# Attributing variation in a regional climate change modelling experiment.

Christopher A. T. Ferro*

April 28, 2004

*Keywords*: analysis of variance, climate change, regional modelling experiment

## 1 Introduction

This note demonstrates a simple *analysis of variance* for quantifying the uncertainty arising from different components of the PRUDENCE experiments. This systematic exploration of experiment results helps to identify the relative importance of the different components, to construct simplifying syntheses of simulation output, and to make inferences about climate change and model differences. The presented analysis quantifies the effects of different forcings, global models (GCMs), regional models (RCMs) and combinations of these factors. The methods used are introduced in Chapter 9 of von Storch and Zwiers (2001), who also cite more detailed references.

The example considers annual mean two-metre air temperatures from four climate-change experiments: the common control (CTL) and A2 scenario (SA2) runs of HIRHAM (DMI) and RCAO (SMHI) driven by both HadAM3H (HAD) and ECHAM4 (ECH). The annual means are computed from the monthly means interpolated onto the CRU grid. The methods are first illustrated with the temperatures averaged over land points and then applied separately at each grid point to reveal the spatial structure.

## 2 Analysis of variance

### 2.1 Basic model

Let $Y_{ijkl}$ be the annual mean two-metre air temperature averaged over land points for scenario $i$, GCM $j$, RCM $k$ and year $l$. Each of the first three *factors* has two *levels*:

*Address for correspondence: Department of Meteorology, University of Reading, Earley Gate, P.O. Box 243, Reading, RG6 6BB, UK. E-mail: c.a.t.ferro@reading.ac.uk

$i = 1$ (CTL) and 2 (SA2); $j = 1$ (HAD) and 2 (ECH); $k = 1$ (DMI) and 2 (SMHI). There are $l = 1, \ldots, 30$ years. Boxplots of the data from the eight integrations are displayed in Figure 1.

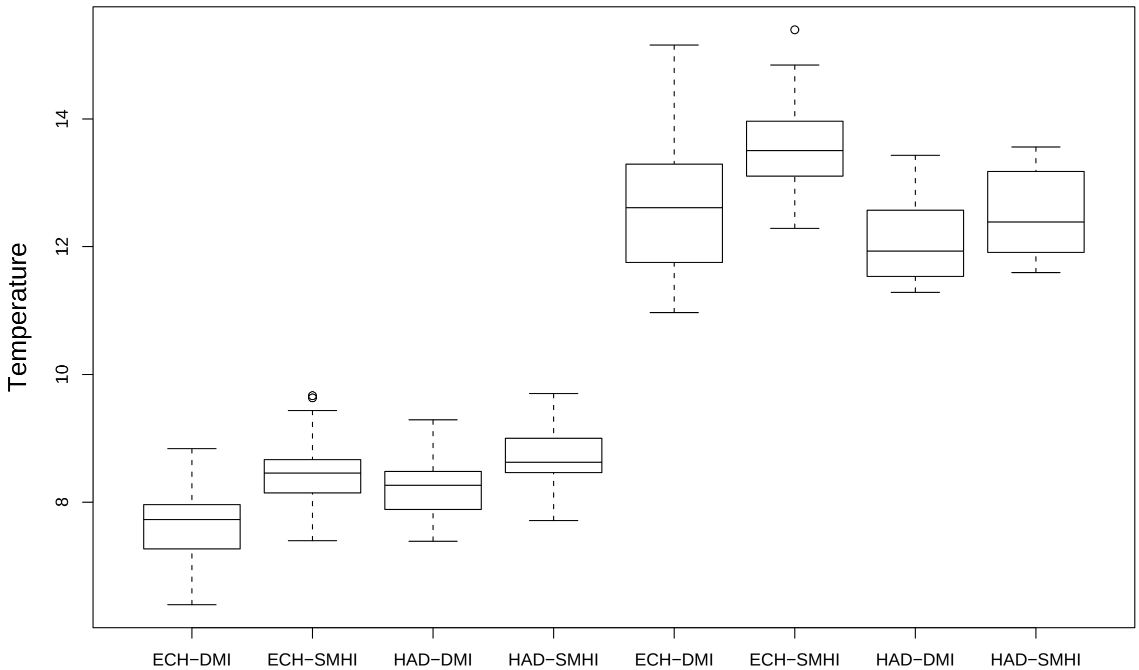


Figure 1: Boxplots of annual mean two-metre air temperatures (°C) averaged over land points for each of eight, 30-year integrations. The first four are the control integrations, the second four are the A2 scenario integrations. Labels indicate the GCM-RCM combination. Each box represents sample quartiles, whiskers extend to the most extreme data points that are outside the box by at most 1.5 times the interquartile range, and other points are plotted as circles.

Suppose that the temperatures can be decomposed according to the model

$$Y_{ijkl} = M + S_i + G_j + R_k + (SG)_{ij} + (SR)_{ik} + (GR)_{jk} + (SGR)_{ijk} + Z_{ijkl}. \qquad (1)$$

Term $M$ is the *overall effect* averaged over all years, scenarios, GCMs and RCMs. Term $S_i$ is the *main effect* (relative to $M$ and averaged over all years, GCMs and RCMs) of scenario $i$; $G_j$ and $R_k$ are defined similarly. Term $(SG)_{ij}$ is the *interaction effect* (relative to $M+S_i+G_j$ and averaged over all years and RCMs) of combining scenario $i$ and GCM $j$; $(SR)_{ik}$ and $(GR)_{jk}$ are defined similarly. Term $(SGR)_{ijk}$ is the *3rd-order interaction effect* (relative to the sum of all previous effects and averaged over all years) of combining scenario $i$, GCM $j$ and RCM $k$. The $Z_{ijkl}$ are independent and identically distributed $N(0,\sigma^2)$ variables representing inter-annual variation. This basic model assumes that within each integration the output $(Y_{ijk1},\ldots,Y_{ijk30})$ is a sequence of independent and identically distributed Normal random variables, the variances of which are the same for all integrations. Extensions of this model are proposed in Section 3.

## 2.2 Parameter estimation

Estimators for the terms in model (1) are given in Table 1, where dot $(\cdot)$ subscripts denote mean with respect to the corresponding indices, for example $Y_{ijk.} = \sum_{l=1}^{30} Y_{ijkl}/30$.

| Effect | Estimator |
|---|---|
| $M$ | $Y_{....}$ |
| $S_i$ | $Y_{i...} - Y_{....}$ |
| $G_j$ | $Y_{.j..} - Y_{....}$ |
| $R_k$ | $Y_{..k.} - Y_{....}$ |
| $(SG)_{ij}$ | $Y_{ij..} - Y_{i...} - Y_{.j..} + Y_{....}$ |
| $(SR)_{ik}$ | $Y_{i.k.} - Y_{i...} - Y_{..k.} + Y_{....}$ |
| $(GR)_{jk}$ | $Y_{.jk.} - Y_{.j..} - Y_{..k.} + Y_{....}$ |
| $(SGR)_{ijk}$ | $Y_{ijk.} - Y_{ij..} - Y_{i.k.} - Y_{.jk.} + Y_{i...} + Y_{.j..} + Y_{..k.} - Y_{....}$ |

Table 1: Estimators for the terms in model (1).

The variances of the estimators in Table 1 follow from the identity

$$\mathrm{var}\left(\sum_i c_i X_i\right) = \sum_i c_i^2 \mathrm{var}(X_i) + \sum_{i\neq j} c_i c_j \mathrm{cov}(X_i, X_j)$$

for constants $c_i$ and random variables $X_i$, and the assumption that the $Y_{ijkl}$ are independent, each with variance $\sigma^2$. For example, the variance of the estimator for $M$ is $\sigma^2/(IJKL)$, where $I=2$, $J=2$, $K=2$ and $L=30$ are the number of levels of the different factors. Estimated variances are obtained by replacing the residual variance $\sigma^2$ with its estimator

$$\hat{\sigma}^2 = \frac{1}{IJK(L-1)} \sum_{ijkl} (Y_{ijkl} - Y_{ijk.})^2.$$

The estimators in Table 1 can be evaluated for the PRUDENCE data using the information in Table 2. The results are displayed in Table 3. By construction, summing any estimator over a single dimension yields zero ($S_1 + S_2 = 0$ for example) so, since the factors each have only two levels, it is sufficient to quote the effects estimated when all subscripts are equal to 1. The residual variance is $\hat{\sigma}^2 = 0.42$.

| | HAD | | | ECH | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | DMI | SMHI | Mean | DMI | SMHI | Mean | DMI | SMHI | Mean |
| CTL | 8.20 | 8.70 | 8.45 | 7.64 | 8.53 | 8.08 | 7.92 | 8.61 | 8.27 |
| SA2 | 12.08 | 12.53 | 12.31 | 12.59 | 13.53 | 13.06 | 12.34 | 13.03 | 12.68 |
| Mean | 10.14 | 10.62 | 10.38 | 10.12 | 11.03 | 10.57 | 10.13 | 10.82 | 10.48 |

Table 2: Means of the annual mean two-metre air temperatures (°C) averaged over land points split by scenario, GCM and RCM. The standard errors for means across 0, 1, 2 and 3 factors are 0.12, 0.083, 0.059 and 0.042°C.

| $M$ | $S_1$ | $G_1$ | $R_1$ | $(SG)_{11}$ | $(SR)_{11}$ | $(GR)_{11}$ | $(SGR)_{111}$ |
|---|---|---|---|---|---|---|---|
| 10.48 | −2.21 | −0.10 | −0.35 | 0.28 | −0.00 | 0.11 | −0.01 |

Table 3: Estimated effects (°C), each with standard error 0.042°C.

## 2.3 Variance decomposition

The *total sum of squares* about the mean, which is a scaled version of the sample variance, can be decomposed into contributions from the different terms in model (1):

$$\begin{aligned}\sum_{ijkl}(Y_{ijkl} - Y_{....})^2 &= \sum_{ijkl} S_i^2 + \sum_{ijkl} G_j^2 + \sum_{ijkl} R_k^2 + \sum_{ijkl}(SG)_{ij}^2 + \sum_{ijkl}(SR)_{ik}^2 \\ &\quad + \sum_{ijkl}(GR)_{jk}^2 + \sum_{ijkl}(SGR)_{ijk}^2 + \sum_{ijkl}(Y_{ijkl} - Y_{ijk.})^2. \qquad (2)\end{aligned}$$

The final term in the decomposition is called the *residual sum of squares*. The contributions are listed in Table 4, often called an 'analysis of variance' table, and expressed as a proportion of the total sum of squares. The model explains 92.7% of the variation in the data, with 7.3% attributed to residual (inter-annual) variation. The interpretation of the other components is deferred to Section 2.5.

| | SS | % | DF | MS | $F$ | $p$ |
|---|---|---|---|---|---|---|
| $S$ | 1170.21 | 88.7 | 1 | 1170.21 | 2817.81 | 0.000 |
| $G$ | 2.21 | 0.2 | 1 | 2.21 | 5.33 | 0.022 |
| $R$ | 28.81 | 2.2 | 1 | 28.81 | 69.38 | 0.000 |
| $SG$ | 18.73 | 1.4 | 1 | 18.73 | 45.10 | 0.000 |
| $SR$ | 0.00 | 0.0 | 1 | 0.00 | 0.00 | 0.994 |
| $GR$ | 2.84 | 0.2 | 1 | 2.84 | 6.83 | 0.010 |
| $SGR$ | 0.03 | 0.0 | 1 | 0.03 | 0.08 | 0.779 |
| Residual | 96.35 | 7.3 | 232 | 0.42 | | |
| Total | 1319.19 | 100.0 | | | | |

Table 4: Analysis of variance table.

## 2.4 Model simplification

The statistical significance of terms in model (1) can be assessed with suitable tests. If terms are found to be insignificant then they can be removed, yielding a more parsimonious model, a simpler interpretation, and more precise estimates of quantities of interest.

The number of independent terms in the model associated with each of the sums of squares (SS) in Table 4 is known as the *degrees of freedom* (DF). The *mean squares* (MS) are the ratios SS/DF. The residual MS is the estimated residual variance $\hat{\sigma}^2$.

Under the hypothesis that all levels of a particular factor are equal ($S_1 = S_2$ for example) and by construction therefore zero, the ratio of the corresponding MS (with $\tau$ degrees of freedom) to $\hat{\sigma}^2$ (with $\nu$ degrees of freedom) has an $F_{\tau,\nu}$ distribution. The $p$-value for the hypothesis test is the mass of the $F_{\tau,\nu}$ distribution lying above the observed value of the ratio. These test statistics and $p$-values are also shown in Table 4. The data from the eight integrations are *balanced*, which means that there is an equal number ($L = 30$) of *replications* for each of the eight possible combinations of levels of the three factors. If this were not the case then Table 4 would need modifying before conducting the hypothesis tests.

Both the terms $SGR$ and $SR$ have large $p$-values. Ignoring these terms leaves a model in which the expectations of climate-change responses, $Y_{2jk.} - Y_{1jk.}$, are independent of the RCM, $k$. Interaction remains, however, between scenario and GCM, and between RCM and GCM. Consequently, the estimated effects in Table 3 must be interpreted carefully. For example, the GCM term $G$ has a small $p$-value and encourages the conclusion that the GCM effects are different from each other, with ECHAM4 warmer than HadAM3H by about $G_2 - G_1 = 0.10 - (-0.10) = 0.20$°C. This conclusion is incorrect, however, as will be shown below. The confusion arises because the GCM effects $G_1$ and $G_2$ average over all possible scenario and RCM combinations, whereas the presence of $SG$ and $GR$ interactions necessitates the inspection of the GCM effects for each combination separately.

## 2.5 Contrasts

The model can be used to make inferences about linear combinations of terms in the model that may be more interesting than the original effects. For notational convenience, denote the expectation of $Y_{ijk.}$ by $\mu_{ijk}$. A *contrast* is a linear combination $\sum_{ijk} c_{ijk}\mu_{ijk}$ in which the coefficients sum to zero: $\sum_{ijk} c_{ijk} = 0$. Two contrasts, $\sum_{ijk} c_{ijk}\mu_{ijk}$ and $\sum_{ijk} c'_{ijk}\mu_{ijk}$ are *orthogonal* if $\sum_{ijk} c_{ijk}c'_{ijk} = 0$. The greatest possible number of mutually orthogonal contrasts is $IJK - 1$.

The estimator for a contrast is $c = \sum_{ijk} c_{ijk}Y_{ijk.}$ with estimated variance $s^2 = \hat{\sigma}^2 \sum_{ijk} c_{ijk}^2/L$. A $100(1-\alpha)\%$-confidence interval is $c \pm t_\nu(1-\alpha/2)s$, where $t_\nu(\alpha)$ is the $\alpha$-quantile of the $t$ distribution with $\nu = IJK(L-1)$ degrees of freedom. The sum of squares associated with a contrast is $\mathrm{SS} = L(\sum_{ijk} c_{ijk}Y_{ijk.})^2/\sum_{ijk} c_{ijk}^2$ with one degree of freedom, and the statistical significance can be quantified by comparing $\mathrm{MS}/\hat{\sigma}^2$ with the $F_{1,\nu}$ distribution. The sums of squares from a set of $IJK-1$ orthogonal contrasts total the *model sum of squares* $\sum_{ijkl}(Y_{ijk.} - Y_{....})^2$, which is the difference between the total and residual sums of squares.

When there are only two levels of each factor, the sums of squares in decomposition (2) correspond to sums of squares for a set of easily interpretable orthogonal contrasts. For example, the contrast $S_2 - S_1 = \mu_{2..} - \mu_{1..}$ has coefficients $c_{ijk} = (-JK)^{-i}$ and sum of squares

$$\frac{L(S_2 - S_1)^2}{I/(JK)} = JKL\sum_i S_i^2 = \sum_{ijkl} S_i^2$$

since $S_2 = -S_1$. The seven orthogonal contrasts are described in Table 5. The first is the mean climate-change response. The others are the differences between the

- GCM main effects,
- RCM main effects,
- mean climate-change responses of the GCMs,
- mean climate-change responses of the RCMs,
- mean effects of the GCMs on the RCMs,
- mean effects of the GCMs on the climate-change responses of the RCMs.

These contrasts help to interpret the information in Table 4. The climate-change response explains 88.7% of the variation in the data and this difference between the scenario main effects is highly statistically significant ($p = 0.000$). Inter-annual variability

| Term | Contrast |
|---|---|
| $S$ | $\mu_{2..} - \mu_{1..}$ |
| $G$ | $\mu_{.2.} - \mu_{.1.}$ |
| $R$ | $\mu_{..2} - \mu_{..1}$ |
| $SG$ | $\Delta\mu_{2.} - \Delta\mu_{1.}$ |
| $SR$ | $\Delta\mu_{.2} - \Delta\mu_{.1}$ |
| $GR$ | $(\mu_{.22} - \mu_{.12}) - (\mu_{.21} - \mu_{.11})$ |
| $SGR$ | $(\Delta\mu_{22} - \Delta\mu_{12}) - (\Delta\mu_{21} - \Delta\mu_{11})$ |

Table 5: The set of orthogonal contrasts corresponding to decomposition (2). Define $\Delta\mu_{jk} = \mu_{2jk} - \mu_{1jk}$ to be the mean climate-change response for GCM $j$ and RCM $k$.

explains the next largest proportion (7.3%) of the variation. Although the differences between the GCM main effects and between the RCM main effects explain only 0.2% and 2.2% of the variation, and so might be considered practically insignificant, they are both statistically significant at the 5% level. The amount of variation accounted for by the difference between the mean climate-change responses of the GCMs (1.4%) is greater than that accounted for by the difference between the mean climate-change responses of the RCMs (0.0%), the latter being statistically insignificant ($p = 0.994$). The difference between the effects of the GCMs on the RCMs accounts for just 0.2% of the variation but is statistically significant at the 1% level. In contrast, the difference between the effects of the GCMs on the climate-change responses of the RCMs is statistically insignificant ($p = 0.779$).

More detailed constrasts can now be considered. The similarity of the mean climate-change responses of the RCMs noted earlier suggests that the following (non-orthogonal) contrasts are sufficient to quantify all of the statistically significant differences.

1. The difference between scenarios (climate-change response) for each GCM, averaging over RCM:
   (a) HAD ($\mu_{21.} - \mu_{11.}$)
   (b) ECH ($\mu_{22.} - \mu_{12.}$)
2. The difference between RCMs for each GCM, averaging over scenario:
   (a) HAD ($\mu_{.12} - \mu_{.11}$)
   (b) ECH ($\mu_{.22} - \mu_{.21}$)
3. The difference between GCMs for each scenario-RCM combination:

(a) CTL-DMI ($\mu_{121} - \mu_{111}$)

(b) CTL-SMHI ($\mu_{122} - \mu_{112}$)

(c) SA2-DMI ($\mu_{221} - \mu_{211}$)

(d) SA2-SMHI ($\mu_{222} - \mu_{212}$)

Confidence intervals and $p$-values for these contrasts are given in Table 6. The following conclusions can be drawn. The A2 scenario is warmer than the control scenario for both GCMs; RCAO is warmer than HIRHAM for both GCMs; ECHAM4 is warmer than HadAM3H for both RCMs in the A2 scenario, but in the control scenario HadAM3H is warmer than ECHAM4 for HIRHAM, and there is little difference for RCAO. These relationships are evident in Figure 1.

| | Contrast | Estimate | CI | SS | $F$ | $p$ |
|---|---|---|---|---|---|---|
| 1a | $\mu_{21.} - \mu_{11.}$ | 3.86 | $(3.63, 4.09)$ | 446.42 | 1074.97 | 0.000 |
| 1b | $\mu_{22.} - \mu_{12.}$ | 4.97 | $(4.74, 5.21)$ | 742.52 | 1787.95 | 0.000 |
| 2a | $\mu_{.12} - \mu_{.11}$ | 0.72 | $(0.48, 0.95)$ | 15.39 | 37.07 | 0.000 |
| 2b | $\mu_{.22} - \mu_{.21}$ | 0.91 | $(0.68, 1.14)$ | 24.87 | 59.88 | 0.000 |
| 3a | $\mu_{121} - \mu_{111}$ | $-0.56$ | $(-0.89, -0.23)$ | 4.72 | 11.36 | 0.001 |
| 3b | $\mu_{122} - \mu_{112}$ | $-0.17$ | $(-0.50, 0.16)$ | 0.45 | 1.08 | 0.301 |
| 3c | $\mu_{221} - \mu_{211}$ | 0.51 | $(0.18, 0.84)$ | 3.90 | 9.40 | 0.002 |
| 3d | $\mu_{222} - \mu_{212}$ | 0.99 | $(0.66, 1.32)$ | 14.75 | 35.51 | 0.000 |

Table 6: Estimated contrasts, 95% confidence intervals (CI) and $p$-values.

The contrasts in Table 6 were considered only after, and indeed motivated by, the initial analysis of the data. When such a set of 'unplanned' contrasts is tested, the significance levels of the tests or, equivalently, the coverages of the confidence intervals should be adjusted to account for the likelihood of obtaining an erroneously significant difference by chance. This *multiple comparisons* problem has not been addressed here, but the literature describes many possible adjustments. See Milliken and Johnson (1984) for example.

## 2.6 Grid-point analysis

Model (1) was estimated at each point on the CRU grid. Figure 2 shows the proportion of variation explained by the different components of the model. These plots provide a summary of the sources of variation in the modelling experiment. The following conclusions can be drawn.

- The proportion of variance explained by the model is at least 75% everywhere, is greater over water and increases from north-east to south-west. The inter-annual variation is around 10% over land and around 5% over water.

- The proportion of variation explained by the differences in scenario main effects increases from 30% in the north-west to near 100% in the south-east, with lower values in mountainous regions and in the Baltic Sea.

- Most of the remaining variation in the Atlantic is explained by the differences between the GCM main effects, with smaller contributions from the differences in GCM climate-change responses and in RCM main effects.

- Almost all of the remaining variation in the Baltic Sea is explained by the GCM main effect.

- Some of the remaining variation over land, the Mediterranean and North Seas is explained by the GCM response, but in mountainous regions, Scandinavia and around the Adriatic Sea a greater part is explained by the RCM main effect.

Figure 3 identifies those interaction effects in the model that are significant at the 5% level. The following conclusions can be drawn.

- The third-order interaction is significant only in the North Atlantic (grey).

- All three second-order terms are significant in central Europe (white).

- In the rest of Europe, the scenario-RCM interaction is significant only in the far north (green).

- The GCM-RCM interaction is insignificant over the rest of Northern Europe, the Alps and the Mediterranean Sea (yellow).

- The scenario-GCM interaction is insignificant over the Baltic Sea and Turkey (red).

Estimated contrasts comparing the two levels of each factor for all combinations of levels of the other factors are shown in Figures 4–6. The following conclusions can be drawn.

- The A2 scenario is everywhere warmer than the control for all GCM-RCM combinations, and the difference typically increases from north-west to south-east and with distance from the coast.

- ECHAM4 is cooler than HadAM3H around the Baltic Sea for all scenario-RCM combinations; in the control, ECHAM4 is also cooler in central and eastern Europe for HIRHAM; in the scenario, ECHAM4 is warmer in Scandinavia, western, central and southern Europe.

- RCAO is cooler than HIRHAM over water and warmer over land, particularly in mountainous regions, for all scenario-GCM combinations.

Estimated contrasts comparing the climate-change responses (SA2 − CTL) of the two GCMs for each RCM and of the two RCMs for each GCM are shown in Figure 7. The following conclusions can be drawn.

- The ECHAM4 response is generally greater than the HadAM3H response, particularly in the west and over land.

- The reponse of HIRHAM is everywhere similar to that of RCAO for HadAM3H, but for ECHAM4 is greater in the north, and lower north of the Alps and west of the Pyrenees.

The final bullet point highlights another potential benefit of performing an analysis of variance. If the main feature of interest were the mean climate-change response with HadAM3H boundary conditions then the analysis suggests that it would be sufficient to plot a map of the response averaged across both RCMs rather than separate maps for RCAO and HIRHAM. Note, however, that this conclusion might change if a more careful analysis, using techniques described in the following section, were conducted.

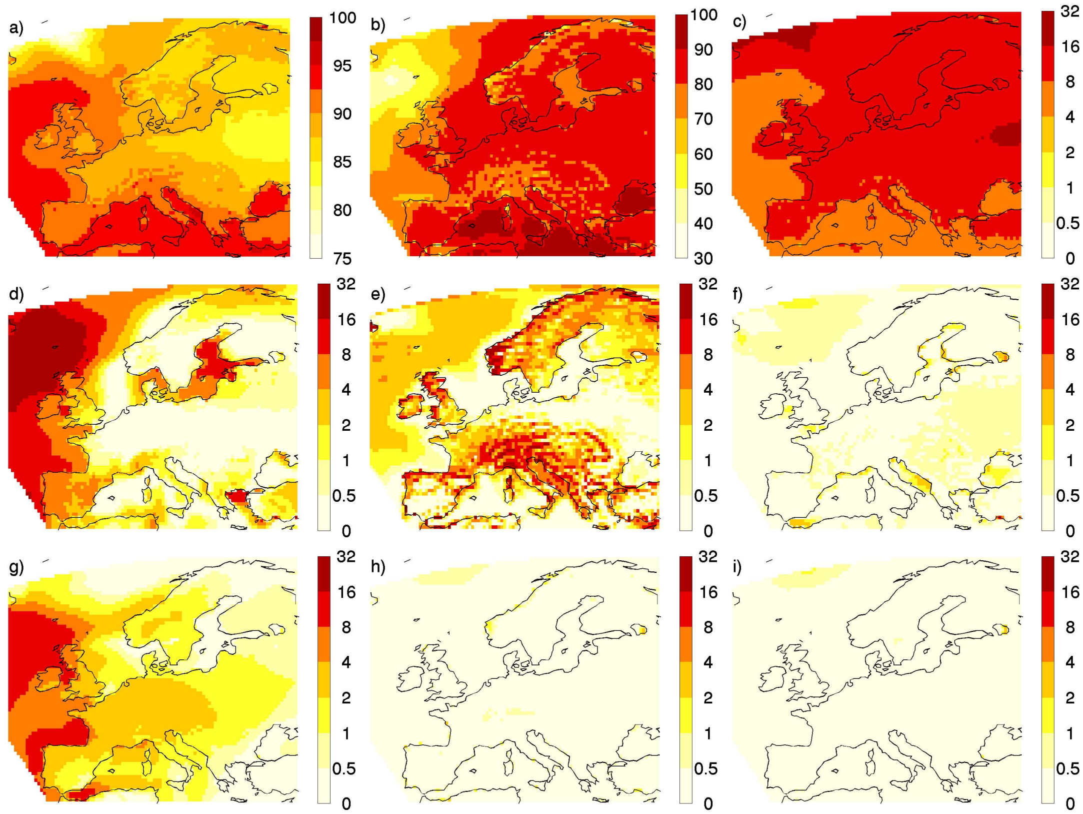


Figure 2: Percentage of variation explained by a) the full model, b) the differences in the scenarios ($S$), c) the inter-annual variation, and by the differences in d) the GCMs ($G$), e) the RCMs ($R$), f) the GCM-RCM interaction ($GR$), g) the GCM responses ($SG$), h) the RCM responses ($SR$), and i) the GCM-RCM response interaction ($SGR$).

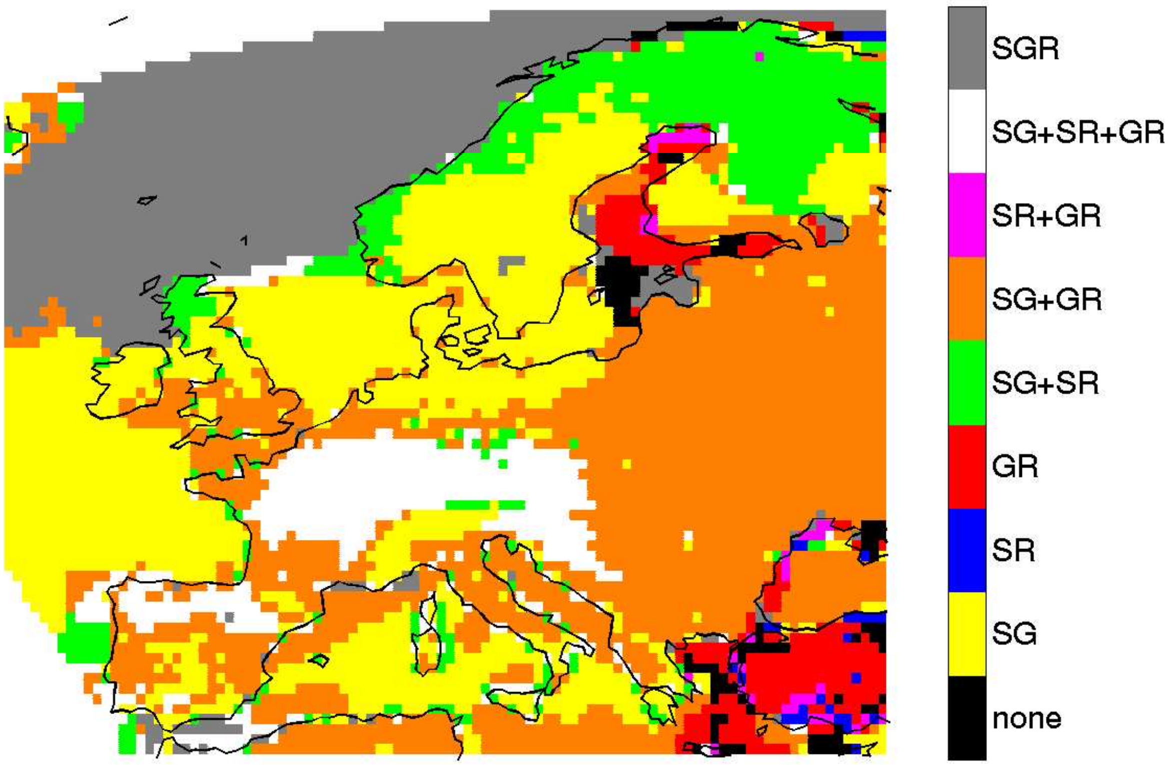


Figure 3: The interactions that are significant at the 5% level. Lower order terms are considered significant if corresponding higher order terms are significant.

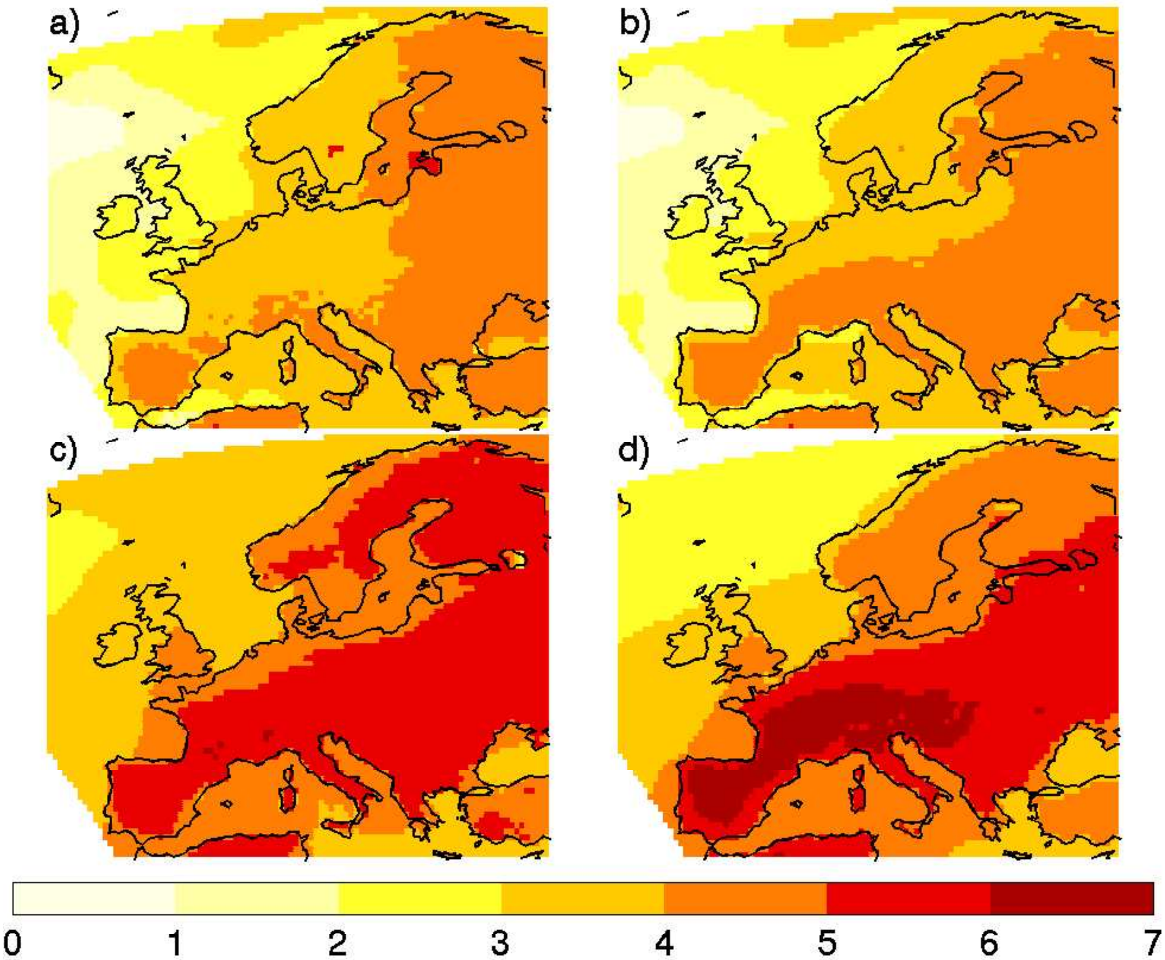


Figure 4: Contrasts (°C) between scenarios (SA2 − CTL) for four GCM-RCM combinations: a) HAD-DMI, b) HAD-SMHI, c) ECH-DMI and d) ECH-SMHI. Grid points at which the contrast is insignificant at the 1% level are masked.

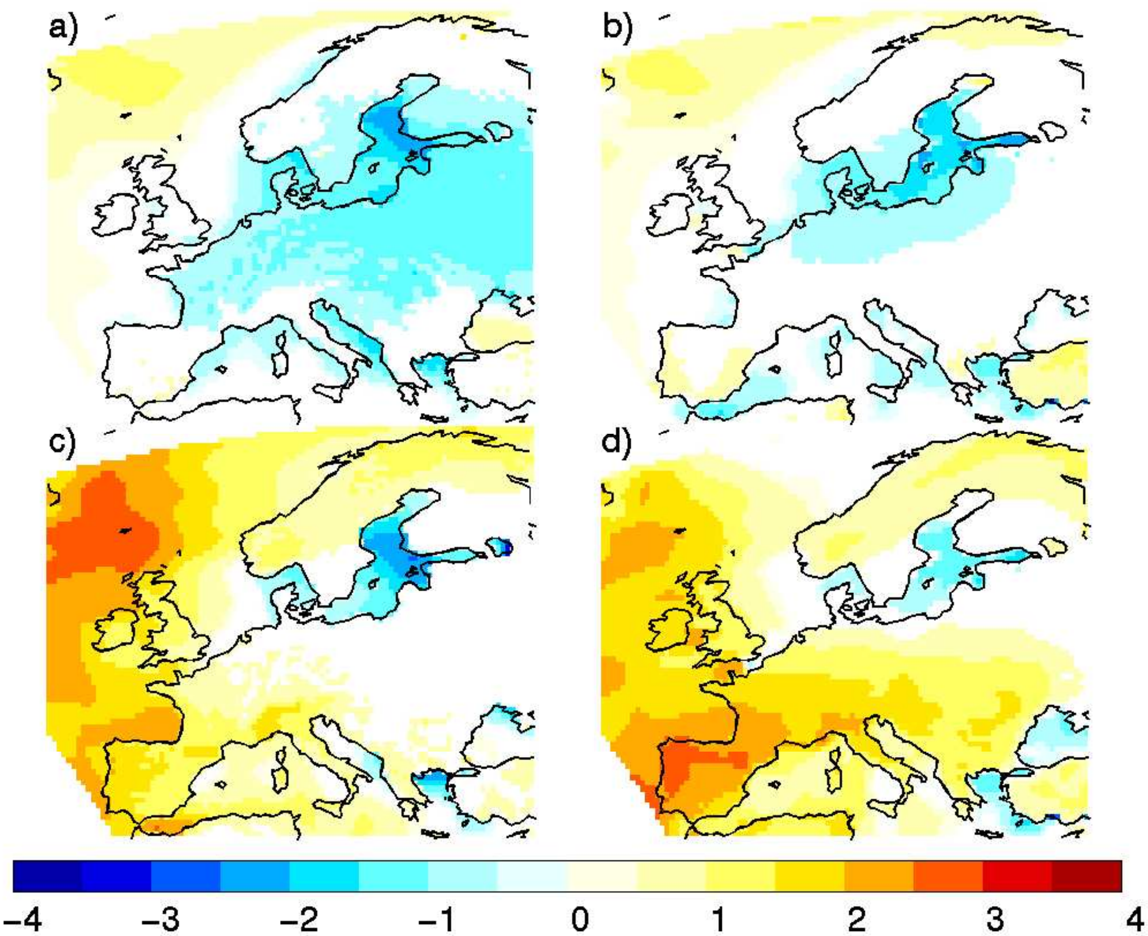


Figure 5: Contrasts (°C) between GCMs (ECH − HAD) for four scenario-RCM combinations: a) CTL-DMI, b) CTL-SMHI, c) SA2-DMI and d) SA2-SMHI. Grid points at which the contrast is insignificant at the 1% level are masked.

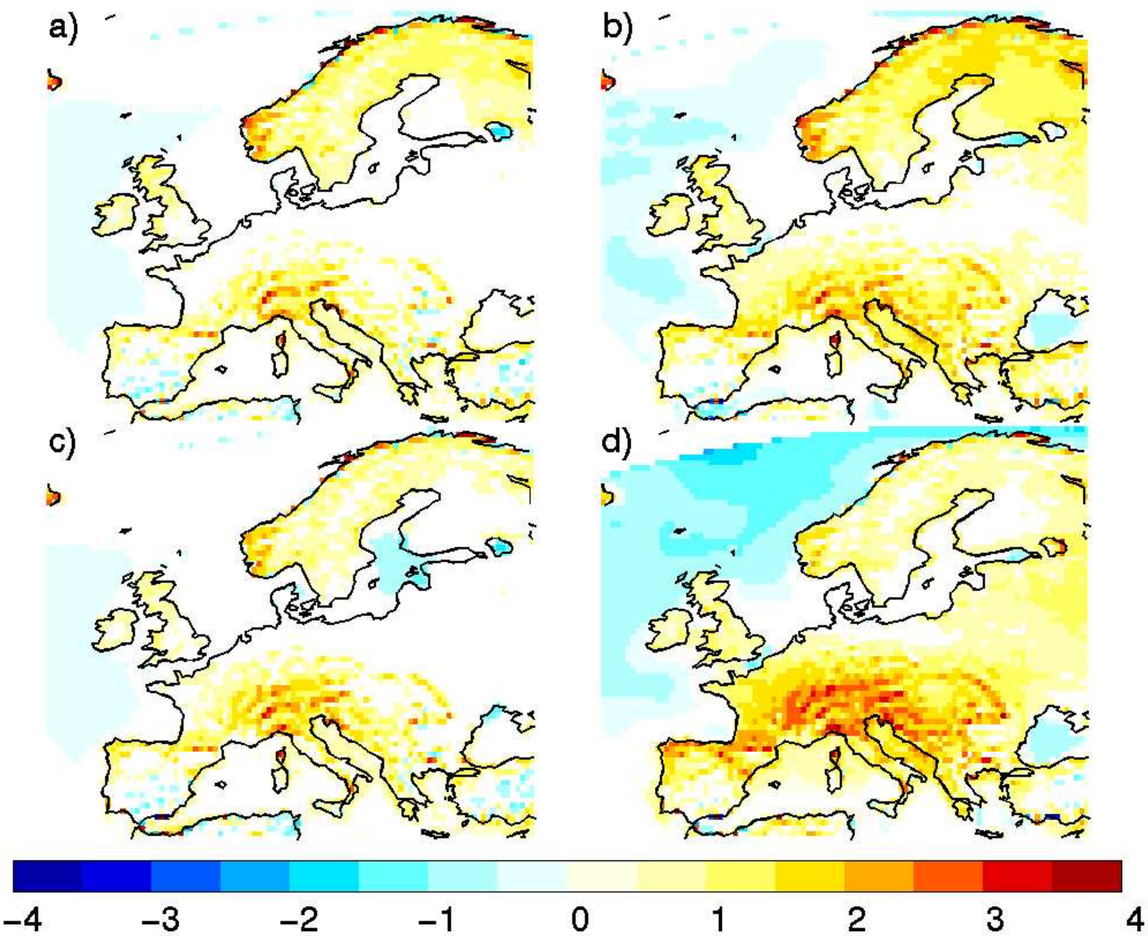


Figure 6: Contrasts (°C) between RCMs (SMHI − DMI) for four scenario-GCM combinations: a) CTL-HAD, b) CTL-ECH, c) SA2-HAD and d) SA2-ECH. Grid points at which the contrast is insignificant at the 1% level are masked.

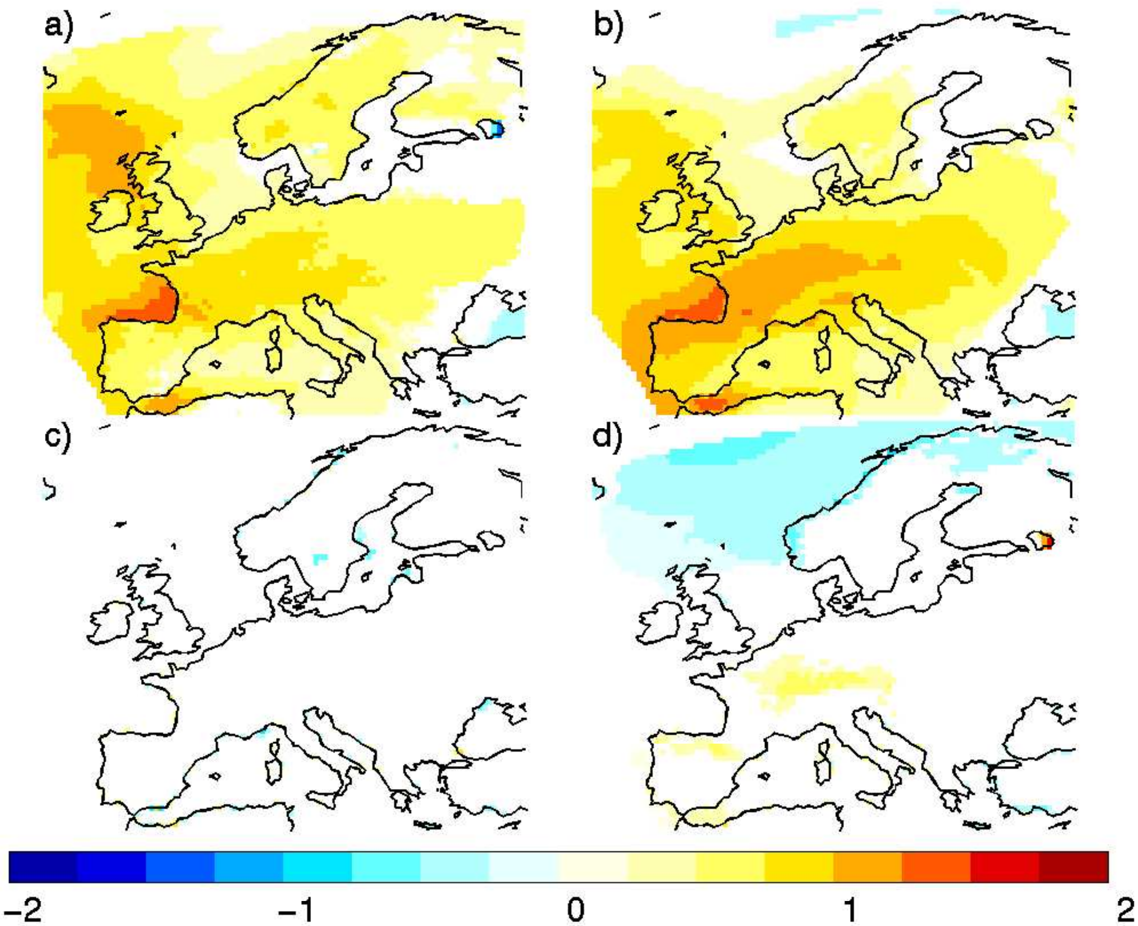


Figure 7: Response (SA2 − CTL) contrasts (°C) between GCMs (ECH − HAD) for a) DMI and b) SMHI, and between RCMs (SMHI − DMI) for c) HAD and d) ECH. Grid points at which the contrast is insignificant at the 1% level are masked.

# 3 Extending the model

This section proposes some extensions to model (1) that represent the data more accurately and support more detailed inferences. The deficiencies of the basic model for the land-averaged data are illustrated with diagnostic plots.

## 3.1 Diagnostics

The quantile-quantile plots in Figure 8 indicate that the assumption of normally distributed residuals is poor: the residuals have lighter tails in the control runs, and a relatively flat distribution in the scenario runs. Some outliers are also evident. The assumption of common variance is also unacceptable: the boxplots in Figure 9 indicate that there is greater variation in the scenario runs. The model fit is assessed in Figure 10 and is inadequate: trends in the scenario runs are overlooked, which is reflected in positive serial correlation (not shown) in the corresponding residuals.

The model assumptions are untenable. Features of the data are not captured by the model and previous inferences may be inaccurate. More elaborate models and inference procedures that avoid these shortcomings are discussed in the following sections.

## 3.2 Model structure

- Trends could be accounted for by including covariates in the model:

  $$Y_{ijkl} = M + \ldots + (SGR)_{ijk} + \beta_{ijk} x_{ijkl} + Z_{ijkl}$$

  for example, where $x_{ijkl}$ could be a measure of time, the HadCM3 boundary conditions, or the scenario emissions forcings. Another approach (Sexton *et al.*, 2003) that includes an extra factor with one level for each time period could also be used, but requires additional observations if the full model is to be estimated.

- More levels of the scenario, GCM and RCM factors could be added, although the data are then likely to be unbalanced and inference becomes more complicated. Different ensemble members could also be added, and considered either as independent replicates or as another factor representing intra-ensemble variation.

- *Repeated measures* models (for example Milliken and Johnson, 1984) that explicitly recognise the time ordering of the data can also be employed.

## 3.3 Error distribution

- Inference becomes complicated if the residuals are not normally distributed. However, it might be hoped that a satisfactory model would achieve approximately normal residuals.

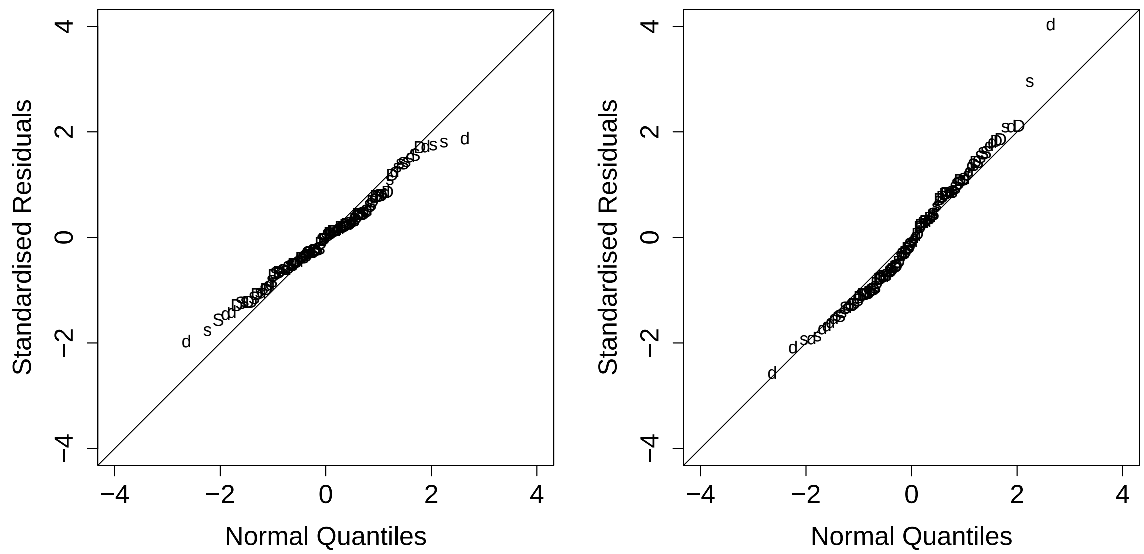


Figure 8: Plots of standardised residuals against standard normal quantiles for the control (left) and scenario (right) integrations. Labels indicate the RCM-GCM combination: DMI-HAD (D), DMI-ECH (d), SMHI-HAD (S) and SMHI-ECH (s).

- Statistical tests for unequal variances are available. If equality of variances is rejected then inference procedures can be adapted. Results are typically robust to unequal variances, however, as long as the data are approximately balanced (Milliken and Johnson, 1984, page 17).

- Even after modelling trends, the residuals might be serially dependent. Modelling this dependence, replacing $Z_{ijkl}$ with an auto-regressive process for example, leads to a general linear (time-series) model, for which inference is more complicated.

## 3.4 Random effects

- All of the models proposed thus far have been *fixed effect* models, that is the terms in the model are assumed to be unknown constants. This is natural if interest

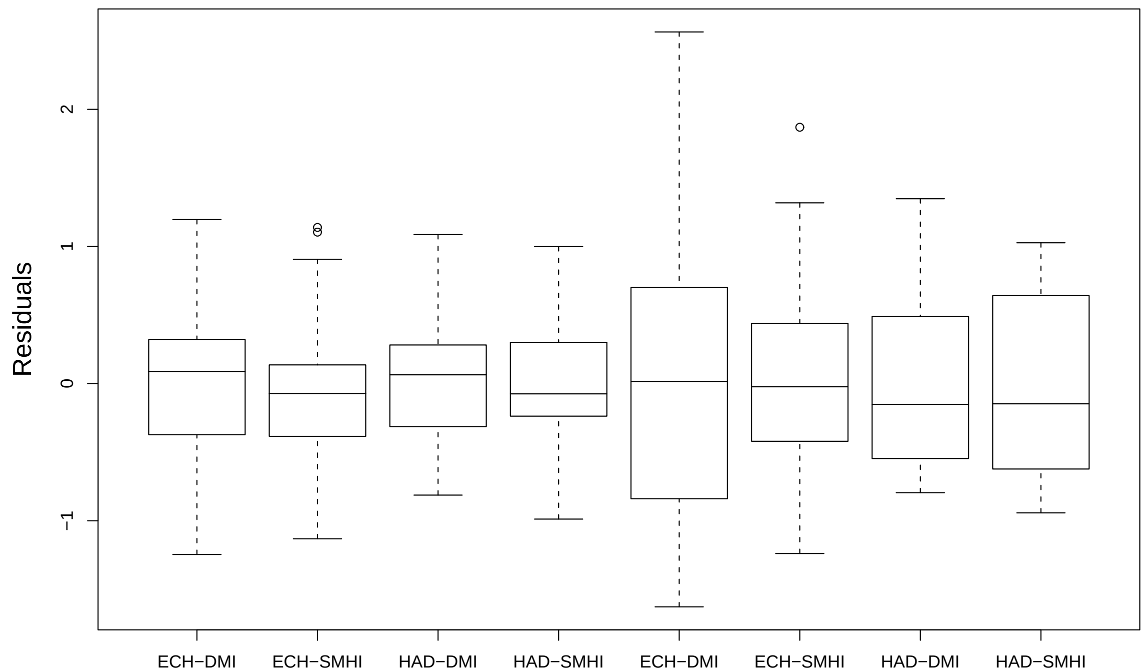


Figure 9: Boxplots of residuals (°C) for each of eight, 30-year integrations. The first four are the control integrations; the second four are A2 scenario integrations. Labels indicate the GCM-RCM combination.

is in those particular levels of the different factors: those particular RCMs for example. For statements about climate change, however, it may be more natural to treat the GCM and RCM effects as random, to assume that they are a sample of all possible GCMs and RCMs, and to make inferences about these GCM and RCM populations and their effects. Models in which all terms are random are known as *random effect* models; models with both fixed and random effects are known as *mixed effect* models. See von Storch and Zwiers (2001) for details.

## 3.5 Multivariate models

- Instead of using only two-metre air temperatures for the response variable $Y$, multivariate techniques could be used to model several variables simultaneously.

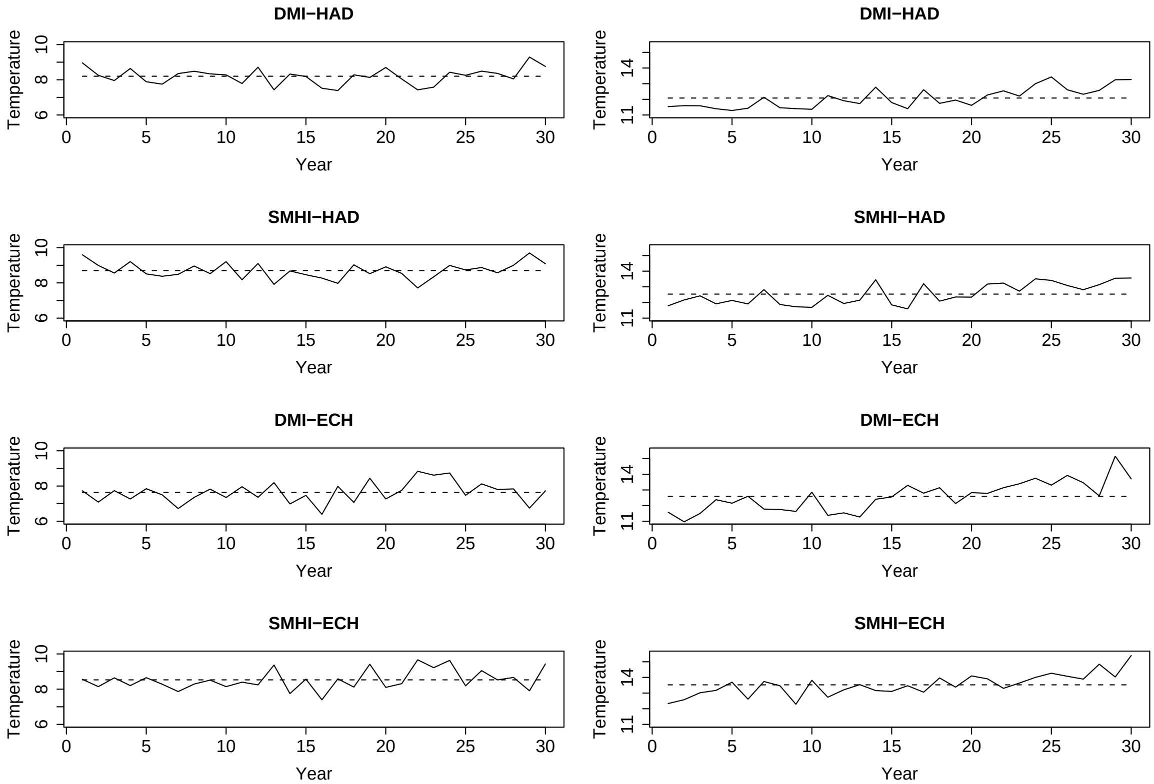


Figure 10: Annual mean two-metre air temperatures (——) in °C averaged over land points for each of eight, 30-year integrations. The control integrations are on the left; the A2 scenario integrations are on the right. Titles indicate the RCM-GCM combination. Fitted values from the model are superimposed (- - - -).

- Multivariate techniques may also be useful for modelling the response variable simultaneously at different grid points, instead of performing a separate analysis at each.


# Acknowledgements

Ian Jolliffe and David B. Stephenson commented usefully on an earlier version.

# References

Milliken GA and Johnson DE (1984) *Analysis of Messy Data Volume I: Designed Experiments*. Wadsworth.

Sexton DMH, Grubb H, Shine KP and Folland CK (2003) Design and analysis of climate model experiments for the efficient estimation of anthropogenic signals. *Journal of Climate*, **16**, 1320–1336.

von Storch H and Zwiers FW (2001) *Statistical Analysis in Climate Research*. Cambridge University Press.