-1, 1]$ and the spread at $x = 0$ is only affected by the uncertainties in the c_0 (intercept) or the c_1 (linear trend) estimates.

An important observation is that single climate model scenarios do not give a reliable description of the warming, in accord with the conclusions of Benestad (2001a) . A multi-model ensemble, on the other hand, may give a good description of the past climatic trends. If future emissions follow the IS92a scenario from the IPCC and the climate models give an equally reliable description for the future as for the past, then future temperature scenarios derived from these models are highly realistic. It is important to keep in mind, however, that there is no guarantee that the future temperature trends will be within the multi-model ensemble range (Allen et al., 2000). This issue is important for the question of how to construct an ensemble.

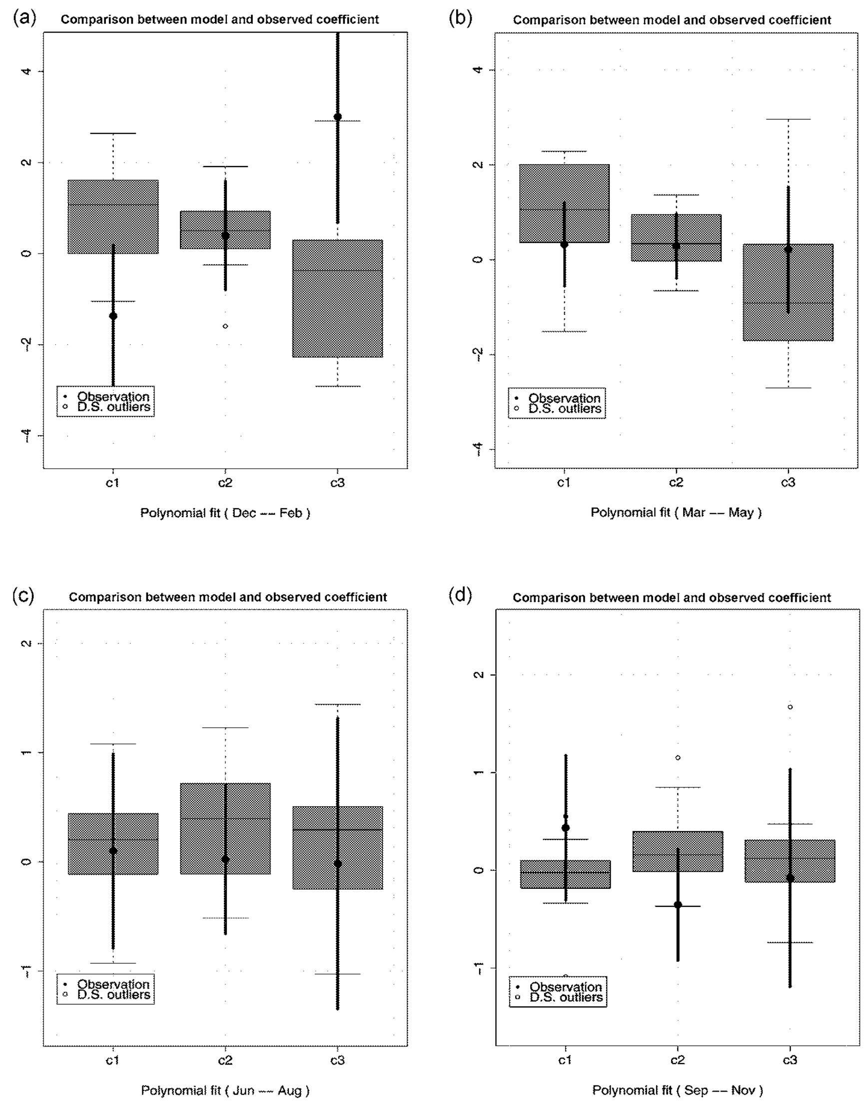

Figure 8. A simple comparison between the linear, quadratic, and cubic terms in the trends derived from the observations and the climate models. These terms are represented by the regression coefficients $c_1, \ldots, c_3$.

On the one hand, an ensemble could consist of extreme climate models which predict an unrealistically wide range (say $[-\infty, \infty]$) of values which is bound to span the future values. Such unrealistic ensembles, however, are not useful. An ensemble may, on the other hand, include only similar models which give a very small spread, but there is greater risk for a very narrow range not to span the future. There is clearly a trade-off between the ‘sharpness’ (usefulness) and the reliability. The spread in the multi-model ensemble may come from different spin-up histories and initialisation (Benestad, 2000), different model formulation, different model resolution and internal variability. It is also important to realize that different choices of downscaling strategy may result in different results (Huth, 2002; Hellström et al., 2001; Benestad, 2001b).

On a regional and local scale, there have been variations in the warming trends with time scales of similar order as the future climate change scenario time horizon (30–50 years) that have only been reproduced by one of the most ‘extreme’ climate models. The rapid winter warming in Scandinavia before 1930 may well be due to internal or natural variability. If the early 20th century warming was a result of internal variability, then the multi-model ensemble size of 14 probably is too small to give a reliable description of the winter warming trends. This warming may also be related to a rapid global warming before 1940–1950, which is difficult to explain in terms of the enhanced greenhouse effect alone (the concentrations were still low then), and it has been speculated about whether changes in the sun may have been responsible for most of this warming (Parker, 2000). If the early 20th century warming was related to solar activity, then the climate models used in this study are not expected to reproduce the rapid warming, since they do not account for changes in the sun. One burning question is then: why would changes in solar activity only affect the winter temperature in Scandinavia? If changes in the sun, volcanism or unresolved oceanic processes contribute to a future climate change, then the climate models may not be able to predict all of the future local warming. These results nevertheless suggest that multi-model ensembles are capable of predicting most of the local climate changes related to an enhanced greenhouse effect. Single-model scenarios (the individual parabolas in Figure 6), on the other hand, are far less reliable.

Acknowledgements

Thanks to Becky Benestad for proofreading and O. Vignes who assisted in retrieving the ECHAM4/OPYC3 GSDIO data. This work was done under the Norwegian Regional Climate Development under Global Warming (RegClim) programme, and was supported by the Norwegian Research Council (Contract NRC-No. 120656/720) and the Norwegian Meteorological Institute. Part of the analysis was carried out using the `R` data processing and analysis language, which is freely available over the Internet (URL http://cran.r-project.org/). I would also like to thank three reviewers who offered critical and useful comments.

Notes

[1] A scaling factor $s = \frac{\max(x)-\min(x)}{\max(\text{year})-\min(\text{year})}$ is required for the estimation of the correct tendency magnitudes.

[2] $s_i = 1.96 \times \sigma_i$, where $\sigma_1, \ldots, \sigma_3$ is the standard error from the ANOVA(Wilks, 1995).

[3] The correlation between symmetric and anti-symmetric functions, such as x and x^2 is close to zero for $x = [-1, \ldots, 1]$, but the correlation between two anti-symmetric functions, such as x and x^3 is close to unity. Hence the errors in coefficients c_1 and c_3 are not independent of each other.

[4] Copies of the R-script can be obtained from rasmus.benestad@met.no.

[5] The term 'regional' will henceforth be used to mean an area covering the range of country- to sub-continental spatial scales, typically with length of 5–20 grid points.

[6] In this case, it is not a requirement that the residual is a white noise process because there may be factors influencing the temperature that are unaccounted for, and have zero trend, and are red noise processes.

References

Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R., and Delworth, T. L.: 2000, 'Quantifying the Uncertainty in Forecasts of Anthropogenic Climate Changes', *Nature* **407**, (October 5), 617–620.

Arrhenius, S.: 1896, 'On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground', *Philosophical Magazine and Journal of Science*, 236–276.

Barnett, T. P.: 1999, 'Comparison of Near-Surface Air Temperature Variability in 11 Coupled Global Climate Models', *J. Climate* **12**, 511–518.

Benestad, R. E.: 2000, 'Future Climate Scenarios for Norway Based on Linear Empirical Downscaling and Inferred Directly from AOGCM Results', *DNMI KLIMA*, 23/00, P. O.Box 43 Blindern, N-0313 Oslo, Norway.

Benestad, R. E.: 2001a, 'The Cause of Warming over Norway in the ECHAM4/OPYC3 *GHG* Integration', *Int. J. Clim.* **21**, 371–387.

Benestad, R. E.: 2001b 'A Comparison between Two Empirical Downscaling Strategies', *Int. J. Clim.* **21**, 1645–1668.

Benestad, R. E.: 2002, 'Empirically Downscaled Temperature Scenarios for Northern Europe Based on a Multi-Model Ensemble', *Clim. Res.*, 21 (2), 105–125.

Cubash, U., Voss, R., Hegerl, G. C., Waszkewitz, J., and Crowley, T. J.: 1997, 'Simulation of the Influence of Solar Radiation Variations on the Global Climate with an Ocean-Atmosphere General Circulation Model', *Clim. Dyn.* **13**, 757–767.

Ellner, Stephen P.: 2001, 'Review of R, Version 1. 1. 1.', *Bulletin of the Ecological Society of America* **82** (2), 127–128.

Flato, G. M., Boer, G. J., Lee, W. G., McFarlane, N. A., Ramsden, D., Reader, M. C., and Weaver, A. J.: 2000, 'The Canadian Centre for Climate Modelling and Analysis Global Coupled Model and its Climate', *Clim. Dyn.*, 16451–16467.

Flury, B.: 1988, *Common Principal Components and Related Multivariate Models*, Wiley, New York.

Gentleman, R. and Ihaka, R.: 2000, 'Lexical Scope and Statistical Computing', *Journal of Computational and Graphical Statistics* **9**, 491–508.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., amd Wood, R. A.: 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.

Gordon, H. B. and O'Farrell, S. P.: 1997, 'Transient Climate Change in the CSIRO Coupled Model with Dynamic Sea Ice', *Mon. Wea. Rev.* **125**, 875–907.

Grotch, S. and MacCracken, M.: 1991, The Use of General Circulation Models to Predict Regional Climate Change', *J. Climate* **4**, 286–303.

Hellström, C., Chen, D., Achberger, C., and Räisanen, J.: 2001, 'Comparison of Climate Change Scenarios for Sweden Based on Statistical and Dynamical Downscaling of Monthly Precipitation', *Clim. Res.* **19**, 45–55.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A.: 2001, *Climate Change 2001: The Scientific Basis*, Contribution of Working Group I to the Third Assessment Report of IPCC. International Panel on Climate Change (available from www.ipcc.ch).

Huth, R.: 2002, 'Statistical Downscaling of Daily Temperature in Central Europe', *J. Climate* **15**, 1731–1742.

Johns, T. C., Carnell, R. E., Crossley, J. F., Gregory, J. M., Mitchell, J. F. B., Senior, C. A., Tett, S. F. B., and Wood, R. A.: 1997, 'The Second Hadley Centre Coupled Ocean-Atmosphere GCM: Model Description, Spin-Up and Validation', *Clim. Dyn.* **13**, 103–134.

Jones, P. D. and Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., and Wigley, T. M. L.: 1998, 'Northern Hemisphere Surface Air Temperature Variations, 1851–1984', *J. Clim. Appl. Meteorol.* **25**, 161–179.

Lorenz, E.: 1967, *The Nature and Theory of the General Circulation of the Atmosphere*, Publication 218, WMO.

Manabe, S. and Stouffer, R. J.: 1996, 'Low-Frequency Variability of Surface Air Temperature in a 1000-Year Integration of a Coupled Atmosphere-Ocean-Land Model, *J. Climate* **9**, 376–393.

Oberhuber, J. M.: 1993, 'Simulation of the Atlantic Circulation with a Coupled Sea Ice-Mixed Layer Isopycnal General Circulation Model. Part 1: Model Description', *J. Phys. Oceanog.* **22**, 808–829.

Palmer, T. N.: 1996, 'Predictability of the Atmosphere and Oceans: From Days to Decades', in Anderson, D. L. T. and Willebrand, J. (eds.), *Decadal Variability*, NATO ASI series, Vol. 44, Springer.

Parker, E. N.: 2000, 'Summary and Perspectives', in *The Solar Cycle and Terrestrial Climate, European Space Agency*, ESA SP-463, 263–269.

Polsky, C., Allard, J., Currit, N., Crane, R., and Yarnal, B.: 2000, 'The Mid-Atlantic Region and its Climate: Past, Present and Future', *Clim. Res.* **14**, 161–173.

Räisänen, J. and Palmer, T. N.: 2001, 'A Probability and Decision-Model Analysis of a Multimodel Ensemble of Climate Change Simulations', *J. Climate* **14**, 3212–3226.

Roeckner, E., Arpe, K., Bengtsson, L., Dümenil, L., Esch, M., Kirk, E., Lunkeit, F., Ponater, M., Rockel, B., Sausen, B., Schlese, U., Schubert, S., and Windelband, M.: 1992, *Simulation of Present-Day Climate with the ECHAM Model: Impact of Model Physics and Resolution*, Tech. Rept., Max Planck-Institute für Meteorologie.

Sengupta, S. and Boyle, J. S.: 1998, 'Using Common Principal Components in Comparing GCM Simulations', *J. Climate* **11**, 816–830.

Tuomenvirta, H., Drebs, A., Førland, E., O. E. Tveitoand, O. E., Alexandersson, H., Laursen, E. V., and Jónsson, T.: 2001, Nordklim data set 1.0 *DNMI KLIMA*, 08/01, P. O.Box 43 Blindern, N-0313 Oslo, Norway.

von Storch, H.: 1999, 'On the Use of "Inflation" in Statistical Downscaling', *J. Climate* **12**, 3505–3506.

Wilks, D. S.: 1995, *Statistical Methods in the Atmospheric Sciences*, Academic Press, Orlando, FL.

(Received 28 January 2002; in revised form 1 November 2002)