An investigation of the pattern scaling technique for describing future climates.

Timothy D. Mitchell

Anno Domini MMI

A thesis,
submitted for examination to the
University of East Anglia in the United Kingdom
pursuant to its requirements for the award of the
degree of Doctor of Philosophy.

Abstract

The technique of pattern scaling was introduced in 1990 to compare equilibrium experiments from global climate models (GCMs). The patterns from GCMs were subsequently combined with simple climate model simulations to allow scenarios of regional climate change to be constructed for a large number of forcing scenarios for which GCM simulations were not available. Although the technique has been widely used in climate change assessments in the last decade, no comprehensive examination of the technique's worth has been made. This thesis provides such an examination.

By using ensemble experiments forced with greenhouse gases from HadCM2 (a fully coupled general circulation model) it is found that the estimates of regional changes in seasonal temperature and precipitation made using pattern scaling techniques are accurate. The accuracy referred to concerns the technique's representation of climate changes modelled in GCMs, rather than the likelihood of the represented changes becoming reality. The accuracy is measured in terms of the statistical and practical significance of the errors introduced by pattern scaling. It is suggested that pattern scaling is extended from the multi-decadal mean to inter-annual variability, for which the technique is also found to be accurate. It is found that when applying pattern scaling to regions, the identification of the pattern and the accuracy of the estimates relies less on the climatic information used to select regional boundaries than on the choice of spatial scale. Using a sample of six fully-coupled GCMs it is confirmed that pattern scaling may be applied to a wide range of GCMs.

Pattern scaling enables an accurate estimate to be made of the regional climate changes that would be simulated by a GCM under different radiative forcing. Pattern scaling facilitates a probabilistic approach to the assessment of future regional climate change.

Acknowledgements

The author wishes to acknowledge the considerable assistance given to him in the preparation, research, and writing of this thesis. In particular he thanks his supervisors Dr. Mike Hulme and Dr. John Mitchell, and also Dr. Tim Osborn, who served on his supervisory committee. However, the author takes full responsibility for all the work described in this thesis, and any faults are his alone.

The author thanks the Natural Environment Research Council and the UK Meteorological Office for providing the financial resources that permitted this thesis to be written. He thanks the Climatic Research Unit of the School of Environmental Sciences at the University of East Anglia for hosting him during his period of research, and the Hadley Centre of the UK Meteorological Office for providing supervisory assistance.

Finally the author thanks his family and friends for their valuable personal support during the period of research. He particularly thanks Suzanne, who was barely a friend when the research began, and by its completion had become his wife!

Contents

Abbreviations

CCSR-NIES	Center for Climate Research Studies and National Institute for Environmental Studies (Japan)
CGCM1	Canadian coupled GCM version 1
CSIRO2	Commonwealth Sci. and Ind. Research Org. (Australia) model v. 2
COP	Conference of the Parties to the UN/FCCC (see below)
CRU	Climatic Research Unit, University of East Anglia, UK
DETR	Department of the Environment, Transport, and the Regions
ECHam3	European Centre / Hamburg model 3
ECHam4	European Centre / Hamburg model 4
ENSO	El Niño Southern Oscillation
GCM	General Circulation Model or Global Climate Model
GFDL	Geophysical Fluid Dynamics Laboratory
HadCM2	Hadley Centre Coupled Model version 2
HadCM3	Hadley Centre Coupled Model version 3
HMSO	Her Majesty's Stationery Office
IAM	Integrated Assessment Model
IPCC	Inter-Governmental Panel on Climate Change
MPI	Max-Planck Institute for Meteorology (Germany)
NCAR	National Center for Atmospheric Research
RCM	Regional Climate Model
RMSE	Root Mean Squared Error
SAR	Second Assessment Report (of the IPCC)
SCM	Simple Climate Model
SRES	Special Report on Emission Scenarios
TGCIA	Task Group on Scenarios for Climate Impact Assessment
UEA	University of East Anglia
UK	United Kingdom
UKCIP	United Kingdom Climate Impacts Programme
UKMO	United Kingdom Meteorological Office
UNDP	United Nations Development Programme
UN/FCCC	United Nations Framework Convention on Climate Change
WWF	World Wide Fund for Nature

Notation

The meanings of the algebraic notations used in the Equations are listed here.

Letters

a	anomaly
D	difference: estimate – simulated
E	estimated pattern
M′	statistical moment
n	sample size
P	response pattern
p	precipitation
S	simulated pattern
s	signal
t	temperature
w	weight
μ	mean
σ	standard deviation

Subscripts

A	forcing scenario A
B	forcing scenario B
b	reference period for anomalies
c	period mean (from control)
g	grid-box within a region
i	grid box
j	period mean (from scenario)
m	simulation
M	ensemble mean
r	region
y	seasonal mean
ν	number of statistical moment

1. Introduction

1.1 Context

It is now generally recognised that climate change is real, and that humans are at least partly responsible. In 1996 a scientific consensus was reached that "the balance of evidence suggests a discernible human influence on global climate" (IPCC, 1996a, p5). In 2001 this scientific consensus was updated:

> "In the light of new evidence and taking into account the remaining uncertainties, most of the observed warming over the last 50 years is likely to have been due to the increase in greenhouse gas concentrations." (IPCC, 2001, p6)

Since climate will continue to change during the 21st century and beyond, a recognition is beginning to dawn that not only must we act to cut anthropogenic emissions, but that we must also adapt to the climate changes that we are already experiencing, and that are yet to come. An old English proverb states that 'forewarned is forearmed', and to arm ourselves appropriately we must be warned of where precisely the dangers lie.

Often it is a single prediction of the likely change in climate that non-scientists seek, whether they are members of the public, 'stakeholders', or government officials. They often treat such a forecast in a similar way to a weather forecast – and perhaps with a similar level of scepticism. However, it is not possible to treat future climate in these 'single-result' terms. Any prediction involves numerous uncertainties, most of which are unavoidable; their sources range from future changes in society to the internal variability of the climate system. Any honest assessment of future climate must take these uncertainties fully into account.

1.2 Uncertainty

It is precisely at this point that a clash appears between the necessity of incorporating uncertainty and the pre-eminent tool for prediction – the numerical model. The numerical weather prediction model has been developed over many decades to do exactly what it says – predict weather. Since the idea that humans might be altering the climate system fired the imagination of the scientific community in the 1970s, the same models have been adapted – with increasing success – to simulate the climate changes of the 20th and 21st centuries, rather than

Thursday's weather. It is results from these global climate models (GCMs)[1] that have motivated the political process towards mitigating emissions, and that have dominated assessments of future climate change and its impacts. However, these models are so computationally expensive that comparatively few simulations can be performed with them at present at present. Although computing speeds are expected to increase further in the near future, the increased resources may be employed as much on improving the models as on increasing the number of simulations that are performed. Without the ability to carry out a very large number of simulations to estimate future climate, it becomes difficult to assess the uncertainties in those estimates.

Faced with this difficulty, many of those who have made estimates of future climate change and its impacts have often ignored the uncertainties. Seeing an already difficult problem of estimating for a century to come, they have made a single estimate of what future climate might look like. This approach has dominated the advice given to the world by the scientific community, particularly the contribution of Working Group II (on impacts, adaptation, and mitigation) to the Second Assessment Report of the IPCC, published in 1996. However, there is a growing recognition of the need to explicitly recognise uncertainty, and frequently multiple 'scenarios' of possible future changes in climate and impacts are presented. We are now moving towards an approach to climate and impacts assessments that is, at heart, probabilistic.

The probabilistic approach treats uncertainty in quantitative terms. It attempts to describe the range of possibilities and attach likelihoods to them, so that a probability distribution may be formed. This may be done with a variety of methods, ranging from literature reviews, through expert judgement solicitations, to highly complex mathematical models. The probability distributions – and thus the sources of uncertainty that they represent – may then be combined through Bayesian logic and Monte Carlo sampling into an overall assessment of future change that includes all the uncertainties of which we are aware.

However, adopting a probabilistic approach does not, in itself, resolve the clash between the need to accommodate uncertainty and the computational expense of GCMs. The warning to adapt must be based on information that is specific to the

[1] GCM denotes 'general circulation model' (*e.g.* IPCC, 1996d, p566), although it is sometimes taken to mean 'global climate model' (*e.g.* Barrow *et al.*, 2000). Some are spectral models and others use a fixed grid, but for convenience we will refer to the spatial element of any GCM as a 'grid-box'.

region under consideration. GCMs are the most credible tool for assessing future *regional* climate change, yet these are the very same models with which we cannot simulate all the possible futures presented to the climate system as a result of anthropogenic emissions and other sources of radiative forcing. We can simulate the future *global* climate changes using simple climate models (SCMs), but not region-specific, season-specific, and variable-specific changes. A technique that attempts to bridge this gap lies at the centre of this thesis.

1.3 Pattern scaling

The technique of pattern scaling attempts to combine the advantages of both SCMs and GCMs, while overcoming their disadvantages. When using this technique we represent the spatially-varying changes in seasonal climate – derived from a GCM – as a time-invariant pattern of response to radiative forcing. We then scale this pattern by the global-mean temperature change simulated by SCMs, for which we have many thousands of simulations. Thus we obtain the region-specific,[2] season-specific, and variable-specific changes that we require for the full range of possible future radiative forcings.

Elements of this technique were first introduced in 1990 by Santer *et al.*, and it was subsequently developed into the form we have described. It has been widely used in the 1990s in climate impact assessments, but it has not received the close scrutiny that such a pivotal technique demands. We attempt to provide such an examination in this thesis. After reviewing the relevant literature (chapter 2) and detailing our data and methods (chapter 3), we will address four questions of particular interest. These questions, we believe, are highly relevant if pattern scaling is to fulfil its potential in climate change assessment. We ask whether it is legitimate to apply pattern scaling to:

- estimate the multi-decadal mean? (chapter 4)
- estimate the inter-annual probability distribution? (chapter 5)
- regions? (chapter 6)
- a variety of models? (chapter 7)

In the final chapter (8) we draw our conclusions and suggest ways in which the technique of pattern scaling may be further developed and applied.

[2] In this thesis 'region' may refer to any spatial area smaller than a continent. Depending on the context it may refer to a single GCM grid-box, or to an amalgam of GCM grid-boxes.

If pattern scaling is a legitimate method of representing possible future changes in regional climate, the consequences for climate assessment may be great. Pattern scaling is not a technique that will necessarily be seen or understood by those outside the scientific community, but it may be crucial in enabling a fully probabilistic approach to climate assessment to be developed. By using SCMs to represent the full range of possible future changes in global climate, and by using GCMs and pattern scaling to represent the full range of possible future changes in regional climate, we may present a full picture of the possible and likely changes in regional climate. Thus much of the gap between managing uncertainty and using numerical models may be bridged.

2. Literature Review

2.1 Introduction

The work of this thesis centres on the technique of pattern scaling. Therefore we begin this literature review by placing the technique in the wider context to which it belongs. We introduce climate scenarios as the most common approach to presenting the regional climate changes that we may encounter in the future (section 2.2). We discuss both the uncertainty arising from model deficiencies and the uncertainty inherent to the system being studied (section 2.3), and review the probabilistic approach to dealing with uncertainty (section 2.4). Climate assessments often begin with the future course of anthropogenic emissions (section 2.5). We introduce simple climate models (SCMs) as a tool to manage this source of uncertainty (section 2.6), and pattern scaling as a technique by which to relate SCM results to GCM simulations of regional climate change (section 2.7).

We then devote four sections to reviewing some aspects of pattern scaling that we intend to address in our four analysis chapters.

- We review some of the issues that are basic to pattern scaling (section 2.8).
- We introduce the concept of applying pattern scaling to variability, rather than just an inter-decadal mean (section 2.9).
- We discuss some important spatial questions that have not been addressed in climate assessments and that have particular relevance for pattern scaling (section 2.10).
- We consider some of the issues involved when we apply pattern scaling to a number of different GCMs (section 2.11).

Finally we draw the discussion to a close by summarising our assessment of the relevance of pattern scaling (section 2.12).

2.2 Climate scenarios

The result of the efforts to detect and attribute climate change is that "the balance of evidence suggests that there is a discernible human influence on global climate" (IPCC, 1996a, p5). Consequently those with an understanding of global climate have a responsibility to make clear the future changes in climate that may be

expected.[3] The most popular method is scenario analysis,[4] a 'climate scenario' being simply "a description of a possible future climate" (Pittock, 1993, p481) – one of a number of "pertinent, plausible, alternative futures" (IPCC-TGCIA, 1999, p24).

The historical development of climate scenarios is tied to the idea of human-induced climate change. The origins of the idea that humans might be enhancing the natural greenhouse effect through emissions of CO_2 stretch back into the 19th century (Tyndall, 1863; Arrhenius, 1896 a,b), but it did not "fire the imagination of the scientific community" until the 1970s (Kellogg, 1987, p113). Because of the concern that climate change might adversely affect humans, climate scenarios were introduced to link possible climate change under enhanced greenhouse warming with its effects on ecosystems.

Hulme and Carter (1999) have documented the subsequent development of climate scenario methods over the last two decades. They describe this development as being "closely related to the availability and sophistication of results from climate change simulations made with global climate models" (p12). It is true that other sources of information have also been used to develop climate scenarios. IPCC-TGCIA (1999) classify these into:

- synthetic scenarios, where climate variables are changed by a realistic but arbitrary amount;
- analogue scenarios, where future climate is equated with a past climate, or with a present climate from a different location.

However, although they have their advantages, the former are arbitrary and the latter equate future climates influenced by an enhanced greenhouse effect with climates caused by other factors. IPCC-TGCIA (1999) conclude that "GCMs offer the most credible tools for estimating the future response of climate to radiative forcing" (p36).[5]

---

[3] This responsibility has been recognised for many years. For example, the IPCC First Assessment Report (Tegart *et al.*, 1990, p424) insisted that "the most essential need is for more reliable and detailed (both in space and time) estimates of future climatic conditions. These estimates must be regionally specific and provide information on both the frequency and magnitude of events."

[4] Katz (1999, p44) is explicit: "The technique of scenario analysis is most relied on today in climate impacts research." Henderson-Sellers (1996) discusses the use of the term 'scenario', many examples of which may be found in IPCC-TGCIA (1999).

[5] There is a body of scientific opinion that challenges the dominance of GCMs in the provision of policy-relevant scientific knowledge, *e.g.* Shackley *et al.* (1998). However, their criticism is chiefly of the dominance of GCMs over other climate models (particularly simple climate models), rather

2.3 Uncertainty

Some aspects of the model-based climate scenarios that have been constructed over the last twenty years have remained remarkably consistent. For example, the equilibrium climate sensitivity under a doubling of CO_2 has consistently been given as the range 1.5 – 4.5 °C in a series of authoritative assessments:

- Two panels convened by the USA National Academy of Sciences (Charney, 1979; Smagorinsky, 1982);
- A summary commissioned by the USA Department of Energy and reviewed by over 300 scientists from 23 countries (McCracken and Luther, 1985);[6]
- The First Assessment Report of the IPCC (Mitchell *et al.*, 1990);[7]
- The Second Assessment Report of the IPCC (IPCC, 1996b).[8]
- The Third Assessment Report of the IPCC (IPCC, 2001).[9]

The same assessments have also been consistent in highlighting their lack of confidence about model projections on regional scales.[10] This lack of confidence arises from model biases and inter-model differences,[11] which are commonly understood to reveal deficiencies in the models themselves. Thus the uncertainty arising from model deficiencies is commonly recognised.[12] Indeed, when the IPCC was asked to prepare a special report on the regional impacts of climate change (Watson *et al.*, 1998), the chair of the IPCC was forced to acknowledge that

---

than of GCMs *per se*: "In highlighting alternative techniques we are not in any sense denying the relative importance of GCMs … Our main argument is that the role of GCMs can only emerge through considering fully the other possible methods that are available." (p183)

[6] "Models indicate that a doubling of the CO_2 concentration would increase the global average surface air temperature by approximately 1.5 to 4.5 °C." (p *xx*)

[7] Under CO_2 doubling, "the global average warming lies between +1.5°C and +4.5°C" (p 135).

[8] "[We use] the range in the estimate of climate sensitivity (1.5 to 4.5°C)" (p39).

[9] "The climate sensitivity … is comparable to the commonly accepted range of 1.5 to 4.5°C." (p8)

[10] For example: "The reader should be aware of the limited ability of current climate models to simulate regional climate change" (Mitchell *et al.*, 1990, p157).

[11] Examples from the assessments cited above include the following: "These models do not yet adequately represent the observed regional features" (McCracken and Luther, 1985, p xx). "The range of simulated scenarios and the model regional biases were still large, so that confidence in the regional scenarios simulated by AOGCMs remains low" (Kattenberg *et al.*, 1996, p339).

[12] Examples that demonstrate how widely it is believed that 'models cannot predict on regional scales' may be taken from a single volume of *Progress in Physical Geography*: Joubert and Hewitson (1997, p52), Schulze (1997, p118), and Wilby and Wigley (1997, p531).

"because of the uncertainties associated with regional projections of climate change" they were unable "to provide quantitative predictions of the impacts of climate change" (Bolin *et al.*, 1998, p *x*).

However, model deficiencies are not the only source of uncertainty. It is important to draw a distinction between deficiencies in scientific tools and uncertainty inherent to the system being investigated (Mitchell and Hulme, 1999), or 'incomplete knowledge' and 'unknowable knowledge' (Hulme and Carter, 1999). The coarse resolution and limited physics in GCMs may be improved or made more complete. In contrast, human actions are not deterministic[13] and climate evolution is not periodic (Lorenz, 1963), so uncertainty is inherent and the future is unpredictable.

Uncertainty tends to make scientists cautious. However, Fowler and Hennessy (1995) argue that since the norm is to assume that the past is a reliable guide to the future, we must challenge the assumption of stationarity by presenting as much information as possible. Moreover, according to Henderson-Sellers (1996), "in the context of information on climate for policy makers scientific (un)certainty is no longer deemed an adequate reason for withholding advice" (p64). That assertion is borne out by the political motivation behind the production of the IPCC report on regional impacts.[14]

It is clear that assessments of future climate change and its impacts are required, but to develop them correctly we must be able to handle uncertainty. Yet the treatment of uncertainty in most assessments has been "quite inadequate" (Katz, 1999, p38). Hulme and Carter (1999) acknowledge that the "suppression of uncertainty" has been "widespread" (p17). In particular:

- significant sources of uncertainty are not considered in the modelling experiments from which scenarios are constructed (Shackley *et al.*, 1998);

[13] Determinism refers to the theory that all events are completely determined by previously existing causes that preclude free will and the possibility that humans could have acted otherwise.

[14] The report (Watson *et al.*, 1998) was commissioned to provide '"a common base of information regarding the potential costs and benefits of climatic change, including the evaluation of uncertainties, to help the COP determine what adaptation and mitigation measures might be justified" (Obasi and Dowdeswell, 1998, vii).

- the flow of information about uncertainty is hindered by a "gulf" (Henderson-Sellers, 1996, p62) between climate modellers and the users of model outputs;[15]
- on occasions no scenario is presented at all (Henderson-Sellers, 1996);
- scenarios are often based on single possibilities among many models, simulations, forcing scenarios, and climate sensitivities (Hulme and Carter, 1999).[16]

Many such objections may be countered by developing multiple climate scenarios (*e.g.* Carter *et al.*, 1991, 1996; Carter, 1998), and Katz (1999) admits that "Alcamo *et al.* [1996] make a compelling case for the usefulness of scenario analysis to policy makers" (p44). However, Katz argues that scenario analysis does not provide a full treatment of uncertainty, and should not be used as a substitute for it. He suggests that an uncertainty analysis should not be limited to just a few scenarios, and that we should take into account the fact that not all scenarios are equally likely.

> "Thus, one is drawn to think about having infinitely many scenarios and weighting the output corresponding to each scenario by its likelihood; that is, a formal probabilistic approach." (Katz, 1999, p41)

Considered in this way, probability analysis becomes the natural outcome of the scenario approach. Indeed, the underlying problem of climatic prediction is itself probabilistic:

> "The migration away from a purely deterministic approach to modelling, emphasises further that all meteorological prediction problems, from weather forecasting to climate-change projection, are essentially probabilistic." (Palmer, 2000)

2.4 Developing a probabilistic approach

The probabilistic approach is an important part of the rationale for the work we present in this thesis. Although a literature has recently developed concerning how a probabilistic approach might be developed, there are only a few practical examples of probability analyses in climate change assessments; current IPCC guidelines do not go beyond the scenario approach (*e.g.* IPCC-TGCIA, 1999).

---

[15] Nonetheless, uncertainties are often explicitly recognised within the assessments, and there is often an overlap in authorship between modellers and others.

[16] This practice has been very common, but it is now being recognised by impacts researchers that this approach is inadequate. A rapporteur from a group of impacts analysts reported as follows: "Using only one GCM scenario for a climate impact assessment misrepresents the uncertainty in the projection of future climate change" (Mortsch, 2000, p113).

We should not apply probability analysis to just a single source of uncertainty, but to each of them.[17] Therefore we must be able to inter-relate the different sources of uncertainty in a climate change assessment.[18] The most common way of conceptualising these relations is as a 'pyramid' (Mitchell and Hulme, 1999), 'cascade' (New and Hulme, 2000), or 'explosion' (Jones, 2000). This is a 'top-down' or 'sequential' conceptualisation of uncertainty that relates external forcings[19] to regional climate change *via* various other sources of uncertainty. This approach is often implemented by including 'one-way'[20] links between different varieties of model, such as radiation models and GCMs.

The sequential approach neglects the feedbacks between different levels in the sequence. It is possible that this neglect might be overcome through integrated assessment models (*e.g.* Alcamo *et al.*, 1996), which offer an alternative to the sequential approach to uncertainty. However, thus far it has not proved possible to use integrated assessment models to treat uncertainty in a non-sequential way (Leemans, 1999),[21] so we adopt the sequential approach.

In order to combine the sources of uncertainty in the sequence it is appropriate to adopt a Bayesian approach (Katz, 1999),[22] which offers two particular advantages:

- we may include subjective assessments of the uncertainty in variables for which we have no simple means of objectively calculating uncertainty;[23]

---

[17] Katz (1999) argues that "anything less than a fully probabilistic approach to uncertainty analysis is inadequate" (p46). However, it is also worth noting that "although integrated assessment is the natural framework to account for uncertainty, it is worthwhile to start by assessing uncertainty in individual model components separately" (p51).

[18] The most important sources of uncertainty to take into account in an assessment of future climate include anthropogenic emissions of greenhouse gases and other pollutants, natural variability in other sources of radiative forcing, atmospheric chemistry, radiative forcing, the global climate response, and regional climate change.

[19] *i.e.* changes in the radiative forcing of the climate system, *e.g.* from anthropogenic greenhouse gas emissions or solar variability.

[20] The term 'one-way' refers to data being passed from model to model in one direction only.

[21] "Unfortunately, a systematic way to include uncertainty in IAMs has not been developed up to date, although some research groups have tried" (Leemans, 1999, p94).

[22] Epstein (1985) provides an introduction to the use of Bayesian statistics in climatology.

[23] For example, in the absence of an authoritative objective assessment of the likelihood of a collapse in the West Antarctic Ice Sheet, Vaughan and Spurge (2001) assess probabilities using expert judgments from a sample of scientists.

- we may combine different sources of uncertainty in the sequence by using the *a posteriori* distribution from one source as the *a priori* distribution for the next source in the sequence.

There are a number of examples in climate change assessments where Bayesian methods have been used to deal with uncertainty. The examples we cite assess uncertainty in:

- model parameters (Shackley *et al.*, 1998);
- climate sensitivity (Tol and de Vos, 1998);
- sea level rise (Titus and Narayanan, 1996);
- regional climate change and hydrological impacts (New and Hulme, 2000);
- irrigation demand (Jones, 2000).

A formal Bayesian approach generally requires the numerical evaluation of very complex integrals, so it is not computationally feasible to solve the probability equations analytically (Katz, 1999). Therefore Monte Carlo techniques are often adopted in which probability distributions are treated through random sampling. The computational demands of Bayesian analysis are such that the examples given above also give examples of Monte Carlo methods.

2.5 Future emissions

The first source of uncertainty that a probabilistic assessment of regional climate change must address is the future course of anthropogenic emissions. By expressing the uncertainty in future emissions – and from natural forcings[24] – in probabilistic terms, we may form the *a priori* distribution with which we must work to develop an assessment of changes in atmospheric chemistry, radiative forcing, global climate change, and regional climate change.

The possible futures for human society are infinite in number and variety; describing them is "a daunting task" that some consider "impossible" (Grübler, 1999, p55).[25] Working Group III of the IPCC recently published the SRES set of emissions scenarios based on four "narrative storylines"; for each storyline a number of modelling approaches were used to develop scenarios, carefully avoiding 'surprise' scenarios and scenarios representing the mitigation of

[24] Ultimately, any probabilistic assessment of future climate ought to include the uncertainty concerning solar and volcanic variability.

[25] With a measure of understatement, the IPCC (2000) noted the complexity of human systems and acknowledged that "their future evolution is highly uncertain" (p3).

emissions (IPCC, 2000). The scenarios numbered no fewer than forty, and the authors were careful to avoid identifying any single scenario as being more likely than another:

> "No judgment is offered in this Report as to the preference for any of the scenarios and they are not assigned probabilities of occurrence … All are equally sound." (IPCC, 2000, pp3-4)

The IPCC (2000) insisted on regarding all forty scenarios as equally likely even when identifying six 'marker scenarios' for the benefit of climate modellers.[26] They also recommended that "more than one family [of scenarios] should be used in most analyses" (p11). This is not the first occasion on which such assertions have been made by those responsible for developing emissions scenarios:

> "We recommend that analysts use the full range of IS92 Scenarios rather than a single scenario as input to atmospheric/climate models … We do not attempt to identify the most likely emission scenario." (Alcamo *et al.*, 1995, pp255, 258)

However, in most climate modelling centres, GCM simulations have been performed for only a few of the scenarios, whether the IS92 or the SRES emissions scenarios. Part of the reason may be that some climate modellers view particular scenarios as more likely than others,[27] but more important are the various pressing demands on computing resources that they face.[28]

2.6 Combining models

Not even the full suite of SRES *marker* scenarios has been represented by simulations from a GCM, still less the set of forty, or scenarios based on 'surprises'

---

[26] "Marker scenarios are no more or less likely than any other scenarios". (IPCC, 2000, p5)

[27] Hulme and Carter (1999) state that "[the SRES] perspective on uncertain world futures has been challenged by natural scientists who are more inclined to view one or two of these futures as more likely than others" (p20). However, it should be noted that the SRES writing team recognise personal preferences – "Preferences for the scenarios presented here vary among users" (IPCC, 2000, p3) – and only insist on a number of scenarios being used from different families.

[28] Authors from the Hadley Centre (Mitchell *et al.*, 1999) state that some modelling centres "do have the computing power" to perform simulations for a range of scenarios (p549), but recommend that this should *not* be done because of their lack of confidence in the models on regional scales, and because of the need for ensembles. "Running climate experiments is expensive in terms of both human and computing resources. It makes little sense to squander these resources to produce detailed scenarios with little credibility when the uncertainties associated with modelling are comparable to those associated with the different emissions scenarios under investigation." (p550)

or mitigation of emissions; nor is this likely in the near future.[29] However, this is not the case with simpler climate models.

Following Harvey *et al.* (1997), we adopt 'Simple Climate Model' (SCM) as a label for the models used to compute CO_2 concentrations and global temperature change. The differences between SCMs and GCMs can be over-emphasised; they have much in common, and they differ mostly in complexity (Harvey *et al.*, 1997).[30] The simplicity of SCMs can be an advantage: Monte Carlo experiments involving many thousands of simulations may be performed because the SCM requires minimal computing resources. For example, MAGICC (Wigley *et al.*, 2000) runs on a PC from a CD-ROM (Hulme *et al.*, 2000) and has been used by New and Hulme (2000) in a Monte Carlo experiment to convert emissions scenarios into global-mean temperatures.

The disadvantage of SCMs is that they are not sufficiently complex to represent regional climate change. Their end product is usually annual global-mean temperature, whereas a climate assessment usually requires regional changes for a variety of climate variables and a variety of time scales. Among climate models only GCMs can provide this level of information.[31]

Therefore, in order to develop a fully probabilistic approach using models, we need some way of combining the computational simplicity of SCMs with the detailed information from GCMs:[32]

---

[29] There are plans to perform experiments with multiple GCMs. "The Hadley Centre is developing a large number (100 or more) climate models, each of which has a different (but plausible) representation of various aspects of the climate system (*e.g.* clouds, carbon cycle)." However, "these models are then run with the same emissions scenario" (UKMO/DETR, 2000, p10).

[30] Harvey *et al.* (1997) give a succinct comparison of SCMs and GCMs: "The essential common features of the models used for climate projection … are that they can calculate the response of surface temperature to radiative forcing, and that they include the ocean, because of its dominant influence on the rate of climatic change. The essential difference between simple and complex models is the degree of simplification, or the level at which parameterization is introduced." (p19)

[31] Even the critical Shackley *et al.* (1998) recognise GCMs as "the only models that are *potentially* capable of producing regional simulations" (p184).

[32] As we noted above, the Hadley Centre plans to run a large multi-model ensemble in the near future. "From this we can build up a picture of the probability of the change being at various levels, such as 10% or 20% more or less rainfall than today." (UKMO/DETR, 2000, p10) However, this does not, by itself, take into account uncertainties concerning forcing scenarios, because among

"The use of AOGCMs for the simulation of regional, time-varying climatic change, and the use of SCMs for more extensive sensitivity and scenario analysis, are both dictated by pragmatic considerations involving computer resources and the level of detail appropriate when coupling various components together." (Harvey *et al.*, 1997, p5)

2.7 Pattern scaling

A decade ago a similar problem was addressed by Santer *et al.* (1990) in an MPI report. They recognised that only GCMs could provide "the space-time resolution and richness of information required by impact analysts" (p1), but that their equilibrium experiments could not satisfy the need for time-dependent information necessary "in the real world" (p6). Their idea was to standardise the spatial response patterns[33] from equilibrium experiments on the basis of their climate sensitivities (thus achieving a measure of comparability between models with very different climate sensitivities), and to use estimated future changes in global-mean temperature to add a time dimension to the equilibrium patterns.

Santer *et al.* (1990) did not envisage a long life for this technique of generating scenarios:

"Such scenarios represent a compromise pending the availability of model output from GCM transient experiments with dynamic oceans." (p7)

However, one of those authors (Wigley) broadened the scope of the technique by using SCMs to generate estimates of the global-mean temperature, and by linking the SCM time series to the GCM pattern *via* global-mean temperature.[34] This idea was employed in the IPCC's First Assessment Report (Mitchell *et al.*, 1990), and was developed further in one of the first integrated assessment models (Rotmans *et al.*, 1994). In this form the technique has become very popular, and has been widely used over the last decade.[35] Thus the lifetime of pattern scaling is extending well beyond the lifetime of equilibrium experimentation, and into the future:

---

the limited number of simulations, the maximisation will be of the number of GCMs rather than of the number of forcing scenarios.

[33] A 'response pattern' is a spatial pattern describing the change in a particular climatic variable in response to radiative forcing (such as that due to increased greenhouse gas concentrations).

[34] Wigley *et al.* submitted a paper to *Climatic Change* in 1990, but the required revisions were never made and the paper was never published (Mike Hulme, *pers. comm.*).

[35] For example, pattern scaling has been widely used in individual country assessments: Hulme *et al.* (1996) for southern Africa, and Hulme and Jenkins (1998) for the UK.

> "In the absence of a larger sample of GCM experiments to draw upon, scaled scenarios will continue to be used in impacts assessments." (Hulme and Jenkins, 1998, pA9)

2.8 Assessing pattern scaling

Although pattern scaling has been widely used in climate change assessments over the last decade, it has received very little attention. The MPI report (Santer *et al.*, 1990) never reached the peer-reviewed literature, and a fuller review of the technique did not appear from any of the original authors for another nine years, until Mitchell *et al.* (1999). Since pattern scaling is an important part of a probabilistic approach to climate change assessments, and since it has many aspects that have not yet received the attention they deserve, we develop this thesis as an examination of the technique of pattern scaling. In our first analysis chapter (4) we go back over some of the ground that has already been partly covered in the literature – chiefly by Mitchell *et al.* (1999) – but in more detail and with fresh perspectives.

There are a number of conceivable ways of defining the response pattern, including:

- a spatial field[36] of anomalies for a particular time slice[37];
- an enhanced signal-to-noise ratio[38] (*e.g.* Jonas *et al.*, 1996);
- linear regression (*e.g.* Mitchell *et al.*, 1999; Hennessy *et al.*, 1999).

The last two methods have the advantage of using information from an extended period; if there is a linear response, they are therefore more likely to capture it in the response pattern. However, the method of Jonas *et al.* (1996) – used to estimate regional temperature changes from early models[39] – has been found to be "sub-optimal" (Mitchell *et al.*, 1999, p550).

---

[36] A 'spatial field' is a set of values, organised spatially, where there is a single value for each spatial location.

[37] A 'time slice' is a period of time during which a particular set of conditions (such as an atmospheric concentration of greenhouse gases) prevail.

[38] A 'signal-to-noise ratio' is the proportion between the numerical magnitudes of the information that is sought (the 'signal') and other information that is present (the 'noise').

[39] They used IMAGE version 1.0 and ECHam1-LSG.

The critical assumption made in pattern scaling is that the climate system responds linearly to radiative forcing.[40] In particular, the regional climate changes must:

- relate *linearly* to global-mean temperature change, not non-linearly or as a step function;
- relate linearly to the *amount* of global-mean temperature change, not to the rate of change.

Examinations of equilibrium (Oglesby and Saltzman, 1992) and transient (Murphy and Mitchell, 1995) experiments have suggested that in general there is a linear relationship. Linearity is also implicit wherever accurate estimates have been made from pattern scaling (see below).

The most important test to apply to pattern scaling is to examine whether a pattern extracted from one GCM experiment gives an accurate estimate of the regional climate change simulated in another GCM experiment. This has been carried out for regional changes in temperature (Jonas *et al.*, 1996; Mitchell *et al.*, 1999) and land surface driving variables (Huntingford and Cox, 2000). In each case favourable results were obtained. In particular, Mitchell *et al.* (1999) found that the errors introduced from estimating temperature were in many cases within the range explicable by the internal variability from a GCM control simulation. Mitchell *et al.* (1999) also investigated precipitation, although they did not carry out a full investigation. They reasoned from a brief signal-to-noise investigation that since the changes in precipitation are "at best marginally significant in many regions" (p574) it would be difficult to identify precipitation responses from internal variability, and they concluded that therefore pattern scaling is not likely to be accurate for precipitation.

The measure by which Mitchell *et al.* (1999) judged the accuracy of the pattern scaling technique is an absolute one. Errors occur when the estimate from pattern scaling is different from the modelled value. Errors may be detected when the difference between the estimated and modelled values is statistically significant at a prescribed level of confidence. However, the statistical is not the only important measure of significance. Errors that are *statistically* significant may be so small

[40] This assumption has been recognised from the outset. Santer *et al.* (1990) remarked that: "The assumption on which this approach is based is that the spatial patterns of temperature change … are stable in the transition from $1xCO_2$ to $2xCO_2$" (p7). Any departures from linearity will introduce errors to the estimates made using pattern scaling. This amounts to an additional source of uncertainty, which is specifically considered in chapter 4 (particularly section 4.5).

(perhaps in relation to the anomaly itself) as to have no *practical* significance. From this perspective we might complement the absolute measure of accuracy by a *relative* measure of accuracy. If we were to reject pattern scaling, the lack of GCM simulations for many forcing scenarios would leave us without any measure of climate change other than the global-mean temperature obtained from a SCM. Therefore it is legitimate to calculate the accuracy of pattern scaling relative to the accuracy of assuming that each grid-box takes the global-mean value. If pattern scaling is more accurate than the global-mean assumption, then there may be cases in which pattern scaling is judged on these grounds to be 'sufficiently accurate' for certain applications.

Thus far we have considered pattern scaling for scenarios of changing greenhouse gas emissions, where the gases are well-mixed and gradually increase over time. Sulphate aerosols are not so well-mixed and concentrations both rise and fall over time. Mitchell *et al.* (1999) found that a pattern from a greenhouse scenario does not give an accurate estimate of regional climate changes under a greenhouse-gas-plus-sulphates scenario.[41] However, it is now believed that the IS92 scenarios adopted by Mitchell *et al.* (1999) contained unrealistically large emissions of sulphur, and the more recent SRES scenarios contain much lower emissions of sulphur (IPCC, 2000).[42] Therefore the effect of aerosols on future climate is also much reduced under the SRES scenarios. Thus the problem posed to pattern scaling by sulphates is not nearly so great as under the IS92 scenarios.

Nonetheless, sulphate emissions are still likely to affect regional climate changes in certain parts of the world. Ramaswamy and Chen (1997) suggested that a response pattern might be decomposed into greenhouse and sulphate components, and that these might be linearly recombined. One of the authors of the original MPI report of 1990 developed this idea by examining whether a sulphate response pattern might be linearly scaled by regional sulphate emissions in a similar manner to the greenhouse response pattern (Schlesinger *et al.*, 1997). It is conceivable that a greenhouse pattern and a number of region-specific sulphate patterns might be individually scaled and linearly recombined to create a single climate scenario.

---

[41] "The results here suggest that scenarios with substantial changes in aerosol emissions are not well represented by a single greenhouse gas pattern." (p576)

[42] "Sulfur emissions in the SRES scenarios are generally below the IS92 range, because of structural changes in the energy system as well as concerns about local and regional air pollution." (p8)

2.9 Pattern scaling the variance

Thus far we have only considered scaling the climatological mean. We have envisaged an approach in which we describe probability distributions for *mean* future emissions, *mean* radiative forcing, *mean* global annual temperature, and *mean* regional climate change. However, many important impacts of climate change may result from changes in extremes.[43] Since a small change in variability has a disproportionately large effect on extreme events (Katz and Brown, 1992),[44] changes in variability are potentially very important. If our climate change assessments, carefully developed in a fully probabilistic manner, exclusively address climatic means, they fail to address a critical area of uncertainty, *i.e.* regional changes in climate variability.

Just as only GCMs can represent regional changes in climate *means*, only GCMs can possibly represent regional changes in climate *variability*. In both cases SCMs are required to represent the full range of possible future courses of radiative forcing. It is conceivable that we could represent a particular change in variability in a GCM – in the interannual variance for example – as a response pattern, expressed as a function of global-mean temperature. We could link the response pattern to the SCMs *via* global-mean temperature, as before. This has not been attempted before.

It is conceivable that by calculating response patterns for both a mean and variance, we might be able to describe how the probability function for a particular climatic variable changes under radiative forcing, for as wide a range of forcing scenarios as we can simulate with a SCM. Impacts researchers are beginning to recognise a need for climate scenarios that include changes in probability distributions.[45] However, since different climatic variables have

[43] For example, Working Group II of the IPCC introduced their assessment of climate change impacts by pointing out that "potentially serious changes have been identified, including an increase in some regions in the incidence of extreme high-temperature events, floods, and droughts, with resultant consequences for fires, pest outbreaks, and ecosystem composition, structure, and functioning, including primary productivity" (1996c, p3).

[44] "A small change in the variability has a stronger effect than a similar change in the mean." (IPCC, 1996b, p44)

[45] A rapporteur from a group of impacts analysts reported that "statistical distributions (pdf's) including their change in time are needed" (de Ronde, 2000, p101).

different distributions,[46] the most appropriate measures by which we may describe probability distributions may vary between climatic variables. Some distributions may, for example, be skewed, and so require a third parameter in order to accurately describe the probability function. In chapter 5 we examine the possibility of applying pattern scaling to higher statistical moments than the mean.

2.10 Spatial questions when pattern scaling

Chapter 6 examines some spatial questions that are particularly relevant when conducting climate assessments in general, and when pattern scaling in particular.

Recent years have seen an increasing number of geographically-specific assessments carried out both for particular countries[47] and for sub-continents[48]. We have already remarked that in many cases climate scenarios are developed for, and motivated by the need for, an assessment of the potential impacts of future climate change: they are aimed at policy-makers,[49] and they are prompted by political concern.[50]

---

[46] *e.g.* monthly temperature often has a Gaussian distribution, and monthly precipitation often has a gamma distribution.

[47] Examples of national assessments include those for the UK (CCIRG, 1991) and Finland (Kuusisto *et al.*, 1996).

[48] An early example of a sub-continental assessment was for Southern Africa (Hulme *et al.*, 1996). Later an IPCC assessment of the regional impacts of climate change (Watson *et al.*, 1998) was prepared for ten continental regions that together cover the earth's entire land surface and adjoining coastal seas.

[49] The editors of the IPCC assessment of the regional impacts of climate change noted that their report "represents an important step forward in the evolution of the impact assessment process for the IPCC … This report analyzes impacts at a continental or subcontinental scale that is of more practical interest to decisionmakers." (Bolin *et al.*, 1996, *ix*)

[50] Political concern is both inside and outside governments. For example, the UK Secretary of State for the Environment employed some predicted regional impacts from research sponsored by his department to argue for reductions in greenhouse gas emissions (Prescott, 1999), whereas the WWF funded the assessment for Southern Africa (Hulme *et al.*, 1996). At the global level, the IPCC assessment of the regional impacts of climate change was prepared at the request of a political body subsidiary to the UN/FCCC (Obasi and Dowdeswell, 1998), something that is true of all IPCC reports: "The IPCC was established as an intergovernmental body with the express purpose of assessing for the policy community a focused set of issues salient to policy formulation." (Moss, 2000, p461)

These impact and policy-related motives influence the methods that are employed in climate assessments. This may be seen in the methods by which different spatial scales are bridged. A major problem in developing climate scenarios from GCM simulations lies in the nature of the GCM outputs, which are rarely at the spatial scale or with the boundaries that are most convenient for the anticipated users of the climate scenarios. An entire area of research has developed around the dynamical or statistical 'downscaling' of GCM output[51] to the spatial scales of hydrological catchments, or even smaller areas. However, it is not always necessary to downscale GCM output before incorporating it into climate scenarios. There are examples in the literature of:

- an assignment being made of the nearest grid-box estimate (Croley, 1990);
- laying a mesh of GCM grid-boxes over the local topography (Hulme *et al.*, 1996);
- representing a grid-box as a geographical region (*e.g.* Eastern England, Hulme and Jenkins, 1998).

The requirements of impact analyses and policy initiatives do not only affect the spatial *scale* at which climate scenarios are constructed. The geographical boundaries of the scenarios are often determined by the same requirements. Thus we find examples of boundaries formed by:

- political states (*e.g.* Carter, 1996; Hulme and Jenkins, 1998);
- topographic features (*e.g.* Croley, 1990).

However, if we move beyond impact-related assessments to the literature of GCM evaluation, we find that GCMs are treated at different spatial scales from the individual grid-box, and with different boundaries. Climate modellers are well aware of the uncertainties in their science,[52] including uncertainties about models on regional scales.[53] This makes them reluctant to use their models in a predictive sense on regional scales,[54] despite their confidence in them at global scales.[55]

---

[51] For reviews of downscaling techniques see Hewitson and Crane (1996), Kattenberg *et al.* (1996), and Wilby and Wigley (1997). 'Dynamical' downscaling includes the use of regional climate models, and 'statistical' downscaling often employs regression techniques.

[52] "We … believe that most, if not all, climate scientists are only too aware of the uncertainties both in 'their' models and throughout policy development." (Henderson-Sellers and McGuffie, 1999, p597)

[53] We remarked upon this and cited key assessments in section 2.3.

[54] A sense of the compulsion required to use GCMs in this way is palpable in the First Assessment Report. Mitchell *et al.* (1990) stated that "In order to assess the impacts of future changes in climate,

Therefore any discussion of regional climate change in GCM evaluations is generally based on spatial scales that are much larger than would generally be described as 'regional'. Two approaches are popular:

- Relatively large geometric shapes are defined, generally representing a crude division into climatic zones.[56] A regional mean for each shape is created by averaging together an aggregation of grid-boxes.
- Results are plotted for each grid-box for the globe, with no commitment to the accuracy of any individual grid-box.[57]

To judge whether the use of individual grid-boxes in the construction of climate scenarios is justified is beyond the scope of this review.[58] It must suffice to make two salient points:

- there is a substantial literature on the subject of uncertainty in climate science from which justifications might be given, as we have demonstrated in previous sections;
- such use is increasing rather than diminishing.[59]

---

one needs to know the changes and rates of change in climate on a regional scale." (p155) Having given a summary of the state of the science from which the unstated conclusion was that this need had not yet been met, they remarked that "Nonetheless, one of the briefs of Working Group I was to provide estimates of changes in 5 selected regions" (p155), and proceeded to do their best to meet their brief.

[55] This may appear to be unjustified, but is in fact based on physical reasoning, as Shackley *et al.* (1998) recognise: "It is assumed by many GCM modellers ('GCMers') that the regional errors do not compromise the validity of the global response to CO_2 forcing. This is because, providing the major, large-scale features are simulated and energy at the top of the atmosphere (TOA) is balanced, then the model will produce a new global equilibrium position which is largely independent of what occurs regionally." (p162)

[56] A continuous thread of development for one such division may be found in Mitchell *et al.* (1990), Kattenberg *et al.* (1996), Kittel *et al.* (1998), and Giorgi and Francisco (2000).

[57] A series of examples may be found in the publications of the Hadley Centre: UKMO / DETR (1997, 1998, 1999).

[58] Note that we refer here to the value of GCMs at the grid-box scale, not the accuracy of pattern scaling at the grid-box scale.

[59] In recent years concerted efforts have been made to improve the supply of grid-box level information to those concerned with the potential impacts of climate change. One example is the IPCC Data Distribution Centre, an IPCC task force charged with the free distribution of climate scenarios and observations. Another example is the circulation by a UNDP subsidiary of software and literature (Hulme *et al.*, 2000) on generating climate scenarios from GCM data that aims to "give countries the tools they need to develop their own solutions" (p4).

It is in this context of a plausible and increasing use of information from the individual grid-box that we choose to employ data from individual grid-boxes in our initial investigations into the possibility of scaling means (chapter 4) and variances (chapter 5).

There is also a case for examining the value of scaling for regions made up of aggregations of grid-boxes. We present a number of relevant points:

- Such regions have an established place in climatology. Analyses of regions have been present in the literature that compares GCMs on sub-global scales for the last decade, and further analyses are expected in the chapter on regional climate change in the IPCC Third Assessment Report.
- Such regions are used to convey sub-global climatic information. They have been one of the principal forms through which the reports of IPCC Working Group I have supplied information on regional climate change during the 1990s.[60]
- Such regions exhibit some agreement between models. It is only for large spatial scales that any claim is made for qualitative agreement between models.[61] (It is recognised that the models could all be wrong in the same way.)
- Such regions are relevant to policy. It is partly potential climate impacts on such spatial scales that give a sense of urgency to the political moves towards mitigation (*e.g.* Prescott, 1999).
- Such regions have already been subject to some evaluation with regard to pattern scaling (Mitchell *et al.*, 1999).

There are unanswered questions about the aggregation of grid-boxes that are relevant to pattern scaling. It is a long-standing problem in the analysis of spatially-aggregated data that "for some analyses the results depend on the definition of the areal units for which data are reported" (Fotheringham *et al.*, 2000, p237). Known as the 'modifiable areal unit problem' (Openshaw, 1984), it has two components:

---

[60] Such regions provided the sole form in the IPCC report of 1990 (Mitchell *et al.*, 1990). Information from statistical and dynamical downscaling was added in the supplementary report of 1992 (Gates *et al.*, 1992). All three forms were included in the Second Assessment Report (Kattenberg *et al.*, 1996) and the supplementary report on the regional impacts of climate change (Giorgi *et al.*, 1998).

[61] Mitchell *et al.* (1990) remarked that there was "agreement between models on the qualitative nature of the large scale changes in temperature and to a lesser extent precipitation" (p155). Kattenberg *et al.* (1996) noted that "several instances occurred in which regional scenarios produced by all models agreed, at least in sign" (p339).

> " 1. *The scale effect*: different results can be obtained from the same statistical analysis at different levels of spatial resolution.
> 2. *The zoning effect*: different results can be obtained owing to the regrouping of zones at a given scale." (Fotheringham *et al.*, 2000, p237)

This problem is not easily solved.[62] Among the possible solutions is the use of spatially disaggregated data, but we are already considering individual grid-boxes, and our interest here is specifically in aggregated data. Another possibility is to construct the regions using some form of optimisation (Openshaw and Rao, 1995). To a certain extent this already occurs, in that the IPCC regions are loosely based on climatic regions; there is potential for a further refining of such regions on the basis of the model topography or its control climate. The ultimate source of information for optimisation would be the response pattern itself, but this would preclude the use of the optimised regions to evaluate the effects of regionalisation on scaling,[63] and introduces a further problem, in that such regions would not be applicable to more than one model.[64] A third possible solution was demonstrated by Fotheringham and Wong (1991): the definition of regions may be carried out a number of times for different methods of regionalisation to assess the stability of a particular result to the regionalisation choices made.

Moreover, the problem has important implications. Mitchell *et al.* (1999) found that the estimated decadal changes using the five original IPCC regions (Mitchell *et al.*, 1990) were an order of magnitude larger than the decadal standard deviation for temperature (their Table II), but not for precipitation (their Table III). They argued that their results suggest that "many of the regional changes in precipitation may not be significant, and hence a large part of the spread in regional precipitation changes from different models is an artefact of internal variability" (p576).

---

[62] Fotheringham *et al.* (2000) comment that "Despite the fact that the modifiable areal unit problem has been observed for decades, we appear little closer in dealing with the problem effectively" (pp237-238), and that "Until a general solution is found, the modifiable areal unit problem will continue to create uncertainty in the widespread applicability of spatial analytical results." (pp239-240)

[63] The reason for the preclusion is the circularity in logic. To define regions on the basis of the model's response to forcing, to apply scaling to those regions, and to evaluate regionalisation on that basis, is circular reasoning.

[64] To create a new set of regions for each GCM does not offer a solution. "If we create one zoning system for one regression model, then include another variable and rezone to get the 'best-fit', then it is not immediately clear how we compare one model with another" (Fotheringham *et al.*, 2000, p238).

However, would those results and conclusions change if the regions were defined differently? If the regions were not defined climatically, would the positive result for temperature still stand? Conversely, if the regions were defined in a way that was more sensitive to the model's precipitation climatology, would the negative result for precipitation still stand? If the spatial scale was reduced would the results still stand?

There are good grounds for examining regions constructed from aggregations of grid-boxes: they are well-established, well-used, climatically-relevant, policy-relevant, and have already been employed in an investigation of pattern scaling. There are questions about the choices we make in aggregating grid-boxes together to form regions that have the potential to influence the conclusions we draw concerning the validity of applying pattern scaling to such regions. These questions concern the spatial scale and the geographical boundaries that are chosen for the regions; they are examined in chapter 6.

2.11 Pattern scaling for multiple models

In chapter 7 we turn our attention to the application of pattern scaling to a sample of six GCMs from the IPCC Data Distribution Centre (DDC). When we attempt to combine different GCMs in a climate assessment we have two alternatives:

- treat the sample of GCMs as a multi-model ensemble, extract a single response pattern from the sample, and scale it for the purposes of a climate assessment;
- treat each GCM as an internally consistent model, extract a response pattern from each model, and use the patterns to give a range of scaled estimates in a climate assessment.

Both are legitimate approaches to take: the former emphasises what the models have in common, whereas the latter tends to emphasise the differences between the models. If our intention is to provide a single best estimate of climate change under a particular forcing scenario, then the former approach may be more appropriate.[65] After all, there is some evidence that the inter-model mean may be closer to the real world than any individual model (Lambert and Boer, 2000). However, if we are concerned to incorporate *all* the uncertainties in our climate assessment then we must recognise that there are substantial uncertainties over

[65] However, we must recognise that although the response patterns from individual GCMs may each be physically reasonable, the inter-model response pattern may not be. This is not necessarily an obstacle to the use of the inter-model mean pattern in pattern scaling, providing that regions are considered individually rather than in relation to each other.

how a GCM ought to represent the climate system. These uncertainties may, at least partly, be incorporated by including the full range of GCM patterns in our assessment.

Differences between models arise partly because of the limitations of human understanding of the climate system, and partly because of the limitations of a mathematical model. For example, there are major uncertainties in the parameterisations that may be fundamental and irreducible (Palmer, 2000). It is important to recognise that all of the models may be regarded as plausible,[66] and that no model may be regarded as the 'best'.[67] All of the models in the DDC sample have weaknesses, and some may be regarded as more plausible than others, but none can be excluded. The range of plausible models that could conceivably be constructed is very great – indeed, it is beyond measurement – and it is beyond the powers of humans to fully occupy that range with models.[68] However, if we are to provide as full an assessment as possible of the uncertainties in climate scenarios, we must represent the plausible range as fully as possible.

This problem parallels the problem of internal variability. In both problems the range of possible cases is beyond measurement, and cannot be fully represented. We cope with the inherent uncertainty from internal variability by constructing random samples called ensembles. In theory we might mimic this strategy by constructing a sample of models whose architectures we vary (Palmer, 2000). This is rather different from anything currently proposed. In the Casino-21 project (Allen, 1999) random variations are made to a few selected variables in a single well-established model. The Hadley Centre plan to vary a number of sub-models within a GCM and run a simulation for each variety (UKMO/DETR, 2000).

---

[66] "Whilst we know Newton's laws of motion extremely well, there is no unique prescription for representing the governing equations of climate computationally, since the process of parameterisation is not a rigorously (or even heuristically) justifiable procedure in regions of mesoscale organisation." (Palmer, 2000) A practical outcome of this is the intention of the Hadley Centre to develop at least 100 "plausible" climate models (UKMO/DETR, 2000, p10).

[67] Lambert and Boer (2000) remarked that "as is characteristic of intercomparison results, different climate variables are simulated with different levels of success by different models and no one model is 'best' for all variables".

[68] This is the case even for the 'Casino-21' project (Allen, 1999), which aims to run many thousands of variants of HadCM2 in which a small number of parameters are changed. Only a *limited number* of variables will be altered in a *single* model that aims at middle-of-the-road behaviour, that represents the climate system in terms of a specific spatio-temporal structure, and that incorporates only a small number of factors.

However, it is a very different proposition to construct entirely new GCMs on a random basis. The latter is almost inconceivable,[69] despite some evidence that an inter-model mean may be a better predictor than any single model (Lambert and Boer, 2000). Therefore we must represent the model-related uncertainty as best we can, using the models that are available.

However, we must recognise that the DDC sample is not a random sample of the range of plausible models that might possibly be constructed. The members of the DDC sample, each formulated using deterministic equations (whether a spectral model or one based on a fixed grid), may well be more similar to each other than they are to nature (Palmer, 2000).[70] Each model has been constructed by a research groups that has developed a 'middle-of-the-road' model that:

- accords with current best practice,
- is tuned to present conditions,
- is limited by present computing constraints,
- has been developed with the aim of winning acceptance for the model and the experiments carried out with it.

The non-random nature of the DDC sample has implications for the statistical approach we take to the DDC sample. In particular, we cannot carry out inferential statistics with the DDC sample, and the median and range are more appropriate measures than the mean and standard deviation.

2.12 Relevance of pattern scaling

Pattern scaling is a valuable technique in the context of climate scenario development (section 2.2), but it reaches its full potential in a probabilistic approach to climate assessment (section 2.7). Under the approach we suggest, Monte Carlo techniques (section 2.4) may be applied to SCMs (section 2.6) to allow a full assessment of the possible global climate changes that may result from future emissions (section 2.5). Pattern scaling allows us to relate the global climate changes to regional climate change. Thus we may be able to incorporate many

[69] It is inconceivable for a number of reasons, including: the vast funds required, the level of co-operation required between presently competing research groups on different continents, and the requirement to avoid matching the model to observations.

[70] Raisanen (2001) adds: "Even when models have been developed to their final form at different institutions, they often have a family relationship in the form of some similar if not identical components".

sources of uncertainty (section 2.3) into climate assessments, without neglecting the models that provide the best tools to study climate change.

We have reviewed the history of pattern scaling (section 2.7) and considered how we might provide a fuller evaluation of it. In particular, we have:

- extended our examination of the basic concept (section 2.8), in preparation for our fuller analysis of the basic concept in chapter 4;
- suggested applying pattern scaling to higher statistical moments than the mean (section 2.9), which we attempt in chapter 5;
- considered some spatial questions that arise in climate assessments generally, and with pattern scaling in particular (section 2.10), which we address in chapter 6;
- discussed how to employ different GCMs in pattern scaling (section 2.11), which we attempt in chapter 7.

Prior to conducting these analyses in chapters 4-7, we introduce some of our key methods and data in chapter 3.

3. Methods and Data

3.1 Introduction

In this chapter we introduce some of the key methods and data-sets that will be used in our analysis. We begin by describing the technique of pattern scaling in more detail (section 3.2). We explain our choices of variable, season, and forcing scenario in section 3.3. In section 3.4 we briefly describe and reference the model simulations we will be using. In section 3.5 we describe permutation testing, a family of distribution-free techniques that we adopt in chapter 6.

3.2 Pattern scaling

In section 2.7 we introduced pattern scaling as a technique by which we may construct regional climate scenarios. The method may be broken down into three steps:

- a GCM response pattern is extracted from a transient simulation and is expressed in terms of the scaler,[71] *i.e.* per °C change in the annual global-mean temperature;
- the selected forcing scenario is simulated in a SCM with the selected climate sensitivity, giving a time series of the scaler;
- for any selected time period, the GCM pattern is multiplied by the value of the scaler from the SCM to give the regional climate scenario.

We illustrate this method in Figure 3.1. A probabilistic approach may be developed by repeating the final two steps a large number of times.

We evaluate the method using GCM simulations. The best test of the method is to:

- extract the response pattern from one forcing scenario;
- obtain the time series of the scaler from a second forcing scenario;
- estimate the anomaly pattern for a period in the second forcing scenario;
- compare the estimated anomalies with the modelled anomalies.

Internal variability complicates the evaluation. Differences might arise between the estimated and modelled anomaly patterns as a result of internal variability, despite having obtained an accurate estimate of the average modelled anomaly pattern from the scaled response pattern. Therefore to evaluate the method we

[71] We use 'scaler' throughout, rather than 'scalar'. The former refers to a ratio of size, whereas the latter refers to a quantity having magnitude but not direction.

need not only multiple forcing scenarios, but an ensemble for each forcing scenario in which the only differences between simulations arise from initial conditions.

There are a number of conceivable ways in which we might calculate the GCM response pattern. Following Mitchell *et al.* (1999) we use a method of linear regression by which we relate the time series in each region[72] to the scaler. Having treated each region individually, we produce a response pattern by combining the regional results.

The time periods that we use for regression depend on the simulations that are available. In chapter 7 we consider a variety of models, so we are restricted to the common period (1990-2099) available from those models, as we describe in section 7.2. However, for most of our analyses it is vital that we have ensembles for more than one forcing scenario. Therefore in chapters 4, 5, and 6 we use the HadCM2 ensembles, for which the period 1860-2099 is available.

In order to maximise the signal-to-noise patterns and construct the patterns as robustly as possible, it will be necessary to smooth the time series. In section 4.4 we briefly investigate the effects on the signal-to-noise ratio of smoothing at different windows. We smooth by calculating the means of overlapping periods in the time-series, upon which we base the linear regression in chapters 4, 6, and 7. In chapter 5 we investigate inter-annual variability, so we use the same overlapping periods but calculate higher statistical moments: the variance, skewness, and kurtosis. Where ensembles are available we base the regression on the individual means from each simulation, rather than the ensemble means.

Before regressing we convert the regional and global time series into anomalies. By using anomalies rather than the absolute values, and by assuming that the long-term mean of the regional anomaly is zero when the scaler anomaly is zero, we may simplify the regression equation in chapters 4, 6, and 7. Rather than relating the regional time series (y) to the scaler time series (x) using $y=ax+b$, we may assume that the intercept (b) is zero and use $y=ax$ instead. We use the gradient (a) to form the response pattern.

[72] Here we use 'region' to refer to the sub-global area that we relate to the global-mean. In practice a sub-global area may be an individual grid-box or the mean of an agglomeration of grid-boxes. We compare these alternatives in chapter 6.

Although we will mostly use linear regression to define our response patterns, on occasion we will use the time slice method to investigate the time-dependence of the response patterns. This method is very simple: we calculate the regional anomaly and the scaler anomaly in the chosen period, and divide the former by the latter to obtain the response pattern. Where ensembles are used we use the ensemble mean anomalies.

3.3 Subjects of investigation

Much of the evaluation of pattern scaling in the literature has been conducted on just a single variable – surface temperature. In many ways this variable may represent the best opportunity for pattern scaling:

- temperature is the same variable as the scaler,
- temperature is spatially continuous,
- temperature is represented well by GCMs,
- temperature has a high signal-to-noise ratio under radiative forcing.

Although pattern scaling is potentially of great value for other variables, its success for temperature may not be repeated for other variables. Therefore throughout our analyses we consider not only temperature – to provide a benchmark and to relate our analyses to previously published work – but also precipitation, which offers a much sterner test for pattern scaling:

- precipitation is *not* the same variable as the scaler,
- precipitation is spatially *dis*continuous,
- precipitation is *not* represented as well as temperature by GCMs,
- precipitation has a relatively *low* signal-to-noise ratio.

Another limitation of the evaluation of pattern scaling in the literature is its restriction – in most cases – to annual means. In many parts of the world the climate changes – particularly for precipitation – will differ between seasons, so we need to apply pattern scaling to individual seasons. Therefore we focus our attention upon a single season (JJA), rather than upon the annual mean.

We have already discussed the importance of sulphate aerosols (section 2.8). However, for a number of reasons any treatment of sulphates will remain outside the scope of our analyses:

- Most of the sulphate-forced simulations that were available when we conducted our analyses used the IS92 forcing scenarios, which are now regarded as including too large a sulphate forcing (IPCC, 2000). Therefore it

would be hard to draw worthwhile conclusions from those simulations about the value of pattern scaling in the context of climate assessments.

- In order to fully examine the accuracy of scaling we need model simulations from different forcing scenarios. Moreover, for pattern scaling to be accurate the forcing must retain the same spatial pattern between scenarios, or else the response pattern is likely to vary between scenarios. However, the only model for which we have ensembles for different sulphate scenarios is HadCM2, and those scenarios do not have the same spatial patterns of sulphate forcing (Mitchell *et al.*, 1999).
- In order to assess pattern scaling for a number of models we need simulations with a common forcing scenario. However, the various climate modelling centres have used different sulphate forcings in their simulations.

Therefore we restrict our analysis to greenhouse-gas-only simulations, which pass each of the requirements that the sulphate-forced simulations fail:

- The greenhouse forcing scenarios for which we have simulations are regarded as plausible.
- We have ensembles from different greenhouse forcing scenarios between which the spatial pattern of forcing remains constant.
- We have simulations from different models with near-identical greenhouse forcing scenarios.

3.4 Model data

It is important to evaluate pattern scaling for a variety of GCMs: one GCM differs from another (section 2.11), and since GCM-related uncertainty should be taken into account in any climate assessment, we need to apply pattern scaling to GCMs in general rather than to a single example. Equally, in order to conduct a thorough examination of pattern scaling we need ensembles of simulations for a variety of forcing scenarios. Our problem is to meet both these requirements: we have simulations for a number of GCMs, but multiple ensembles only for HadCM2. Our solution is to restrict most of our analysis (chapters 4, 5, and 6) to the HadCM2 simulations, but to broaden the relevance of our results by comparing a wider variety of GCMs in chapter 7. The reasons for our selection of the particular GCMs used in chapter 7 are given in section 7.2, together with appropriate references (Table 7.1).

The second version of the Hadley Centre coupled model (HadCM2) has probably been used in more climate impact assessments than any other GCM,[73] so it is a highly relevant model for which to examine the accuracy of pattern scaling. Many aspects of the model have received detailed assessment.[74] In addition, detailed assessments of the model have been made for particular regions, notably for the North Atlantic storm track region.[75]

HadCM2 is a coupled atmosphere-ocean model and includes representations of both the land surface and sea ice. It is one of a series of models that have been successively developed by the Hadley Centre, and it runs under the Unified Model system (Cullen, 1993) for both meteorological and climatological models. The spatial structure of the model takes the form of a grid with a horizontal resolution of 2.5° latitude and 3.75° longitude. There are 19 vertical levels in the atmosphere and 20 vertical levels in the ocean. Conditions in the model atmosphere and ocean are calculated on time scales of minutes, and the two are coupled on a daily time scale. Flux adjustments are required to prevent model drift.

Our interest is particularly in the long control simulation (Tett *et al.*, 1997),[76] in the 'Ga' and 'Gd' four-member ensembles of greenhouse-gas-only simulations (Mitchell *et al.*, 1999), and in the two stabilisation simulations – s550 and s750 (Mitchell *et al.*, 2000). We are interested in this set of simulations because they have been performed with a single GCM, and because as a group they offer extensive opportunities for evaluating pattern scaling. We describe the simulations below:

- The four simulations in the Ga ensemble receive identical radiative forcing, but their initial conditions were taken from points 150 years apart in the control simulation. The radiative forcing in Ga represents estimated changes in greenhouse gas concentrations for the period 1860-1990, and a 1% per annum

[73] An important reason for this is the dissemination of the model output via the LINK project.

[74] The list is lengthy even if we restrict our list to Hadley Centre assessments of aspects of the model on a global scale, which include: the model itself and its current mean climate (Johns *et al.*, 1997); its variability (Tett *et al.*, 1997); its climatology in the mid-Holocene (Hewitt and Mitchell, 1998); its response to radiative forcing (Tett *et al.*, 1996); the contribution of sulphate aerosols (Mitchell and Johns, 1997); the ensembles for four forcing scenarios (Mitchell *et al.*, 1999); the simulations of stabilisation scenarios (Mitchell *et al.*, 2000).

[75] For example, Osborn *et al.* have compared the modelled North Atlantic Oscillation (1999a) and downscaled UK climate (1999b) with observations, while Carnell and Senior (1998) have examined the storm track under increased radiative forcing.

[76] We have 1400 years of continuous control simulation.

increase in concentrations for the period 1990-2099 that approximates the IS92a scenario (Mitchell and Gregory, 1992).

- The four simulations that make up the Gd ensemble begin in 1990, having been initialised from their respective Ga simulations in 1990. However, when we analyse the Gd simulations we add the period 1860-1989 from their respective Ga simulations to give the full length of 240 years. The radiative forcing for the period 1990-2099 is weaker in Gd than in Ga, and represents concentrations approximating to the IS92d scenario (Mitchell and Gregory, 1992).
- The s550 simulation (1990-2259) is initialised from Ga 2 (John Mitchell, pers. comm.), so when we analyse it we add the period 1860-1989 from Ga 2.[77] The carbon dioxide concentrations stabilise at 550 ppm[78] by 2150, but other greenhouse gases remain at 1990 levels throughout the simulation.
- The s750 simulation (1990-2249) is similar to s550, but the carbon dioxide concentrations stabilise at 750 ppm by 2250.

3.5 Permutation testing

In a number of places in chapter 6 we use a distribution-free method of statistical inference to compare methods of selecting regions. Statistical tests have been widely used for many decades to permit inferential reasoning on the basis of a sample. However, many tests – such as Student's t-test – make assumptions that strictly restrict their application, frequently including the assumption that the reference and test distributions are normally distributed.[79] There will be occasions in chapter 6 where this assumption is not justified, so we turn to a family of distribution-free tests called 'permutation tests' or 'randomisation tests'.[80]

The logical basis for permutation testing is the principle of randomisation, first conceived by Sir Ronald Fisher in the 1930s.[81] The principle is easiest to grasp

[77] 'Ga 2' denotes the second member of the Ga ensemble.

[78] 'ppm' denotes parts per million

[79] Such assumptions are often breached in practice, but at the cost of statistical accuracy. In the official journal of the American Heart Association, Glantz (1980) claimed that: "Approximately half the articles published in medical journals that use statistical methods use them incorrectly"; most of the errors tabulated concerned Student's t-test.

[80] Sprent (1998) provides a good introduction to permutation testing.

[81] It was Pitman (1937a-c) who provided the first theoretical framework for randomisation tests, but he did not claim priority of the idea: "The main idea is not new, it seems to be implicit in all Fisher's writings" (1937a, p119). Fisher (1935) developed the principle to test the null hypothesis

when we consider two random samples that have been drawn from the same population. We may regard them as a single sample, the individuals of which have been randomly divided into two sub-samples of the sizes actually drawn. In theory the process of randomly dividing the sample into two sub-samples might be done for every possible permutation of individuals. We might calculate the means of the sub-samples and the difference between the means for a particular permutation. If we repeated this for every permutation we would obtain a theoretical probability distribution for the difference between the sub-samples.

Now consider an experiment where we are interested in the significance of the difference between two random samples, of unequal size and unknown distributions. Our null hypothesis is that the two samples have been randomly chosen from a single population (of unknown distribution). By adopting Fisher's principle of randomisation and following the procedure above we may construct a null distribution against which to test our null hypothesis. If the null hypothesis is true, the difference between sample means that we obtained will not be 'unusual' compared to the null distribution of differences between sample means. It is this method of statistical testing that we adopt in certain places in chapter 6.

Although the principle and its applications have been well known for much of the 20th century, it has never been very widely used. This is not because of any fault in the method itself, but rather because of the immense number of calculations that must be made in order to establish the null distribution.[82] This obstacle has been overcome with the advent of computers. Where there are too many permutations to calculate each, we may obtain a very good approximation to the theoretical null distribution by sampling the permutations with Monte Carlo methods.

Although no assumption is made about the parent distributions, it is essential that the samples are random. In the past, confusion about the value of permutation techniques has arisen because temporal auto-correlation has tampered with this

---

that seeds from two populations would produce the same sized plants under the same conditions. See Edgington (1987, p17-21) for a discussion of the early history of the principle.

[82] Fisher (1935) considered a reference set of 32,768 data permutations, and Edgington (1987, p18) comments that "It is a credit to Fisher's analytical ability that he could determine what proportion of the data permutations provided as large a difference between totals as the obtained value without computing the test statistic for all 32,768 data permutations." In 1968 Bradley described randomisation tests as "among the best tests to be found" (p83), being "superior or equal to their parametric counterparts in the generality of cases" (p83), but concluded that they were "little more than statistical curiosities" (p84) because of the computational problem.

assumption.[83] Ben Santer's PhD thesis (1988) and the papers published following it (Santer and Wigley, 1990; Wigley and Santer, 1990) used randomisation methods without making allowance for temporal autocorrelation, and were justly criticised on those grounds (Santer, pers. comm.). Model simulations with lengths of 10, 20, and 35 years were used, and adjacent single year time slices were permuted. Barnett *et al.* (1998) also recognised the temporal auto-correlation problem when they used overlapping periods to estimate null distributions.[84] Nonetheless, when properly applied the technique is still a valuable one.[85] We do not have the problem of temporal auto-correlation because we sample regions, not periods. If we used every region in a permutation test we would have the parallel problem of spatial auto-correlation, but we only employ a few widely-separated regions in any single test, so spatial auto-correlation is not a problem in this thesis.

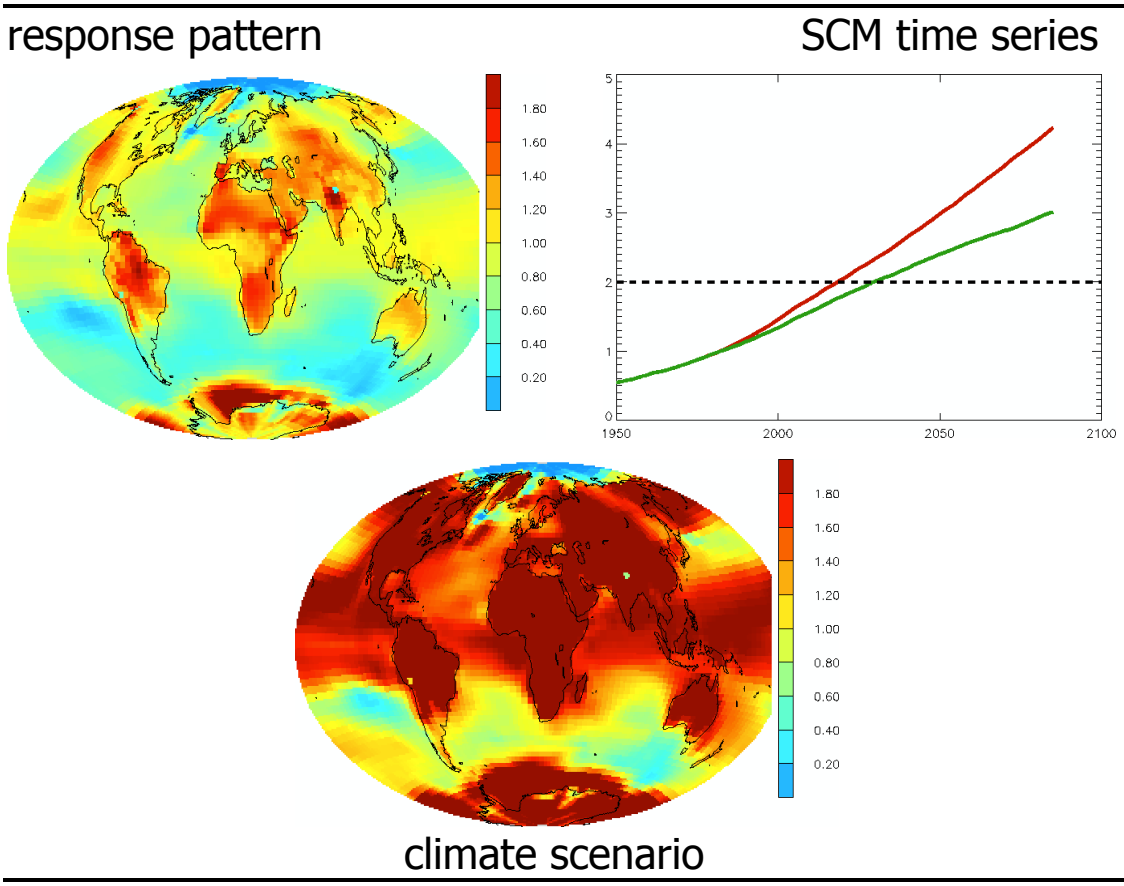
[83] Livezey (1995) concluded that: "serial correlation … is perhaps the most serious difficulty the analyst faces in the application of permutation or Monte Carlo techniques. Without attention to its effects it is completely debilitating to the effectiveness of the procedures." (p171).

[84] Barnett *et al.* (1998) deliberately added a caveat to their estimates: "Obviously these estimates are not independent due to the temporal autocorrelation" (p663). They also felt constrained to use speech marks when referring to a 'significance threshold' (p667).

[85] Livezey (1995) recognised the value of permutation tests in a wide variety of situations: "In these instances the permutation and Monte Carlo techniques described in this chapter are generally effective alternatives for hypothesis testing." (p159) "A knowledge of the material in this chapter, consultation of the listed references when needed, and some creativity will permit the application of permutation techniques in a wide variety of test situations." (p175)

Figure 3.1

The technique of pattern scaling.



1. A response pattern is extracted from a GCM, and each grid-box response is expressed in °C per °C of global annual temperature change (**top left**).
2. A number of SCM simulations (**green**, **red**) are performed for various emissions scenarios, and the global annual temperature change ('the scaler') in those simulations is expressed in °C as a function of time (**top right**).
3. We select a particular time period under a particular emissions scenario, and determine the value of the scaler from (2). For example, the scaler is **2°C** under both the red scenario in 2020 and the green scenario in 2030. We multiply the response pattern from (1) by the scaler from (2) to obtain an estimate of the temperature anomaly pattern (**bottom**) for the selected period and emissions scenario, in °C.

Thus there is a perfect correlation between the GCM response pattern and the climate scenario. The only difference between them lies in the magnitude of the climate scenario.

4. May we use pattern scaling?

4.1 Introduction

In this chapter we examine the possibility of multiplying climate response patterns by a scaler (global-mean annual temperature) to estimate changes under unmodelled forcing scenarios. We use linear regression to construct the spatial patterns. Linear regression has been used before to construct patterns of change on centennial time scales (*e.g.* Hennessy *et al.*, 1999). We build upon the earlier work of Mitchell *et al.* (1999) in evaluating pattern scaling, and considerably extend their work. We make a number of choices at the outset concerning the data that we use to evaluate pattern scaling. In particular, we consider:

- a single model (HadCM2),
- not only temperature but also precipitation,
- not only the scenarios with ensembles (Ga and Gd), but also – where possible – the stabilisation scenarios,
- not annual but seasonal (JJA) changes.

We do not merely consider how accurate scaling is, but we also examine the errors introduced by pattern scaling in some detail, together with the sources of those errors. We consider both the statistical and practical significance of the errors. We consider variations in the accuracy of pattern scaling both over time and spatially.

We begin by discussing the relationships between global-mean climatic change and the scaler, as the basis for all that follows (section 4.2). Non-linear behaviour that arises in that discussion is treated further in section 4.3. We construct signal-to-noise ratios to establish the extent to which there are patterns of response to radiative forcing that may be identified against the background noise of internal variability (section 4.4). These discussions are instructive preliminaries to the main tasks of constructing the response patterns (section 4.5), and using them to estimate climate changes. We examine the errors introduced by pattern scaling in terms of both spatial patterns (section 4.6) and time-series (section 4.7).

We evaluate these errors in the context of the alternatives to pattern scaling by comparing the errors introduced by pattern scaling with those introduced by assuming that the global-mean applies at each grid-box (section 4.8). We also calculate the additional errors that would be introduced by constructing a response pattern from a single simulation rather than an ensemble (section 4.9).

Following the conclusions (section 4.10) we add two appendices referred to in the text.

4.2 Global-mean behaviour

We begin with the time series of global-mean JJA temperature and precipitation[86] (Figure 4.1, top). The global climate changes vary among the four forcing scenarios, so these scenarios provide a good opportunity to examine whether it is reasonable to use pattern scaling to estimate changes under unmodelled forcing scenarios. Because both precipitation and temperature data are available we may examine whether it is legitimate to scale response patterns for a variable other than temperature. However, in order for pattern scaling to be accurate there must be an approximately linear relationship between the variable in question and the scaler. We consider this relationship at the spatial scale of the global-mean by plotting the same set of time-series (from Figure 4.1, top) against the scaler (Figure 4.1 bottom), and find that the relationship is indeed approximately linear.

Since the relationships between the scaler and seasonal climatic responses to radiative forcing are close to linear for the global-mean, it is reasonable to consider the same relationship at the spatial scale of the grid-box. For anomalies in °C or mm/day the global-mean expresses the global average of the relationships between grid-box anomalies and the scaler, so on the basis of Figure 4.1 we conclude that on average, grid-box anomalies of temperature (°C) and precipitation (mm/day) are linearly related to the scaler. However, this only applies to anomalies expressed in °C and mm/day, not when the anomaly is a percentage.

In Figure 4.2 we compare the relationships with the scaler of precipitation anomalies calculated as:

- a global-mean percentage change (left), and
- the percentage change for each grid-box, averaged over the globe (right).

This comparison shows that although the former relationship is linear, the latter is non-linear, which suggests that we will introduce errors by scaling response patterns representing percentage changes in precipitation. We return to this subject in section 4.3.

[86] Precipitation anomalies are expressed both in mm/day and as percentages, because either or both may be useful in a particular region.

For pattern scaling to be relevant, the scaler must relate to the grid-box anomalies in a similar way across the range of forcing scenarios. Figures 4.1 and 4.2 (bottom) suggest that in general, the relationships with the scaler are reasonably similar from one forcing scenario to another. We quantify the similarities by calculating a best-fit line[87] for each variable and simulation (Figure 4.3).

Figure 4.3 shows that the relationships between the scaler and grid-box temperatures in Ga and Gd are very similar; both scenarios warm in boreal summer (JJA) with approximately 85% of the annual warming (Figure 4.3, top right). However, the stabilisation scenarios diverge from Ga and Gd, suggesting that as annual temperature is stabilised, boreal summer (JJA) continues to warm relative to the other seasons. Therefore any JJA response pattern derived from the stabilisation scenarios will be warmer than a JJA response pattern derived from non-stabilisation scenarios. Senior and Mitchell (2000) reached a similar conclusion from a simulation where CO_2 was stabilised after doubling, and made a similar connection with pattern scaling.

The various forcing scenarios exhibit relationships between global-mean precipitation and the scaler that, although similar, are distinct from each other (Figure 4.3, left). It is unlikely that the differences between scenarios are due to internal variability. Firstly, the range between Ga 4 and s550 is 0.25% (or 0.007 mm/day) per degree of annual warming,[88] whereas the range within the Ga ensemble is only 0.04% (or 0.001 mm/day) per degree of annual warming. Secondly, the order between the forcing scenarios corresponds to their rates of increase in radiative forcing: The rate of enhancement of global precipitation (per degree of annual warming) increases from Ga, through Gd and s750, to s550, whereas the rate of increase in radiative forcing decreases from Ga, through Gd and s750, to s550.

How might the rate of increase in radiative forcing have an inverse effect on the rate of enhancement of global precipitation? This is an important question with implications beyond pattern scaling, but because of the computational expense of performing simulations for a variety of forcing scenarios it has not been closely

[87] The best-fit line ($y=ax$) is calculated by least squares regression and is constrained to pass through the origin.

[88] Hulme *et al.* (1998) made similar calculations, but the values calculated here are only broadly comparable with those from Hulme *et al.* (1998), since they examined annual (not seasonal) precipitation, for the period 1900-1996 (not 1860-2099).

examined. However, it has been noted that the magnitude of surface temperature change depends on not just the magnitude, but on the rate of increase in radiative forcing (Kattenberg *et al.*, 1996). We also know that much of the heat resulting from any increase in radiative forcing goes into evaporating surface moisture (Trenberth, 1999). The importance for response patterns of this temperature-moisture connection may be judged by the fact that it is the limited moisture available for evaporative cooling of the land that is partly responsible for the simulated land-sea temperature contrast.[89]

It seems reasonable to suggest that the slower the rate of increase in radiative forcing, the greater will be the warming of the ocean as a proportion of the total warming; therefore evaporation and hydrological intensification will also be greater. We demonstrate that the first element of this hypothesis is true for HadCM2 from the Ga and Gd ensemble means by using the 1.5m air temperature over the oceans as a proxy for sea surface temperatures. The reduced rate of increase of radiative forcing in the Gd compared to the Ga scenario means that the land warms less (per degree of global warming), and the oceans warm more.[90]

In Figure 4.4 we compare global-mean temperature and precipitation (in JJA) with radiative forcing. It is interesting that the amount of warming itself depends on the *rate* as well as the *magnitude* of the increase in radiative forcing (bottom left). However, we are particularly interested in the comparison between temperature and precipitation. We find that whereas the global-mean JJA *temperature* anomaly increases by 11.9% between equivalent periods in Ga and Gd,[91] the *precipitation* anomaly increases by 18.4%.[92] A similar contrast may be found in Figure 4.1 (bottom), where the dependence on the rate of increase in the scaler is – although still slight – greater for precipitation than it is for temperature. Therefore we conclude that not only does the amount of warming depend on the rate of change in radiative forcing, but the hydrological intensification is *even more* dependent.

---

[89] Other important factors include the thermal inertia of the oceans, stronger feedbacks over land, and changes in cloudiness (Kattenberg *et al.*, 1996).

[90] The numbers for Ga [Gd] are as follows: per degree of change in the scaler, the change in annual temperature for 2070-2099 is 1.255°C [1.237°C] over land, and 0.873°C [0.881°C] over the oceans.

[91] We make the comparison for the periods 2070-2099 (Gd) and 2027-2056 (Ga), when the additional radiative forcing in each scenario is 4.52 Wm^{-2}.

[92] The temperature anomalies (°C) are 3.01 (Gd) and 2.69 (Ga). The precipitation anomalies (%) are 5.28 (Gd) and 4.46 (Ga).

This evidence suggests that our hypothesis is correct: the slower the rate of increase in radiative forcing, the greater will be the warming of the ocean as a proportion of the total warming; therefore the evaporation and hydrological intensification will also be greater. On this basis we expect to find that errors are introduced when scaling a pattern from one scenario to estimate precipitation changes in another scenario, because the intensification of the hydrological cycle will depend on the rate of global warming as well as its magnitude.

Furthermore, we have identified from Figure 4.4 that the global *temperature* change is also dependent upon the rate of increase in radiative forcing. Senior and Mitchell (2000) found that in a stabilisation simulation the warming under a doubling of CO_2 increased over *time* due to cloud feedbacks in the southern hemisphere; it appears from Figure 4.4 (bottom left) that the warming under a doubling of CO_2 may also vary *between forcing scenarios*. As Senior and Mitchell (2000) point out with regard to variations in the effective climate sensitivity over time, such variations are likely to be an important source of errors introduced by pattern scaling. However, these particular errors cannot be investigated without carrying out a full comparison of the values of the scaler simulated by GCMs and simple climate models, which is beyond our scope here.[93]

4.3 Non-linear precipitation response

In section 4.2 we considered global-means, but we noted in passing that the global average of grid-box precipitation percentage anomalies is a different statistic. We plotted this statistic against time and global annual temperature (Figure 4.2, right), and we calculated corresponding best-fit lines (Figure 4.3, bottom right). We have already noted the non-linearity of this statistic's relationship with the scaler (Figure 4.2, bottom right), which is characteristic of each of the forcing scenarios and appears to be consistent between them. It is worth noting that the non-linearity makes it inevitable that the best-fit lines will differ between forcing scenarios (Figure 4.3, bottom right),[94] so the differences between best-fit lines do not imply that the relationship between precipitation and the scaler is different among the various forcing scenarios.[95]

---

[93] The interested reader is referred to Raper and Cubasch (1996) and Raper *et al.* (2001).

[94] The non-linearity means that the gradients are chiefly dependent on the amount of radiative forcing in the scenario, rather than on the relationship between precipitation and the scaler.

[95] However, if we calculate a best-fit line for each grid-box in order to generate a response pattern, the differences in gradient displayed in Figure 4.3 will be manifested in the derivation of different response patterns from different forcing scenarios.

Why is the global-mean linear when the grid-boxes are non-linear (Figure 4.2, bottom)? We might express the same problem in a different way. Why are the global-mean mm/day anomalies linear (Figure 4.1, bottom centre) when the global-mean percentage anomalies are non-linear (Figure 4.2, bottom right)? It is evident that we cannot offer an explanation while assuming that the precipitation at each grid-box is linearly related to the scaler.[96] Therefore the non-linearity implies that – for at least some of the grid-boxes – the precipitation response changes over time in relation to the scaler, and therefore in relation to radiative forcing. It is these changes that make the percentage anomalies non-linear. However, the non-linear changes at the grid-box level must balance over the globe as a whole,[97] because the global-mean changes remain linear.

We may represent the energy that drives the global hydrological cycle in terms of global temperature, which is linearly related to radiative forcing, so it is not surprising that global-mean precipitation is also linearly related to radiative forcing (Figure 4.4). However, the regional manifestations of the hydrological cycle are often characterised by discontinuities and thresholds, so neither would it be surprising if a continuous global-scale increase in energy were to cause discontinuous changes in precipitation in particular regions. The rate of global precipitation change (as a function of the scaler) would not vary, but the rate of precipitation change in individual regions would vary. We might hypothesise that the initial response would be to increase precipitation rates in regions of high precipitation, and that the later response would be to increase rates in regions of previously low precipitation. In this case the percentage changes in precipitation in the regions affected would initially be low, but then increase much more sharply. Averaged over the globe, these hypothesised responses would give the non-linear relationship described.

Is our hypothesis plausible? If surface temperature is increased due to increased radiative forcing, among the consequences will be an increase in the water-holding capacity of the atmosphere and an increase in evaporation. These hydrological changes should increase the actual moisture content of the atmosphere and

[96] If the precipitation at each grid-box (as a percentage anomaly) were linearly related to the scaler then the global average of the grid-box percentage anomalies would also be linearly related to the scaler.

[97] *i.e.* a disproportionate change in one region must be balanced by a disproportionate change in one or more other regions.

increase global precipitation. These expectations are consistent with observations: the former has been widely observed in recent decades (*e.g.* Gaffen *et al.*, 1991), and the Second Assessment Report of the IPCC (Nicholls *et al.*, 1996) reported a small positive trend in precipitation over the 20th century. However, the *regional* manifestations of these global trends are highly variable. On the one hand, increased atmospheric moisture suggests more intense rainfall events, reasoning that is consistent with observations (*e.g.* Karl and Knight, 1998). On the other hand, naturally-occurring droughts are likely to be exacerbated by enhanced potential evapotranspiration (*e.g.* Gregory *et al.*, 1997). This reasoning supports the first half of our hypothesis, that the initial increases in precipitation might be located in regions of high precipitation.

The second half of our hypothesis, that the later response might be located in regions of lower precipitation, is suggestive of dynamical changes bringing precipitation to different regions, rather than intensity changes within the same regions. Trenberth (1998) notes that dipole-like changes in precipitation patterns should occur if mid-latitude storm tracks shift meridionally (*e.g.* Hurrell, 1995), or if there are changes in the behaviour of ENSO. It is uncertain why a change in emphasis from intensity changes to dynamical changes should take place.

We test our hypothesis using the Ga ensemble mean. By calculating the precipitation anomaly (mm/day) for each grid-box for the periods 2010-2039 ('2020s') and 2070-2099 ('2080s') and dividing them by the global-mean temperature anomalies for these periods, we obtain response patterns for the early and late 21st century.[98] We compare the difference between the response patterns with the precipitation rates in the control simulation. We find that over the 21st century the response in the wettest grid-boxes (in the control) diminishes, but the response in the drier grid-boxes (in the control) is enhanced.[99] Moreover, the diminution and enhancement almost balance.[100]

---

[98] This piece of analysis anticipates our full description of the construction of response patterns in section 4.5 (particularly Figure 4.17), so we do not present detailed reasoning or figures at this point.

[99] There are 324 grid-boxes (4.8%) with at least 8.0 mm/day in the control simulation, and 3501 grid-boxes (51.4%) with less than 2.0 mm/day. Between the 2020s and the 2080s the change in response to radiative forcing in the wettest grid-boxes averages –0.217 mm/day (per degree of global warming), and in the drier grid-boxes averages 0.017 mm/day.

[100] The change in the wettest grid-boxes represents a change of –0.0103 mm/day (per degree of global warming) in global-mean precipitation, whereas the change in the drier grid-boxes represents a change of 0.0087mm/day.

If our analysis is correct, then difficulties will be introduced in estimating precipitation[101] by scaling from one point in a forcing scenario to another point in the same scenario. Any response pattern derived from an early part of a forcing scenario might be different from one derived from a later part. Consequently the application of the early response pattern to a later part of the forcing scenario will introduce errors.

It is conceivable that if we derive a response pattern from a set of time series rather than a single period,[102] the pattern may even distort the real changes, by displaying weakened versions of precipitation responses that are non-existent for part of the scenario, and prominent features for another part. In other words, estimates constructed by scaling the pattern are likely to err in one direction early in a scenario, and err in the other direction later in the scenario. Since the majority of the radiative forcing takes place after the 20th century, we expect the errors to be greatest in the early and late parts of the 21st century, and smallest in the mid-21st century. However, the cost to accuracy of constructing a response pattern using a single period at the end of the forcing scenario rather than using a time series would be even larger, because the pattern would represent a single extreme period, rather than an average among the periods.

4.4 Signal-to-noise ratios

We have examined the relationship between climate changes and the scaler. We may take our analysis a step further by examining the climate responses in the context of internal variability, in the form of signal-to-noise ratios. By this method we may evaluate the likelihood of being able to extract from the perturbed simulations a pattern of response to increased radiative forcing that is distinct from internal variability.

4.4a Spatial heterogeneity of response

We follow the approach adopted by Mitchell *et al.* (1999), who calculated signal-to-noise ratios for temperature and precipitation response patterns.[103] Their ratio

[101] It is only for precipitation that we have examined percentage anomalies. It is possible that percentage anomalies of temperature might show similar features.

[102] *i.e.* We use one time series per grid-box, rather than one period per grid-box, and calculate a best-fit line over time rather than simply dividing the period mean by the scaler.

[103] They describe their method in their Appendix; the results are plotted in their Figures 1c (for temperature) and 12 (for precipitation).

provides a single global measure of the extent to which there is a spatially varying response to increased radiative forcing. Under any such response, as radiative forcing increases the grid-box anomalies will increasingly diverge from the global-mean anomaly; thus the signal – the squared differences between the grid-box and global-mean anomalies – increases. As radiative forcing increases and the response pattern increasingly dominates internal variability, so the signal-to-noise ratio increases.

4.4b Method of calculation

Mitchell *et al.* (1999) chose to use ten-year means in order to "filter out major oscillations such as those due to El Niño / Southern Oscillation", and in order not to "smooth out any secular change in the [response] pattern" (p555). Mitchell *et al.* (1999) also chose to calculate anomalies relative to the 1961-1990 mean "because it is changes from the present day that are of interest to the impacts community, rather than changes from the control simulation which represents pre-industrial climate" (p555). However, the *climatologist's* problem is to identify the response pattern as robustly as possible. A strength of the pattern scaling technique is that having identified that response pattern and scaled it, the problem of presenting the estimated changes with reference to the desired base period is trivial. Therefore we believe that *if* we can identify a response pattern more robustly than Mitchell *et al.* (1999) were able to do, then we should do so.

Our initial objective is to identify the method of calculating anomalies that gives the highest signal-to-noise ratios, so that we may identify response patterns as clearly as possible. Therefore at this point we calculate annual signal-to-noise ratios for two forcing scenarios (Ga and Gd) for the last part of the 21st century. Except where we state otherwise, we calculate the signal (Equation 4.2.1), noise (Equation 4.2.2), and signal-to-noise ratio (Equation 4.2.3) in precisely the same way as Mitchell *et al.* (1999), using the same method of weighting grid-boxes according to the spatial area they represent (Equation 4.2.4). We explain in Appendix 4.1 why the results we obtained are different from those plotted in Mitchell *et al.* (1999). Here we examine the dependence of the signal-to-noise ratios we obtain on the choices we make in anomalising.

Mitchell *et al.* (1999) reported that the maximum signal-to-noise ratio for precipitation in the Gd ensemble (in the 2090-2099 period)[104] "barely achieves a value of 2 to 3" (p572, *cf.* their Figure 12). On this basis they suggest that "many of

[104] The period mean was anomalised against the 1960-1989 period in the Ga ensemble.

the regional changes in precipitation may not be significant" (p576). We carried out a number of similar calculations (Figure 4.5, top centre):

(a) We repeated their calculation and obtained a value between four and five (left-hand green bar). [105]

(b) We increased the period length from 10 years to 30 years, and considered the signal-to-noise ratio for the 2070-2099 period.[106] Despite the mean forcing over that period being smaller than for the 2090-2099 period, the precipitation signal-to-noise ratio increased by about 15% in Gd to five and a half (central green bar). The ratio increased because the decrease in noise due to a longer period (25%)[107] exceeded the decrease in signal (11%) from using a period with smaller averaging forcing.

(c) Therefore we continued to use the 30-year period length, but this time we anomalised the 2070-2099 period against the first 240 years of the control.[108] We found that the Gd precipitation signal-to-noise ratio increased by a further 80% to almost ten (right-hand green bar), because not only did the noise decrease further (by a further 27%),[109] but the signal also increased (by a further 29%).

We conducted similar comparisons for temperature (Figure 4.5, top left), and for unweighted temperature and precipitation (Figure 4.5, bottom), and found a similar dependence of the signal-to-noise ratio on the period length and on the base from which anomalies are constructed. These results indicate that we may

---

[105] Appendix 1 examines the differences between the Mitchell *et al.* (1999) and our own results.

[106] We still anomalised the period mean against the 1960-1989 period in the Ga ensemble.

[107] These bracketed percentages refer to the changes in weighted precipitation (mm/day) in the Gd ensemble, but the same argument applies to all the variables, weighted or unweighted, and to both ensembles.

[108] Because Mitchell *et al.* (1999) anomalised against the period 1960-1989, they constructed their control anomalies using bases within 140-year periods, so model drift did not influence the noise they calculated. Here we constructed control anomalies for the final 1160 years using the first 240 years as our base (see footnote in Equations 4.2), so before anomalising we detrended the full 1400 years in each grid-box by removing the gradient from the best-fit line (by least squares regression) through the time series. Thus we removed any model drift before calculating the anomalies on which our estimate of the noise (Equations 4.2) is based.

[109] In other words, the noise from the Mitchell *et al.* (1999) method (using the anomalies in the final 30-year periods (*relative to* the initial 30-year periods) in 140-year segments of the 1400-year control), is reduced by 27% when we use our method of calculating noise (using the anomalies in 30-year periods (separated with 10-year gaps) in the final 1160 years of the 1400-year control, *relative to* the first 240 years). This is probably due to a reduction in the noise from *the base period itself* when we move from a 30-year base to a 240-year base.

improve on the signal-to-noise ratios calculated by Mitchell *et al.* (1999) by using 30-year averaging and anomalising against the first 240 years of the control. It was on the basis of the weak signal-to-noise ratios they found for precipitation that Mitchell *et al.* (1999) argued that the regional precipitation responses were "at best marginally significant" (p574). Therefore we conclude that the larger ratios we have found suggest that many of the regional changes in precipitation may in fact be significant after all, and that it is reasonable to attempt to scale them.

4.4c Examination of spatial heterogeneity

We now apply our modified method to time series of temperature and precipitation for the full suite of forcing scenarios. We display the results in Figures 4.6 and 4.7 for individual simulations and ensemble means respectively. We may compare the seasonal (JJA) and annual signal-to-noise ratios for the 2080s by comparing the final points plotted in Figure 4.7 (JJA) with the right-hand bars in each panel of Figure 4.5 (annual). The differences between them are consistent between variables: in each case the seasonal signal-to-noise ratio is slightly less than its annual equivalent, as one might expect. The most critical decrease is in Gd precipitation (%), where the 2080s ratio decreases from nine (annual) to six (JJA). Since in the most critical case the signal-to-noise ratio at the end of the simulation is as high as six, we expect to be able to identify seasonal precipitation response patterns in a similar manner to the identification of annual temperature response patterns by Mitchell *et al.* (1999).

The signal-to-noise ratios of the individual simulations (Figure 4.6) are much smaller than those of the ensemble means (Figure 4.7), but they permit a fair comparison between the ensembles and the stabilisation scenarios. It is particularly noteworthy that the signal-to-noise ratios are very similar between forcing scenarios at similar levels of global warming: the inter-scenario differences appear to be no larger than those due to internal variability (Figure 4.6 bottom). In particular, the signal-to-noise ratios of the unstabilised scenarios (Ga and Gd) closely approximate to those of the stabilisation scenarios (s550 and s750) (Figure 4.6 bottom).

In section 4.2 we found that the relationship between seasonal global-mean anomalies and global annual temperatures was approximately linear (Figure 4.1 bottom). That is not the case here, where the variables exhibit varying degrees of non-linearity between seasonal signal-to-noise ratios and global annual temperatures (Figure 4.6 bottom). The most obviously non-linear are the precipitation (%) anomalies, but a careful visual inspection shows that the

precipitation (mm/day) anomalies are also non-linear; it is possible that the temperature anomalies are also non-linear, but if so the departure from linearity is very small.

The differing degrees of non-linearity among the variables in Figure 4.6 complicate any comparison of them, but it is reasonably accurate to state that at the same scaler value, the temperature signal-to-noise ratios are between 25% and 50% greater than the precipitation (mm/day) signal-to-noise ratios. In the lower part of the scaler range, the precipitation signal-to-noise ratio for mm/day anomalies exceeds that for percentage anomalies, but in the upper part of the range the percentage ratio increases more swiftly, so that the percentage and mm/day ratios are nearly equal by the end of the Ga scenario. The non-linear increases in precipitation signal-to-noise ratios are consistent with our hypothesis (section 4.3) that initial increases in precipitation rates in regions of high precipitation are followed by increased rates in regions of low precipitation.

4.4d Grid-box responses

The signal-to-noise ratios described above provide us with a single global measure of the extent to which there is a spatially heterogenous response to radiative forcing. We may also treat the grid-boxes individually and calculate for each grid-box the extent to which a response to radiative forcing can be identified from the noise of internal variability. If we plot a single signal-to-noise ratio for each grid-box, we may produce a pattern indicating the location of the regions where the responses to radiative forcing are most clearly distinguished from noise. To construct such a ratio (Equation 4.3.3) we use an ensemble mean as the signal (Equation 4.3.1), and construct the noise (Equation 4.3.2) from the control standard deviation for that grid-box.

The method of construction of these spatially-varying signal-to-noise ratios means that the ratios are identical whether based on percentage or mm/day anomalies, so we only plot the ratios for temperature (Figure 4.8) and precipitation (mm/day, Figure 4.9). The need for ensembles restricts us to the Ga (top) and Gd (bottom) forcing scenarios. We are also restricted to a single ratio for each grid-box on a single plot,[110] but we summarise the early and late 20th century in two separate plots using ratios based on the 2020s (left) and 2080s (right). The magnitudes of some of these ratios – when globally averaged – are much larger than the 'pattern'

[110] If we were examining a particular region, it might be of greater interest to restrict the number of grid-boxes, and to plot a time series for each grid-box of the ratios of overlapping 30-year periods.

ratios calculated above because of the method of construction, but our concern here is with the spatial differences rather than the magnitude.[111]

Zonal differences dominate the temperature signal-to-noise ratios (Figure 4.8); the tropics and parts of the subtropics have the largest ratios.[112] As the radiative forcing increases through the forcing scenarios, so the signal-to-noise ratios increase across most of the globe. The precipitation signal-to-noise ratios (Figure 4.9) do not have the same zonal structure as the temperature ratios; they vary within latitudinal bands to the extent that in many places the polarity of the response flips from one side of a continent or ocean to the other. The similarity between these signal-to-noise patterns and the response patterns (introduced below) indicates that for precipitation the pattern of the signal is a more important influence on the signal-to-noise ratios than the pattern of the noise. As with temperature, the increase in radiative forcing through the forcing scenarios increases the magnitudes of the signal-to-noise ratios across most of the globe.

4.5 Response patterns

Our overall objective is to estimate climate changes by linearly scaling a response pattern. We have shown that although grid-box temperature and precipitation are linearly related to the scaler, for individual grid-boxes the precipitation response may be non-linearly related to the scaler. We have also identified the regions where those responses are most clearly identifiable from the noise of internal variability. In this section we construct response patterns.

4.5a Initial patterns

---

[111] Figure 4.8 displays temperature signal-to-noise ratios greater than 90 for particular grid-boxes. These are genuine results. Annual ratios are similarly large. In the final 30-year period of the 21st century the annual anomaly in some grid-boxes (relative to the first 240 years of the control) amounts to 10°C (for confirmation see Mitchell *et al.*, 1998, Figure 11b). The standard deviation of decadal means of annual temperature estimated from the control simulation is mostly 0.2°C or less (for confirmation see Mitchell *et al.*, 1999, Figure 5a). Thus the literature implies ratios of 50; these ratios increase further when we reduce the noise to the level of a *30-year* period, and when we allow for the use of an *ensemble*.

[112] There are two possible reasons for the zonal structure. One possible reason is that it results from the noise; it is plausible that the noise might be larger in the extra-tropics than in the tropics, and the signal is certainly dominated by the land-sea contrast (*e.g.* Mitchell and Johns, 1997). The other possible reason is that the noise also has a land-sea contrast that cancels out the land-sea contrast in the signal.

Initially we construct response patterns using the entire length of run in the Ga ensemble, by applying least squares regression to each grid-box in turn. Following the results of our signal-to-noise analysis (section 4.4) we use 30-year period means and anomalise against the first 240 years of the control. For each grid-box we calculate a constant rate of change (a) in the variable (y) per degree of global annual warming (x), using least squares regression. Rather than fitting to *absolute* values of x and y, which would require an equation with two constants ($y=ax+b$), we fit to the anomalies of x and y, for which we do not need an intercept ($y=ax$). Our response pattern consists of the value of a in each box.

We construct response patterns for temperature[113] and precipitation (Figure 4.10). These response patterns are similar to the patterns of change considered in the IPCC Second Assessment Report (Kattenberg *et al.*, 1996), and are broadly consistent with other coupled models. The warming over land is greater than over the ocean. Warming is minimised in the North Atlantic and Southern Ocean (except in the sea-ice areas) due to deep convection. The differences between the response patterns for mm/day and percentage precipitation anomalies illustrate the value of using both in any assessment of the response of precipitation to radiative forcing. Most of the global increase in precipitation occurs in the tropics, but not uniformly; in some tropical regions there are large increases in precipitation while other regions become drier. There are also substantial increases in precipitation in mid-latitudes, particularly in the southern hemisphere (for which JJA is winter).

4.5b Sources of differences between temperature patterns

We recognise that if the response to radiative forcing changes during the course of a forcing scenario, a response pattern constructed using the entire length of run will not represent the response for particular periods of time as accurately as it might otherwise do. Therefore we examine whether there are changes in the response pattern during the 21st century. We extract response patterns for two periods in the 21st century from the Ga ensemble and calculate the statistical significance of the differences between them (Equations 4.4). We display the results in Figure 4.11.[114]

---

[113] Mitchell *et al.* (1999) plotted (their Figure 2) a similar pattern for annual (not seasonal) temperature using the period 1990-2099 (not 1860-2099).

[114] We plot the 2020s pattern (top left). We also plot the differences between the patterns (top right), the statistical significance of the differences (bottom left), and a measure that we call the 'strengthening of the pattern'. For details of this final measure see Table 4.1.

Figure 4.11 shows that the differences (top right) between the patterns from the 2020s (top left)[115] and the 2080s are statistically significant in nearly half the grid-boxes (bottom left). The largest changes during the 21st century are a relative warming of the tropics at the expense of extratropics. As a measure of the 'strengthening of the pattern' we express the change in response during the 21st century as a fraction of the 2020s pattern, but plot only those grid-boxes where the change is statistically significant (bottom right). This pattern represents the statistically significant strengthening (and weakening) of the response pattern during the 21st century. (Table 4.1 gives further details of the meaning of this pattern.) These changes in response during the course of the forcing scenario imply that estimating changes by scaling response patterns will introduce errors. The summary statistics in Figure 4.11 (caption) present two ways of sub-dividing these changes:

i. The statistically significant changes may be sub-divided into a *reversal* of the direction of the response in 6.9% of the grid-boxes, a *diminution* of the pattern's strength in 18.3%, and an *enhancement* in 17.6%.[116]

ii. The statistically significant changes may be sub-divided into those that are small relative to the response itself, and those that are large. This is one measure of the *practical* significance of the changes: a statistically significant change may be practically insignificant if it is only a small change. Here we arbitrarily select a change of a fifth as the boundary between small and large changes: under this definition 32.2% of the grid-boxes show changes that are both statistically and practically significant.

A further potential source of errors lies in the differences in response between one forcing scenario and another. However, since there are changes in response pattern *within* a single forcing scenario, we cannot compare response patterns derived from the *entire lengths* of different forcing scenarios without contaminating our comparison. Therefore we based our comparison on 30-year periods within the forcing scenarios when the scaler is approximately the same:

[115] Note that the range of the scales is 2°C in *both* of the top plots.

[116] The changes expressed here are to the spatial pattern of the seasonal temperature changes accompanying a change in annual global-mean temperature. Thus a reversal in direction refers to a change in the response from being smaller than the scaler to being larger (or *vice versa*). A diminution [enhancement] of the pattern's strength refers to a change in the temperature response of the individual grid-box such that it becomes closer to [further from] the value of the scaler.

- To compare s550 and Ga we selected four independent periods from s550 and four equivalent periods from the four Ga members.[117]
- To compare s750 and Ga we repeated this procedure, selecting different periods.
- To compare Gd and Ga we selected a period from Gd (for all four realisations) and an equivalent period from Ga (for all four realisations).

For each comparison we constructed a single response pattern from the four periods from each forcing scenario, and subtracted the Ga pattern from the other pattern (Figure 4.12).

Thus each plot in Figure 4.12 is a comparison of two patterns derived from responses with the same scaler, but the scalers are not the same between the various plots. Given the changes in response that occur within a forcing scenario, the similarity between the three plots[118] is therefore all the more remarkable (*e.g.* Africa). The large-scale effect of stabilising global annual temperature, no matter which forcing scenario is used, is to increase the rate of warming over the southern oceans relative to the northern continents. In other words, the general effect of stabilisation is to smooth out the regional contrasts in the Ga response pattern (Figure 4.10 bottom), as recognised by Stouffer *et al.* (1989) and specifically for HadCM2 (Mitchell *et al.*, 2000; Senior and Mitchell, 2000). However, there are regions where the stabilisation has the opposite effect and further enhances the already enhanced warming (*e.g.* Amazonia and south-eastern Africa), or further diminishes the already below-average warming (*e.g.* the Kuro Siwo area of the Pacific).

We cannot calculate the statistical significance of the differences between the response patterns because ensembles are not available for the stabilisation scenarios (see Appendix 4.2 for details). However, we may express the statistical significance of the differences between the Ga and Gd response patterns in a similar manner to Figure 4.11 (Figure 4.13). It is noteworthy that the differences between the Ga and Gd response patterns are statistically significant in only 13.6%, and the changes are greater than a fifth in only 8.0%, of the grid-boxes. These proportions are much smaller than the 42.8% and 32.2% (respectively) obtained for the differences between two periods from Ga (Figure 4.11). This may

[117] We have only one simulation for the stabilisation forcing scenario s550, so a pattern based on a single period would be heavily dependent on internal variability.

[118] In the top two plots, 88.3% of the grid-boxes have the same sign. In the left-hand plots, 78.6% of the grid-boxes have the same sign.

also be confirmed by a visual inspection: compare the set (i) magnitudes displayed in Figure 4.11 (top right) with the set (ii) magnitudes in Figure 4.13 (top right).

We have already identified two sets of changes in the temperature response pattern:

i. from one period to another within a single forcing scenario, and
ii. between one forcing scenario and another at periods with equivalent global annual temperatures.

However, it is our intention to apply a single response pattern from a single forcing scenario to different periods within different forcing scenarios. Therefore we complete our examination of the temperature response patterns by comparing the response patterns drawn from the entire Ga and Gd scenarios (Figure 4.14).

When we compare two patterns drawn from the full lengths of two different forcing scenarios we expect both sets of changes (i. and ii. above) to contribute to the differences between the patterns. A visual inspection of Figure 4.14 (top right) reveals that the differences between the patterns are an amalgam of the differences displayed in Figures 4.11 (*e.g.* Africa), and 4.13 (*e.g.* Europe). As we noted above, the changes of set (ii) are smaller than those of set (i), and a cursory analysis suggests that the changes of set (ii) are indeed slightly less influential here than those of set (i).[119] The number of grid-boxes with significant differences (22.6%) lies between the examples already given, being less than for set (i) (Figure 4.11, 42.8%), but more than for set (ii) (Figure 4.13, 13.6%). The proportion of grid-boxes with significant changes greater than a fifth is only a tenth (10.6%, Figure 4.14), which suggests that scaling a pattern from one scenario to estimate temperature changes in another scenario may be accurate in most cases.

4.5c Sources of differences between precipitation patterns

We repeat our comparison of response patterns for precipitation, both as mm/day anomalies (Figure 4.15) and as percentage anomalies (Figure 4.16). The number of grid-boxes with significant differences between responses under Ga and Gd is remarkably consistent between temperature and precipitation: in each case the proportion is in the range 22.5% – 22.7% (Figures 4.14, 4.15, 4.16). However, more of the statistically significant precipitation changes (about 18%) are greater than a fifth than was the case for temperature. The differences between the precipitation

[119] The signs of each grid-box in Figure 4.14 (top right) were compared with those in Figures 4.11 (top right) and Figure 4.13 (top right). In the case of set (i) (Figure 4.11), 71.0% of boxes had the same sign in Figure 4.14. For set (ii) (Figure 4.13), the statistic was 62.4%.

responses under Ga and Gd are of a similar variety to those for temperature discussed above. Some regions have a response to radiative forcing that is diminished under a weaker rate of radiative forcing (*e.g.* the Caribbean dries less, the Atacamba Desert moistens less). In other regions the response is enhanced under a weaker rate of radiative forcing (*e.g.* the Arabian Peninsular dries more, the Sahel moistens more).

To examine the differences between precipitation response patterns drawn from different periods in the same forcing scenario, and from different forcing scenarios, we calculated precipitation response patterns similar to the temperature response patterns.[120] We repeated our comparison of different periods in Ga (Figure 4.17), and between periods in Ga and Gd with equivalent changes in the scaler (Figure 4.18). Since precipitation is not the same variable as the scaler (as was the case in the temperature patterns discussed above), the meaning of the 'strengthening of the pattern' plots (bottom right) is slightly different for precipitation compared to temperature.[121]

It is most interesting that during the 21st century there are fewer significant changes in the precipitation response pattern (34.0%, Figure 4.17) than in the temperature response pattern (42.8%, Figure 4.11). Similarly, there are fewer significant changes between the Ga and Gd patterns for precipitation (8.8%, Figure 4.18) than for temperature (13.6%, Figure 4.13). However, where significant changes do occur they are much larger in proportion to the original pattern. Hence the proportions of grid-boxes where the changes are both statistically significant and greater than a fifth are almost identical for temperature and precipitation, both for 21st century changes in Ga (32.2% and 33.0% respectively) and for changes between Ga and Gd (both 8.0%).

4.6 Patterns of estimation error

[120] Specifically, we calculated response patterns from the 30-year periods in Ga centred on 2025, 2045, and 2085, and from 2073 in Gd.

[121] Among the grid-box changes in precipitation are both moistening (positive) and drying (negative) responses to radiative forcing. The seasonal (JJA) precipitation changes average 0.05 mm/day (or 2%) per degree of change in the scaler (Figure 4.1), but this global-mean does not translate well into grid-box changes, which vary much more widely. Therefore we identify any reversals from moistening to drying (or *vice versa*), and any diminution or enhancement of the moistening (or drying). A reversal corresponds to ratio<-1 in Figures 4.15-4.18 (bottom right), and a diminution [enhancement] corresponds to –1<ratio<0 [0<ratio].

We have constructed response patterns for both temperature and precipitation, and examined the variations in those patterns within and between forcing scenarios. Now we estimate regional temperature and precipitation by scaling the response patterns by global annual temperature ('the scaler'), and we calculate errors from pattern scaling by comparing the estimates with the model values. In this section we examine the spatial patterns of error, and in the next (4.7) we examine time-series of error.

We may express our statistical model as follows: we estimate the grid-box temperature or precipitation anomaly (y) by multiplying the response (a, as calculated in section 4.5) by the scaler (x) through the relationship $y=ax+\beta+\varepsilon$, where the error in estimation is divided into the systematic error (β) and noise (ε). By this method grid-box anomalies may be estimated for any period in any forcing scenario, by scaling a GCM response pattern by the annual global-mean temperature from a simple climate model. The magnitude of ε will depend on two factors in particular:

i. The variations in the spatial pattern of response over the course of the forcing scenario from which the response pattern is extracted.

ii. The variations in the spatial pattern of response between the forcing scenarios from which it is extracted, and for which the estimate is made.

We begin by extracting a response pattern from an entire forcing scenario (Ga, Figure 4.10), and scaling it to estimate the regional temperature and precipitation for particular periods under the same forcing scenario. By using an ensemble we reduce ε to a minimum, and by estimating for the same forcing scenario from which the response pattern was drawn we eliminate component (ii) of β. The remaining error is ε plus component (i) of β, which we plot for temperature (Figure 4.19, top) in a comparable manner to Mitchell *et al.* (1999).[122] We isolate component (i) of β from ε by calculating the statistical significance of the error (Figure 4.19, bottom). Ours is a slightly different method from that of Mitchell *et al.* (1999, their Figure 5), who simply divided the error by the standard deviation in the control. We calculated the theoretical sampling distribution of the error to permit us to describe the statistical significance of the errors in terms of the

[122] See Figure 4 in Mitchell *et al.* (1999). Although they estimated annual temperatures for decadal means, the magnitudes and patterns of error that we obtain are comparable to theirs.

number of standard deviations from 'no error' that the errors represent (Equations 4.5).[123]

We described in section 4.5 (Figure 4.11) how the temperature response pattern changes during the 21st century. When we calculate a response pattern based on the entire length of simulation under the scenario,[124] the pattern may be thought of as an average between the responses at the beginning and end of the 21st century. Therefore when – as here – we scale such a response pattern to estimate regional temperatures during the 21st century, the estimate errs in one direction early in the century, and errs in the other direction late in the century. This explains the anti-correlation between the top two plots in Figure 4.19,[125] and the correlation between those plots and the plot of the difference between response patterns in the 2020s and 2080s (Figure 4.11, top right).[126] In section 4.5 we found that the response patterns for the 2020s and 2080s were significantly different from each other (at the 2□ level) for over 40% of the grid-boxes (Figure 4.11). Since here we are subtracting the 2020s and 2080s temperature patterns from a scaled response pattern that averages between those temperature patterns, the proportion of grid-boxes with significant estimation errors is much smaller than 40% (Figure 4.19, bottom).

In the method described above we followed Mitchell *et al.* (1999) in estimating the level of internal variability under perturbation ($\square_M$ in Equations 4.5) from the control. However, if there is a response to radiative forcing not only in the mean but also in higher order statistics, it may be more appropriate to estimate $\square_M$ from the perturbed ensemble itself. Therefore in Figure 4.20 (left) we plot two versions of the statistical significance of the error in estimating 2080s temperature anomalies: that obtained by calculating $\square_M$ from the control (top left),[127] and from

[123] Here we calculate $\square_M$ (Equations 4.5) from the control. We discuss below the calculation of $\square_M$ from the ensemble.

[124] It might be argued that the response pattern would be defined more robustly using simply the 2070-2099 period. We do not believe that to be true. Firstly, if there are any non-linearities in the response over time (as we know there are), using solely the 2070-2099 period would optimise the response pattern to *that period* in *that forcing scenario*. This would make the pattern less accurate for other periods in the same forcing scenario, and for other forcing scenarios. Secondly, by using data for the full 240 years we are greatly increasing the quantity of data that we are using to define the response pattern, which should make the pattern less dependent upon internal variability and therefore *more* robust than merely using the 2070-2099 period.

[125] The correlation coefficient between the top two plots in Figure 4.19 is –0.86.

[126] The correlation coefficient between Figures 4.19 (top left) and 4.11 (top right) is 0.97.

[127] This plot is a duplicate of Figure 4.19 (bottom right).

the Ga ensemble (bottom left). The effect of calculating $\square_M$ from the ensemble is to increase the proportion of boxes with significant errors by about 4% of the total, implying that the estimate of internal variability from the ensemble is smaller than that from the control.[128] There are two possible causes for this that we discuss further below:

(a) part of the climatic response to radiative forcing is a decrease in temperature variability (at 30-year time scales);
(b) the intra-ensemble variance is an under-estimate of temperature variability.

In Figure 4.20 (right) we express the error as a fraction of the estimate and mask this proportion by the error significance. This statistic has a similar meaning to the 'strengthening of the pattern' statistic calculated in section 4.5. The proportion of grid-boxes where the error is both statistically significant and more than a fifth of the estimate itself is very small: less than 6%. If we accept the arbitrary boundary of a fifth as determining – in some sense – the practical significance of the error, we may conclude that scaling the Ga pattern gives an accurate estimate of the temperature anomalies in Ga in the 2080s, for the errors are both statistically and practically significant in fewer than 6% of the grid-boxes.

We repeat this analysis for precipitation anomalies (mm/day, Figure 4.21).[129] We find a similar anti-correlation between the error patterns for precipitation in the early and late 21st century (not plotted) to that we found for temperature (Figure 4.19). The proportion of grid-boxes with statistically significant errors (Figure 4.21) is slightly smaller – by 3-5% – than for temperature (Figure 4.20). However, the proportion of grid-boxes with both statistically and practically significant errors[130] is 21.3% (Figure 4.21), which is much higher than for temperature. As with temperature, the effect of employing the Ga ensemble to estimate internal variability rather than the control simulation is to reduce internal variability and hence increase the proportion of grid-boxes with statistically significant errors, in this case by about 2% (Figure 4.21).

[128] It is not true to state that the estimate of internal variability is smaller from the ensemble than from the control because of the different sample sizes, for these are population estimates that have been adjusted for sample size bias.

[129] The equations to calculate the significance of the errors are identical for mm/day and percentage anomalies, so we only calculate for mm/day.

[130] The practical significance is defined by our arbitrary measure of a fifth of the estimate.

We now apply the Ga response patterns to the task of estimating temperature (Figure 4.22) and precipitation (Figure 4.23) patterns under the Gd forcing scenario. A visual inspection of the 2080s anomalies from the Gd ensemble (top left) and the estimate from scaling the Ga pattern (top right) gives a strong indication of the accuracy of pattern scaling. The statistically significant errors (bottom left) are confined to a small proportion of the grid-boxes (13.8% for temperature and 10.7% for precipitation). A still smaller proportion have errors that are both statistically significant and larger than a fifth of the modelled anomaly (1.8% for temperature and 8.9% for precipitation).

4.7 Time-series of estimation errors

4.7a Estimates of internal variability

We begin with a slight but necessary diversion: we consider whether to use estimates of internal variability ($\square_M$ in Equations 4.5) based on the perturbed ensembles or based on the control simulation. We found from Figures 4.21 and 4.22 that the estimate of internal variability from the Ga ensemble was less than that obtained from the control simulation, and suggested two possible reasons:

(a) part of the climatic response to radiative forcing is a decrease in temperature variability (at 30-year time scales);
(b) the intra-ensemble variance is an under-estimate of temperature variability.

We investigate this further by plotting $\square_M$; we globally average $\square_M$ at each grid-box (Figure 4.24) for:

- temperature (left) and precipitation (centre),
- each overlapping period for the Ga (blue) and Gd (green) ensembles, and
- the set of 30-year periods in the control (dotted black).

There is no obvious trend in temperature variability (left), which is consistent with a shift in the entire probability distribution in response to radiative forcing. Since the internal variability of temperature is consistently less in the ensembles than in the control, it is likely that the intra-ensemble variance is an under-estimate of the temperature variability in the model on this 30-year time scale.

The ensembles also underestimate the internal variability of precipitation early in the scenarios, but the level of precipitation variability rises during the scenarios. Because of the nature of precipitation, its variability generally varies in proportion to its mean (a property that statisticians call heteroscedasticity). In order to examine whether or not the increase in precipitation variability is heteroscedastic,

we calculate the coefficient of variation[131] for each grid-box and period in the ensembles, and for each grid-box in the control, and globally average the coefficient (Figure 4.24, right). If the coefficient of variation were constant, any changes in variability would be heteroscedastic; however there is a trend in the coefficient of variation, which suggests that the heteroscedasticity of the precipitation only explains part of the increase in variability under the scenarios.

Since the intra-ensemble variance appears to be an under-estimate of the model variability, we calculate $\square_M$ (Equations 4.5) from the control in the analyses that follow. The changes in precipitation variability under radiative forcing imply that the significances at the end of the scenarios will be slightly too large relative to the start of the scenarios. Our discussion of the changes in variability under increased radiative forcing will be taken further in chapter 5.

4.7b Time-series of estimation errors

In section 4.6 we used pattern scaling to estimate climate changes for particular time periods, and displayed the results as spatial plots. In this section we estimate climate changes for overlapping 30-year periods, summarise the results into a single measure, and plot them as time-series.

We present a set of four figures (4.25-4.28). We constructed four response patterns for each variable using the four forcing scenarios; in each case we constructed the pattern from the entire set of overlapping 30-year periods available from the simulation or ensemble. For each figure we estimated:

- the temperature (4.25-4.26) or precipitation (4.27-4.28),
- in Ga (4.25, 4.27) or Gd (4.26, 4.28),
- for each overlapping 30-year period,
- by scaling each of the four patterns.[132]

We present three different statistics, each of which is plotted against time (top) and against the scaler (bottom):

- The left-hand plots show the global-mean of the error at each grid-box (the 'mean error'). When the error equals zero (indicated by the dotted black line in

[131] The coefficient of variation is calculated using absolute values, not anomalies, and is the standard deviation divided by the mean.

[132] In each figure the colour of the lines represents the pattern that has been scaled to provide the estimate, not the estimate itself. Ga is blue, Gd is green, s550 is red, and s750 is brown.

the left-hand plots) the implication is not that there are no errors, but that the sum of the positive errors is balanced by the sum of the negative errors.

- The central plots show the global-mean of the magnitude of the error at each grid-box (the 'error sizes').
- The right-hand plots show the global-mean of the significance of the error at each grid-box (the 'significances', Equations 4.5). The significance of the grid-box error is expressed in standard deviations. Significances are only calculated for Ga and Gd, the only scenarios with ensembles.

The first point to note is that pattern scaling introduces errors even in the absence of radiative forcing. This may be seen by inspecting the first decades in Figures 4.25 and 4.27 (left and centre). These errors stem from two principal causes:

(a) Internal variability introduces differences between any generalised pattern and the pattern at a particular point in time;

(b) Differences between the patterns of climate change accompanying global-mean temperature changes arising from natural variability (as in the control or early in the perturbed simulations) and arising from anthropogenic increases in greenhouse gases (later in the perturbed simulations).

On the basis of the left-hand plots we estimate the global-mean errors to amount to ±0.03°C or ±0.005 mm/day, which exhibit variability on all time scales, including the centennial. The average magnitudes of the grid-box errors are 0.1°C or 0.07 mm/day (central plots).

The errors under additional radiative forcing are much larger than under unperturbed conditions. Global-mean errors (left-hand plots) may be multiplied many times over to ±0.1°C or ±0.03 mm/day, and average magnitudes may triple to 0.3°C or 0.2 mm/day (central plots). Again there are two principal causes for these *increases* in error:

(c) Non-linearities in the global-mean climate changes that accompany a particular change in the scaler.

(d) Non-linearities in the responses of individual grid-boxes (which may or may not cancel out to leave a linear change in the global-mean).

4.7b Estimates made for scenario X by scaling a pattern from X

We consider in particular the estimates made using a pattern from the same scenario.[133] In each case (Figures 4.25-4.28, top left) there is an oscillation in the error during the course of the 21st century as radiative forcing increases. We

[133] The blue lines in Figures 4.25 and 4.27, and the green lines in Figures 4.26 and 4.28.

overestimate the temperature changes early in the 21st century, and underestimate them late in the 21st century. In other words, seasonal (JJA) temperature has a non-linear relationship with the scaler that is not unlike an exponential curve. When we turn to precipitation we find that the estimates of Ga and Gd changes (Figure 4.27 blue, 4.28 green, both top left) also show this oscillation, but *in opposite directions*: for Ga [Gd] the estimate is too dry [moist] in the early 21st century, and too moist [dry] in the late 21st century. The explanation may be found in Figure 4.1 (bottom centre): the scaling errors result from the deviations of these lines from linearity. It appears from Figures 4.27 and 4.28 that these errors in estimating precipitation may not exceed those that would be obtained from sources (a) and (b).

We found in section 4.5 that there are differences between the response patterns at different periods within a forcing scenario (*e.g.* Figures 4.11, 4.17), and we noted then that these would introduce errors to pattern scaling. In section 4.6 we found such errors when we scaled a pattern to estimate changes in the same scenario (*e.g.* Figure 4.19). Above we found a similar oscillation in the global-mean error during the 21st century. The figures[134] show the changes in the average magnitudes of the grid-box errors, in which the oscillation may be seen again. Pattern scaling approximates the non-linear responses in linear terms, and since the response patterns are constructed from the entire length of simulation the best-fit line passes between the non-linear extremes. Therefore the average error at individual grid-boxes is maximised at the beginning and end of the 21st century, and minimised in the 2050s.

When we calculate the statistical significance of the grid-box errors,[135] we find that averaged over the globe as a whole these errors do not exceed two standard deviations. The error significances are maximised at the end of the 21st century, so we use Figures 4.20 and 4.21 to assess the errors that are most statistically significant. We find that the proportion of grid-boxes where the error is statistically significant is only a quarter.[136] There are even fewer grid-boxes where the error is both statistically significant and larger than a fifth of the estimated value.[137]

---

[134] See the blue lines in Figures 4.25 and 4.27, and the green in 4.26 and 4.28 (all top centre).

[135] See the blue lines in Figures 4.25 and 4.27, and the green in 4.26 and 4.28 (all top right).

[136] The proportions are 27.5% for temperature (Figure 4.20) or 24.3% for precipitation (4.21).

[137] The proportions are 5.6% for temperature (Figure 4.20) or 21.2% for precipitation (4.21).

4.7c Estimates made for scenario X by scaling a pattern from Y

We found in section 4.5 that there are significant differences between response patterns extracted from the Ga and Gd ensembles (*e.g.* Figures 4.14 and 4.15). We noted in section 4.6 the significant errors that were consequently introduced by pattern scaling (*e.g.* Figures 4.22 and 4.23). Here we present time-series of the errors introduced by pattern scaling from one scenario to the other,[138] and also by pattern scaling from the stabilisation scenarios to the ensembles.[139]

The global-mean errors are larger than when scaling a pattern from the same scenario. This is because the non-linearities in the responses mean that the response in each scenario is best approximated by a pattern extracted from *itself* rather than a *different* scenario. The Ga patterns are warmer and drier than the Gd patterns (Figure 4.3, top), so when the Ga pattern is too warm in the early 21st century (Figure 4.25, top left, blue), the Gd pattern actually provides a better estimate (green). However, as the 21st century progresses and the Ga simulations warm faster, the Ga pattern underestimates the warming and therefore the Gd pattern underestimates it even more. This transition may be seen even more clearly in the average magnitudes of the grid-box errors (Figure 4.25, top centre). A similar transition occurs for precipitation (Figure 4.27).

The rapid growth in error at the end of the Ga scenario is because the Ga pattern – and even more the Gd pattern – is unable to represent the exponential change. However, because the exponential response does not advance as far under Gd as under Ga (for an illustration see Figure 4.6, bottom), the Ga pattern gives a better approximation of the Gd response than is the case *vice versa*. Therefore although the Ga pattern does not approximate the Gd response as well as the Gd pattern itself, it is still reasonably accurate (Figures 4.26 and 4.28). The globally averaged error significance (top right) is only a single standard deviation at the end of the 21st century. The proportion of grid-boxes with statistically significant errors that this represents is only a tenth,[140] and the proportion where the errors are also larger than a fifth of the estimated value is even smaller.[141] The greater accuracy

[138] See the green lines in Figures 4.25 and 4.27, and the blue in Figures 4.26 and 4.28.

[139] See the red (s550) and brown (s750) lines in Figures 4.25-4.28. We cannot calculate error significances because these require ensembles of patterns (see Appendix 4.2).

[140] The proportions are 13.8% for temperature (Figure 4.22) or 10.6% for precipitation (Figure 4.23).

[141] The proportions are 1.8% for temperature (Figure 4.22) or 8.9% for precipitation (Figure 4.23).

found when scaling from Ga to Gd than *vice versa* may be described as the superiority of interpolation over extrapolation.[142]

Under stabilisation the rate of global-mean change (relative to the scaler) further warms and moistens (Figures 4.1, 4.3), particularly the moistening in s550; so the global-mean errors from scaling the stabilisation patterns are also warmer and moister than those from scaling the ensemble patterns.[143] However, there is a contrast between s550 and s750 in the growth of the errors in individual grid-boxes (Figures 4.25-4.28, top centre) because of the different amounts of global warming in them as they stabilise. Because s750 warms more than s550 it progresses further along the exponential response (*e.g.* Figure 4.6, bottom), so the errors it introduces are larger early in the 21st century than late in the 21st century, whereas the opposite is the case for s550. Although we are unable to calculate the statistical significance of these errors because of the lack of ensembles for the stabilisation scenarios, the similarities between the top centre and top right plots for Ga and Gd (Figures 4.25-4.28) lead us to suggest that there may not be substantially more grid-boxes with significant errors than when scaling the Ga and Gd patterns.

4.8 Added value from scaling

We have evaluated pattern scaling principally in terms of the errors in its estimates. However, often regional climate scenarios must be developed for forcing scenarios for which there are no GCM simulations.[144] In such cases some form of scaling cannot be avoided, the only alternative (other than refraining from presenting any information at all) being to assume that there is no pattern to the global climatic change and to present the user with a global temperature anomaly. In these circumstances it is reasonable to evaluate the accuracy of scaling in estimating regional climate changes in the model against the accuracy of this assumption.[145]

---

[142] We extrapolate when we construct a pattern from a relatively weak forcing scenario and scale it to estimate changes under a relatively strong forcing scenario. To interpolate is to estimate in the opposite direction.

[143] Compare the red/brown lines with the blue/green lines in Figures 4.25-4.28 (top left).

[144] Although the Casino-21 project (Allen, 1999) and the Hadley Centre (UKMO/DETR, 2000) have plans to perform hundreds of model simulations, the emphasis is on different models rather than different forcing scenarios.

[145] We recognise that the pattern of regional climate changes exhibited by the GCM may not accurately reflect the actual response of the climate system, and that the GCM pattern may actually

For temperature we construct an additional response pattern to those constructed from the four forcing scenarios in HadCM2, in which each grid-box is set to 0.86°C per degree of global warming.[146] Similarly, for precipitation we construct a pattern where each grid-box is set to 0.043 mm/day per degree of global warming. We scale these patterns and calculate the absolute errors in an identical manner to that employed in Figures 4.25-4.28, and we plot the results in Figure 4.29 (temperature) and 4.30 (precipitation). For both temperature and precipitation the absolute errors from the new patterns (black lines) greatly exceed the errors from estimating by scaling response patterns (coloured lines).

We acknowledge that a small number of statistically significant errors are introduced in scaling from one forcing scenario to another, particularly when extrapolating rather than interpolating. However, the errors are much less than under the assumption that there is no pattern to global climate change.

4.9 Single simulations

We follow Mitchell *et al.* (1999) in asking "how accurately can one approximate the ensemble mean response using a single simulation" (p562)? This is relevant not just for our analysis in this chapter, where we have scaled patterns derived from single stabilisation scenarios (section 4.7), but for a later chapter (7) when we scale patterns for individual simulations available from a number of GCMs.

Here we construct response patterns for each individual ensemble member, scale them, and calculate the error between each estimate and the modelled ensemble mean. In Figures 4.31 and 4.32 we plot the globally averaged magnitudes of the individual grid-box errors,[147] both for the individual ensemble members (thin lines) and for the ensemble means (thick lines). The difference between the errors obtained from individual members and the ensemble means gives an indication of the component of the error from internal variability ($\Box$ in section 4.6) when estimating by scaling a pattern from a single simulation. The figures (4.31, 4.32)

---

be no more accurate than one of these assumptions. However, our aim here is to evaluate the accuracy of scaling in approximating the GCM responses, whether or not the GCM accurately represents the response of the climate system. The analysis described here should not be understood as an evaluation of the climate scenarios developed under different methods, but as an evaluation of scaling as a method in developing climate scenarios.

[146] 0.86 is the global-mean JJA warming per degree of global annual warming in Ga (Figure 4.3).

[147] This statistic is equivalent to that plotted in Figures 4.25-4.28 (centre).

show that δ only contributes a small proportion of the error. On this basis it appears reasonable to scale patterns constructed from single simulations.

4.10 Conclusions

In this chapter we have examined the possibility of scaling climate change patterns by annual global-mean temperature to estimate changes under unmodelled forcing scenarios. In many respects we have followed the lead of Mitchell *et al.* (1999), particularly in:

- calculating signal-to-noise ratios that represent the spatial heterogeneity of the climate response to additional radiative forcing;[148]
- using least squares regression to calculate response patterns;[149]
- calculating the estimation errors from scaling a response pattern from the same forcing scenario;[150]
- calculating the estimation errors from scaling a response pattern from a different forcing scenario.[151]

Therefore it is not surprising that in many respects we have obtained similar results to Mitchell *et al.* (1999):

- the grid-box errors when scaling the Ga temperature pattern to estimate Ga changes have typical magnitudes of 0.1-0.2°C, and are maximised at the end of the 21st century;[152]
- the grid-box errors when scaling the Ga temperature pattern to estimate Gd changes also have typical magnitudes of 0.1-0.2°C.[153]

However, we have shown that pattern scaling is accurate in many more ways than Mitchell *et al.* (1999) showed:

- pattern scaling is accurate for seasonal means, not just annual means;[154]
- pattern scaling is accurate for most grid-boxes, not just sub-continental regions;[155]

[148] Compare their Appendix with our section 4.4, their Figures 1c and 12 with our Figure 4.7.

[149] Compare their Appendix with our section 4.5, their Figure 2 with our Figure 10 (bottom left).

[150] Compare their Figure 4 with our Figure 4.19 (top).

[151] Compare their Figure 7 with our Figures 4.22 and 4.23 (top).

[152] Compare their Figure 4 with our Figure 4.25 (top centre, blue).

[153] Compare their Figure 7 with our Figure 4.27 (top centre, blue).

[154] Compare their Figure 7 with our Figure 4.27 (top centre, blue).

[155] Compare their Table II with our Figure 4.22 (bottom left).

- signal-to-noise ratios that represent the spatial heterogeneity of the climate response to additional radiative forcing are large for precipitation, not just temperature;[156]
- pattern scaling is accurate for precipitation, not just temperature.[157]

We have also paid close attention to the errors that are introduced by pattern scaling. In section 4.6 we described the errors in terms of the equation: $y=ax+\square+\square$.[158] In section 4.7 we outlined four potential sources of error, the first being $\square$, and the final three contributing to $\square$.

(a) Internal variability introduces differences between any pattern from a long period of time, and a pattern from a particular point in time;
(b) There may be differences between the patterns of climate change accompanying global-mean temperature changes arising from natural variability (as in the control or early in the perturbed simulations) and those arising from anthropogenic increases in greenhouse gases (later in the perturbed simulations).
(c) There may be non-linearities in the global-mean climate changes that accompany a particular change in the scaler.
(d) There may be non-linearities in the responses of individual grid-boxes (which may or may not cancel out to leave a linear change in the global-mean).

We found that internal variability – (a) in the list above – made a relatively small contribution to the total errors from pattern scaling.[159] Therefore we are reasonably confident that internal variability will not greatly add to the errors when we use single simulations – such as for the stabilisation scenarios in this chapter, or for patterns derived from various GCMs in chapter 7.[160] We found in section 4.7 that combining (a) and (b) gives global-mean errors of ±0.03°C or ±0.005 mm/day; the average magnitudes of the individual grid-box errors are 0.1°C or 0.07 mm/day. These errors may increase under radiative forcing because of sources (c) and (d).

[156] Compare their Figure 12 with our Figure 4.7 (centre).

[157] Compare Figures 4.20 and 4.21, and compare Figures 4.26 and 4.28 (both blue).

[158] The algebraic signs represent the modelled value (y), the estimate (ax) derived from the response (x) multiplied by the scaler (a), the systematic error ($\square$), and the error from internal variability ($\square$).

[159] See section 4.9 and Figures 4.31, 4.32.

[160] In the case of chapter 7, this assumes that internal variability plays a similar role in other models to that which it plays in HadCM2.

The global-mean climate changes that accompany a particular change in the scaler are not entirely linear. This is demonstrated by the global-mean errors that are introduced by scaling a pattern from one scenario to estimate changes in another,[161] and more directly by Figure 4.3. Consequently the response patterns err in one direction early in the 21st century, and in the other direction late in the 21st century.

There are also strong non-linearities in the responses of individual grid-boxes, and therefore there are also changes in the spatial pattern of response during the 21st century. This was demonstrated by the non-linear relationships between the signal-to-noise ratios and the scaler (Figures 4.6, 4.7), and later by calculating response patterns for particular periods in the 21st century and comparing them (Figures 4.11, 4.17). This latter experiment showed that more than a third of the grid-boxes experience statistically significant changes in response during the 21st century: 42.8% (temperature) and 34.0% (precipitation). These changes in response add further to the errors introduced by scaling a response pattern from Ga to estimate changes in Ga: at the end of the 21st century a quarter of the grid-boxes have statistically significant errors (27.5% for temperature and 24.3% for precipitation).

However, we recognise that not all statistically significant errors may be practically significant. Unlike statistical significance, practical significance is context-dependent, so we cannot make a universal definition of practical significance. Therefore we arbitrarily define an error that exceeds a fifth of the estimated value as being practically significant, and identify the proportions of grid-boxes that are simultaneously statistically and practically significant. In practice this means that pattern scaling for all other grid-boxes may be treated as being accurate with an error margin of ±20%. We find that when scaling a response pattern from Ga to estimate changes in Ga at the end of the 21st century, the errors are both statistically and practically significant in only 5.9% (temperature) or 21.3% (precipitation) of grid-boxes (Figures 4.20, 4.21).

The differences between response patterns extracted from equivalent periods in different forcing scenarios are much smaller than the differences between patterns within a forcing scenario.[162] Such differences are due to the dependence of the spatial climate response on the *rate*, not just the *amount*, of change in the scaler.

[161] See Figures 4.25 and 4.27 (top left).

[162] Compare Figures 4.13 and 4.18 with Figures 4.11 and 4.17.

The differences between response patterns extracted from the entire length of simulation in different forcing scenarios are a combination of all the different sources we have described above.[163] The same sources introduce errors into scaling a pattern from one forcing scenario to estimate changes in another forcing scenario. Such errors are minimised when interpolating from a stronger forcing scenario (*e.g.* Ga) to a weaker (*e.g.* Gd), rather than extrapolating in the opposite direction.[164] When interpolating, the errors at the end of the 21st century are both statistically and practically significant in only a small proportion of the grid-boxes (Figures 4.22, 4.23): 1.8% (temperature) or 8.9% (precipitation).

We were hampered in our assessment of the stabilisation scenarios by the lack of any ensembles for them, but we did ascertain that compared to the Ga and Gd scenarios:

- their global-means have slightly less linear relationships with the scaler (Figure 4.1);
- their rates of global-mean response are slightly more dependent on the rate of increase in the scaler (Figure 4.3);
- their spatial patterns of response are similarly heterogenous (Figure 4.6);
- their differences from Ga in the spatial pattern of response are similar to Gd (Figure 4.12);
- the global-mean errors introduced by pattern scaling are larger, but it is possible that there are not substantially more grid-boxes with statistically significant errors (Figures 4.25-4.28).

On the basis of our work we may give an affirmative answer in response to the question posed in the chapter title, namely 'May we use pattern scaling?' We also add that pattern scaling is much more accurate than assuming that the global-mean response holds for individual grid-boxes (section 4.8), which may be the only alternative in some cases. However, we also recognise the limitations of our work:

- we have only examined a single season (JJA);
- we have only considered a single model (HadCM2);
- we have not attempted to include sulphate aerosols;
- we have only considered temperature and precipitation;

[163] See Figures 4.14 and 4.15.

[164] For example, compare the blue lines in Figures 4.26 and 4.28, with the green lines in Figures 4.25 and 4.27.

- we have only four members in each ensemble;
- we have not considered any period after climate has been fully stabilised.

The most important limitation is that we have evaluated pattern scaling on the assumption that HadCM2 perfectly represents the climate system.

There are a number of ways in which our work might be extended. Some opportunities are offered by the list of limitations given above. Other opportunities are taken in the chapters that follow.

It is possible that further work would reveal similarities between the evolving climatic responses to increased radiative forcing under different forcing scenarios. For example, the changes over time in the pattern of response in one stabilisation scenario might correspond to those in another. In that case a superior form of scaling might use two parameters: one parameter might estimate the unstabilised response as a function of the scaler (as we have done here), and an additional parameter might estimate the stabilising component of the response, perhaps as a function of the gradient of the scaler.

The suggestion above would be a complex method of scaling. A simpler method, but still more complex than that employed in this chapter, might be to expand the linear equations, by which the response is estimated and the pattern scaled, to quadratic equations. Scaling quadratically would enable us to incorporate many of the non-linear changes within a single forcing scenario. Perhaps the critical factor in both these suggestions might be the extra amount of information (*i.e.* ensemble members) required to construct stable estimates of the second constants.

Appendix 4.1

In this appendix we compare the signal-to-noise ratios that we obtained with those plotted by Mitchell *et al.* (1999). The method that we followed (Equations 4.2) was exactly that described by Mitchell *et al.* (1999, Appendix), so one might expect the results we obtained (Figure 4.5) to be identical to those plotted by Mitchell *et al.* (1999, Figures 1c, 12), but they are not. For temperatures, they obtained signal-to-noise ratios for Ga and Gd of 15.5 and 6.5 respectively, whereas we obtained 11.0 and 7.2. For precipitation (mm/day anomalies) they obtained 7.4 and 2.7, whereas we obtained 6.4 and 4.0. Having checked our method and calculations carefully and found them to be correct, we examined those of Mitchell *et al.* (1999).

According to the Appendix and the text in Mitchell *et al.* (1999), "all the temperature changes are calculated relative to the mean over 1961 to 1990" (p555). However, the graphs (*e.g.* Figure 1) express the changes relative to the 1990-1999 decadal mean, because the time series were "reset" to make the 1990s zero (John Mitchell, pers. comm.). A side-effect of this "zeroing" was to reduce the *plotted* global temperature change over 1990-2099 from that *calculated* by about 10%, because effectively only 1995-2095 is plotted (John Mitchell, pers. comm.). This is confirmed by other published results (*e.g.* Mitchell and Johns, 1997, Figure 2b; Mitchell *et al.*, 2000, Figure 1) where the global temperature changes for the full period 1990-2099 are 10% greater than those plotted in Mitchell *et al.* (1999, Figure 1b), and with which our calculations are in agreement. Since a similar "zeroing" is applied to the signal-to-noise ratios (Figure 1c), a similar side-effect will be introduced to the signal-to-noise ratios. This will affect the comparison of the results we obtained with those plotted by Mitchell *et al.* (1999).

However, we suspect that the "zeroing" is not the only difference between the published method and plotted results in Mitchell *et al.* (1999). It is suspicious that although the ratio between the Ga and Gd global mean temperatures is only 1.7 (their Figure 1b, in approximate agreement with the values we obtained), the ratio between the Ga and Gd weighted signals increases to 2.4 (their Figure 1c), whereas the ratio we obtained remained close to 1.7. A similarly large unexplained increase is apparent for precipitation (their Figure 12). Although it is conceivable that the large increase in the ratio is a genuine result, the implication would be that under a faster rate of increase in radiative forcing, the regional contrasts are much greater than at the same global mean temperature under a slower rate of forcing.

However, when we test this implication, we find that the regional contrasts in temperature vary little between different radiative forcing scenarios. We established this by taking the temperature anomaly patterns for Ga and Gd for the 30-year periods centred on 2045 and 2073 respectively[165], and calculating the weighted global-mean and global weighted signal for each ensemble (Equation 4.2.1). As expected, the weighted global-mean anomaly was almost identical for the two forcing scenarios (with a ratio between them of 1.00656); the interesting result was that the global weighted signals from the two forcing scenarios were even closer to being identical (the ratio was 1.00314). We conclude from this result that the regional temperature contrasts vary little between different forcing scenarios, when periods of equivalent temperature are considered. Hence we consider it unlikely that the ratio between the global temperature means of Ga and Gd should be so greatly exceeded by the ratio between their signals, as suggested by the plots in Mitchell *et al.* (1999).

Therefore we consider it probable that the signal-to-noise ratio plots in Mitchell *et al.* (1999) are erroneous. The source of the errors is a matter for conjecture, but we suggest a single deviation from the published method that may explain why the signal-to-noise ratios differ so widely from those that we obtained. Our suggestion is simply that the square-root operation in their Equation A.2 (or our Equation 4.2.1) was accidentally omitted. By using their published sampling error for the temperature pattern (their Figure 3b, which is close to our estimate), and by adopting our own estimate of the sampling error for the precipitation (mm/day) pattern, we obtained to a reasonable approximation the signals used by Mitchell *et al.* (1999). When we applied the square rooting operation to each of the signals (Ga and Gd for temperature and precipitation) and recalculated the signal-to-noise ratios, we obtained values for which Ga and Gd were in the correct proportion to each other, and values that were within 10% of those that we obtained. At least some of the remaining differences may be explained by the "zeroing" discussed above.

We raised these matters with John Mitchell personally. After an initial investigation he concluded that the signal-to-noise ratios plotted in Mitchell *et al.* (1999) were definitely erroneous, although he was unable at that time to precisely identify the source of the errors or the correct values (John Mitchell, pers. comm.). Therefore we conclude that the differences between the signal-to-noise ratios that

[165] These periods have equivalent global mean temperatures under both forcing scenarios.

we obtained and those plotted by Mitchell *et al.* (1999) should not cause us concern.

Appendix 4.2

In this appendix we demonstrate that our method of calculating the statistical significance of differences between response patterns (Equations 4.4) requires that ensembles are used to construct the response patterns. Our method (Equations 4.4) hinges on our ability to estimate σ_A^2 and σ_B^2, for which we require a sample (*i.e.* ensemble) with more than one member. If only a single simulation is available for *B* (as in the case for the HadCM2 stabilisation scenarios) then we are unable to estimate σ_B^2 using that simulation.

A solution would be possible if σ_A^2 and σ_B^2 were very similar, in which case we might use σ_A^2 (which we are able to estimate using the *A* ensemble) as an approximation for σ_B^2. If we are able to assume that for all practical purposes $\sigma_A^2 = \sigma_B^2$,[166] then the pattern difference when A_{bar} is subtracted from B_i is distributed as follows:

$$D \sim N(\mu_B - \mu_A, (\sigma_A^2/4 + \sigma_A^2))$$

or $$D \sim N(\mu_B - \mu_A, 5\sigma_A^2/4)$$

We cannot investigate the validity of the assumption for the HadCM2 stabilisation scenarios – for which it is required – precisely because of the lack of any ensembles. However, we can investigate the validity of the assumption using the temperature response patterns of the Ga and Gd ensembles.

We calculated σ_A (Ga) and σ_B (Gd) for each grid-box using the entire 240-year simulations, and plotted them and the differences between them in Figure 4.A2.1. There are substantial differences between σ_A and σ_B over the globe as a whole; when averaged over the globe, σ_B is 29.3% greater than σ_A. Of individual grid-boxes, nearly a third have values of σ_B that are twice as large or only half as much as their values of σ_A. We repeated our comparison for response patterns drawn from periods in Ga and Gd with equivalent global annual temperatures, and obtained similar results (Figure 4.A2.2). Since σ_A (Ga) and σ_B (Gd) are not equivalent, it does not seem reasonable to assume that the variances of the stabilisation scenarios are equivalent to that of Ga.

We may complete our comparison by quantifying the effects of making the assumption that $\sigma_A^2 = \sigma_B^2$ on the significance of the differences in response pattern.

---

[166] In other words, we assume that the internal climate variability under the *A* forcing scenario is similar to that under the *B* forcing scenario.

The statistic against which we measure the differences in response pattern is D, the pattern difference when one ensemble mean pattern is subtracted from another (Equations 4.4). Here the variance of D without the assumption is ($\sigma_A^2/4 + \sigma_B^2/4$), and with the assumption it is ($\sigma_A^2/2$). We calculate the significance of D with and without the assumption, and plot the differences between them (Figure 4.A2.3 top). The differences between σ_A (Ga) and σ_B (Gd) described above suggest that the variance of D will depend on whether or not the assumption is made. Making the assumption does indeed introduce an error: when globally averaged, the number of standard deviations represented by D decreases by 16.8%. Therefore when we divide the response pattern difference by D to obtain the significance of the difference in standard deviations from the mean, the number of significant differences is erroneously large (Figure 4.A2.3 bottom). By adopting the assumption, the number of grid-boxes that fall outside the $\pm 2\sigma$ limits and are therefore stated to have climate responses in Ga and Gd that are significantly different at the 95% level, rises from 22.6% to 30.4%. Therefore we conclude that $\sigma_A^2 \neq \sigma_B^2$, and that any assumption to the contrary is likely to introduce substantial errors into any estimation of the significance of the differences in response patterns.

A second conceivable solution is to divide up the single simulation into a number of consecutive periods, and to estimate σ_B^2 by treating them as a pseudo-ensemble. This is feasible if we may assume that the periods are statistically independent and a random sample of a single distribution. The first assumption is warranted if we split the periods in time; for s550, we might select 30-year periods centred on 2120, 2160, 2200, and 2240, as we have already done (section 4.5). However, the second assumption is only warranted if the climate system is not responding to any perturbation. Although we commonly make this assumption for a control simulation, it is not so easily justified for the HadCM2 stabilisation scenarios, since the climate system is still responding to radiative forcing and is not stabilised until the very end of the scenario, if at all.[167] Therefore it does not seem reasonable to apply this second solution to the HadCM2 stabilisation scenarios.

[167] For s550, the climate system is still responding to radiative forcing in the selected periods: global annual temperature anomalies (against the control) are still rising (Figure 4.1, top left), from 2.79°C in the period centred on 2020, to 3.06°C in 2160, 3.16°C in 2200, and 3.30°C in 2240. Senior and Mitchell (2000) found that temperatures continued to rise for 800 years after stabilisation of CO_2 concentrations at double pre-industrial levels.

Equations 4.1

1. Global temperature anomaly (a, °C) for a single 30-year period mean (j) in a single simulation (m).

$$a_{jm} = \frac{1}{n_i} \sum_{i=1}^{n_i} \left(t_{ijm} - t_{ib} \right)$$

2. Global precipitation anomaly (a, mm/day) for a single 30-year period mean (j) in a single simulation (m).

$$a_{jm} = \frac{1}{n_i} \sum_{i=1}^{n_i} \left(p_{ijm} - p_{ib} \right)$$

3. Global precipitation anomaly (a, %) for a single 30-year period mean (j) in a single simulation (m).

$$a_{jm} = 100 \left[\frac{\sum_{i=1}^{n_i} p_{ijm}}{\sum_{i=1}^{n_i} p_{ib}} - 1 \right]$$

4. Globally-averaged regional precipitation anomaly (a, %) for a single 30-year period mean (j) in a single simulation (m).[168]

$$a_{jm} = \frac{100}{n_i} \sum_{i=1}^{n_i} \left[\frac{p_{ijm}}{p_{ib}} - 1 \right]$$

[168] The difference between (3) and (4) is that in (3) we calculate the global-mean before we anomalise it, whereas in (4) we calculate the anomaly of each grid-box before averaging into a global-mean. For absolute anomalies (not percentage anomalies), the methods are identical.

Equations 4.2

1. Temperature (°C) or precipitation (mm/day) **signal** (s_{Mj}) for the pattern of 30-year period mean (j) anomalies (a) from an ensemble mean (M).

$$s_{Mj} = \sqrt{\frac{1}{\sum_{i=1}^{n_i} w_i} \sum_{i=1}^{n_i} w_i a_{Mij}^2 - \left(\frac{1}{\sum_{i=1}^{n_i} w_i} \sum_{i=1}^{n_i} a_{Mij} w_i \right)^2}$$

2. Pattern **noise** corresponding to 4.2.1. A signal (s_{cj}) is constructed for each period[169] ($c=1 \ldots n_c$) in the 1400-year control, using 4.2.1 but with subscript c replacing subscript M. The noise is based on the mean of those signals.[170]

$$\sigma_j = \frac{\sum_{c=1}^{n_c} s_{cj}}{n_c}$$

3. The pattern **signal-to-noise ratio**, for a single 30-year period (j) from an ensemble mean (M), is simply[171]:

$$ratio = \frac{s_{Mj} \sqrt{4}}{\sigma_j}$$

4. Either of two sets of **weights** (w) may be applied. In the 'unweighted' set an equal weighting is given to all grid-boxes ($w_i=1$). The 'weighted' set gives a weight to each grid-box that is proportional to its surface area:

$$w_i = \cos(l_i) \frac{n_i}{\sum_{i=1}^{n} \cos(l_i)}$$

---

[169] ***If*** the anomalies (a) in 4.2.1 are based on the first 240 years of the control, the remaining 1160 years of the control are divided into 40-year sequential segments, and the periods in 4.2.2 ($c=1 \ldots n_c$) are the first 30 years of each segment. ***If*** the anomalies (a) in 4.2.1 are referenced to the period 1960-1989 from the Ga ensemble, the control is divided into ten 140-year segments, and the periods in 4.2.2 ($c=1 \ldots n_c$) are the periods in each segment corresponding to that in 4.2.1 (*e.g.* last 30 years in segment when period in 4.2.1 is 2070-2099).

[170] We understand this to be the method of Mitchell *et al.* (1999, p578); their temperature sampling error (their Figure 3b) appears sufficiently close to our result to confirm this.

[171] The noise is divided by two because the signal is averaged over four ensemble members.

Equations 4.3

The grid-box signal-to-noise ratio is constructed differently to the pattern signal-to-noise ratios described in Equations 4.2.

1. Temperature (°C) or precipitation (mm/day) grid-box (*i*) **signal** (s_{Mij}) for a single 30-year period mean (*j*) from an ensemble.

$$s_{Mij} = \frac{1}{n_m} \sum_{m=1}^{n_m} a_{mij}$$

2. Grid-box (*i*) **noise** (σ_{ij}) corresponding to 4.3.1, using the set of periods[172] ($c=1 \ldots n_c$) in the 1400-year control.

$$\sigma_{ij} = \sqrt{\frac{n_c}{n_c - 1} \left(\frac{\sum_{c=1}^{n_c} a_{ijc}^2}{n_c} - \left(\frac{\sum_{c=1}^{n_c} a_{ijc}}{n_c} \right)^2 \right)}$$

3. The grid-box (*i*) **signal-to-noise ratio** for a single 30-year period (*j*) from an ensemble is simply[173]:

$$ratio = \frac{s_{Mij} \sqrt{4}}{\sigma_{ij}}$$

[172] If the anomalies (*a*) in 4.2.1 are referenced to the first 240 years of the control, the remaining 1160 years of the control are divided into 40-year sequential segments, and the periods in 4.2.2 ($c=1 \ldots n_c$) are the first 30 years of each segment. If the anomalies (*a*) in 4.2.1 are referenced to the period 1960-1989 from the Ga ensemble, the control is divided into ten 140-year segments, and the periods in 4.2.2 ($c=1 \ldots n_c$) are the periods in each segment corresponding to that in 4.2.1 (*e.g.* last 30 years in segment when period in 4.2.1 is 2070-2099).

[173] The noise is divided by two because the signal is averaged over four ensemble members.

Equations 4.4

To evaluate the statistical significance of the difference between two response patterns, it is necessary to calculate the theoretical distribution obtained when one random variable (a single response pattern) is subtracted from another. Thus we may calculate the number of standard deviations that the difference represents. This is a fairly simple procedure when the patterns are derived from ensembles.

We may calculate a temperature response pattern from a single simulation (m) in either one of two different ensembles (A and B), and represent it as belonging to a random variable (P_m). Since the patterns are constructed from 30-year means, we assume (under the central limit theorem) that they are distributed normally. By estimating the mean (μ_m) and variance (σ_m^2) of the population of patterns, we may quantify the random variables to which the patterns belong.

$$P_{mA} \sim \mathrm{N}(\mu_{mA}, \sigma_{mA}^2) \qquad P_{mB} \sim \mathrm{N}(\mu_{mB}, \sigma_{mB}^2)$$

However, we are more interested in the ensemble mean patterns, which we represent as belonging to new random variables (P_M):

$$P_{MA} \sim \mathrm{N}(\mu_{mA}, \sigma_{mA}^2/4) \qquad P_{MB} \sim \mathrm{N}(\mu_{mB}, \sigma_{mB}^2/4)$$

We may define the difference between the ensemble mean patterns as another random variable ($P_{(B-A)}$). Under our null hypothesis that there is no systematic difference between the two ensemble-mean patterns:

$$P_{(B-A)} \sim \mathrm{N}(0, (\sigma_{mA}^2/4 + \sigma_{mB}^2/4))$$

Since we may estimate the mean (μ_m) and variance (σ_m^2), we may quantify the random variable $P_{(B-A)}$. Therefore we may calculate the significance of the difference between ensemble mean response patterns in terms of the standard deviation of the difference expected from internal variability.

Equations 4.5

To evaluate the statistical significance of the error in estimating by pattern scaling, it is necessary to calculate the theoretical distribution obtained when one random variable (the simulated pattern) is subtracted from another (the estimated pattern). Thus we may calculate the number of standard deviations that the difference (the error) represents, which is a fairly simple procedure when estimating for an ensemble.

We may represent the simulated pattern for a single simulation (m) as belonging to a random variable (S_m). When this pattern is a 30-year mean we assume (under the central limit theorem) that it is distributed normally:

$$S_m \sim \mathrm{N}(\mu_{mS}, \sigma_{mS}^2)$$

We define our estimate of the simulated pattern as another random variable (E_m):

$$E_m \sim \mathrm{N}(\mu_{mE}, \sigma_{mE}^2)$$

However, we are more interested in estimating ensemble means, for which we define new random variables:

$$S_M \sim \mathrm{N}(\mu_{mS}, \sigma_{mS}^2/4)$$

$$E_M \sim \mathrm{N}(\mu_{mE}, \sigma_{mE}^2/4)$$

We may define the difference between S_M and E_M as another random variable (D_M). Under our null hypothesis that there is no systematic error in the estimate from pattern scaling:

$$D_M \sim \mathrm{N}(0, (\sigma_{mS}^2/4 + \sigma_{mE}^2/4))$$

Therefore we may calculate the significance of any errors that we find in terms of the standard deviation of the error expected from internal variability.

We estimate μ_{mE} by estimating for each individual ensemble member: we multiply the response pattern from a single member by the scaler from the same member. There are two options for estimating μ_{mS}:

1. from the control, using a sample of 30-year periods (with 10-year gaps);
2. from the ensemble, using the sample of four members.

Table 4.1

The changes represented by the 'strengthening of pattern' plots.

2020s	2080s	diff.	stren.	meaning
− −	− − −	−	>0	negative response becomes more negative
− −	−	+	<0	negative response becomes less negative
− −	+	+++	<-1	negative response becomes less negative to the extent of changing sign
+ +	−	− —	<-1	positive response becomes less positive to the extent of changing sign
+ +	+	−	<0	positive response becomes less positive
+ +	+ + +	+	>0	positive response becomes more positive

The comparison may be between response patterns from two 21st century periods in the same forcing scenario (as here), or between response patterns from two different forcing scenarios (in which case substitute 'Ga' and 'Gd' for '2020s' and '2080s' respectively). If the comparison is for *temperature* response patterns the values in the patterns are first reduced by one, so that a zero value in the pattern represents a response equal to the annual global-mean temperature.

The first two columns represent the sign of the response in a particular grid-box and the number of symbols represents the magnitude. The third column represents the sign of the change in response during the 21st century.[174] The fourth column identifies whether the response is becoming stronger (+) or weaker (–) during the 21st century.[175] The fourth column is plotted with a mask derived from the statistical significance of the third column; so as plotted it represents the *statistically significant* strengthening (and weakening) of the response pattern during the 21st century.

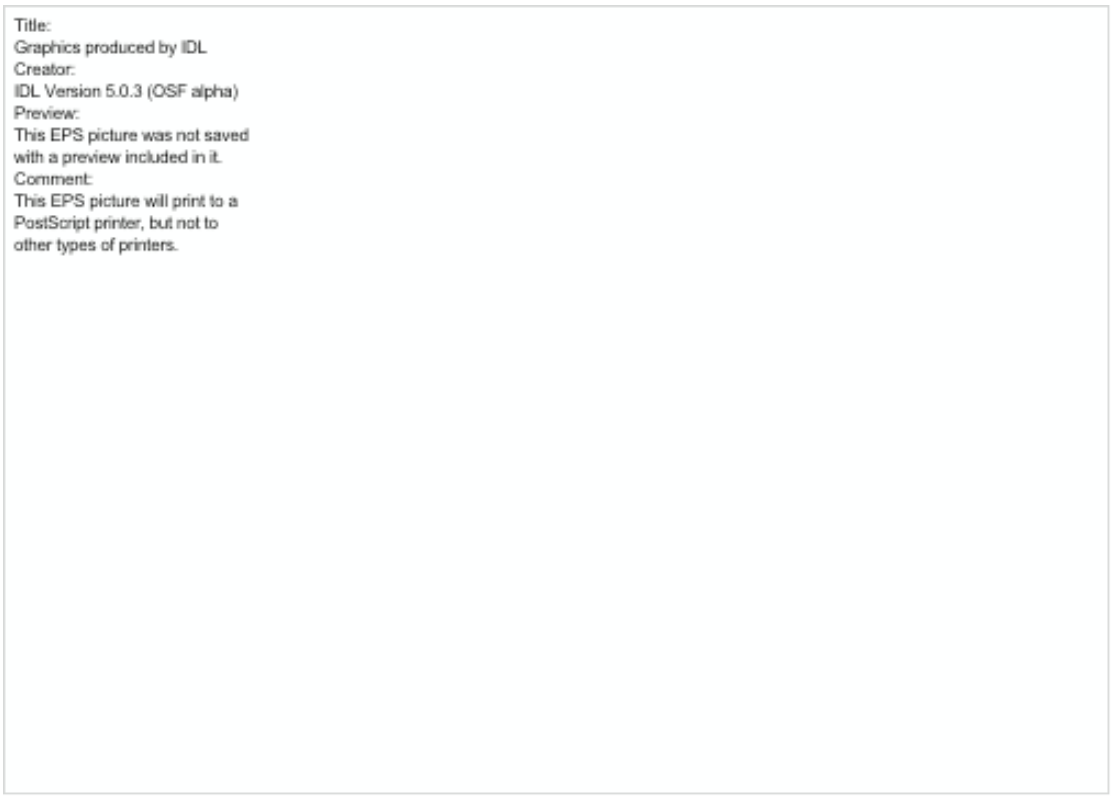
[174] In relevant figures the difference between the 2080s and 2020s is plotted in the top right.

[175] This measure is the 2080s minus 2020s difference, divided by the 2020s pattern. We call it the 'strengthening of the response' and plot it in the bottom right.

Figure 4.1

Time-series of global-mean JJA temperature and precipitation.

temperature (°C) precip (mm/day) precip (%)



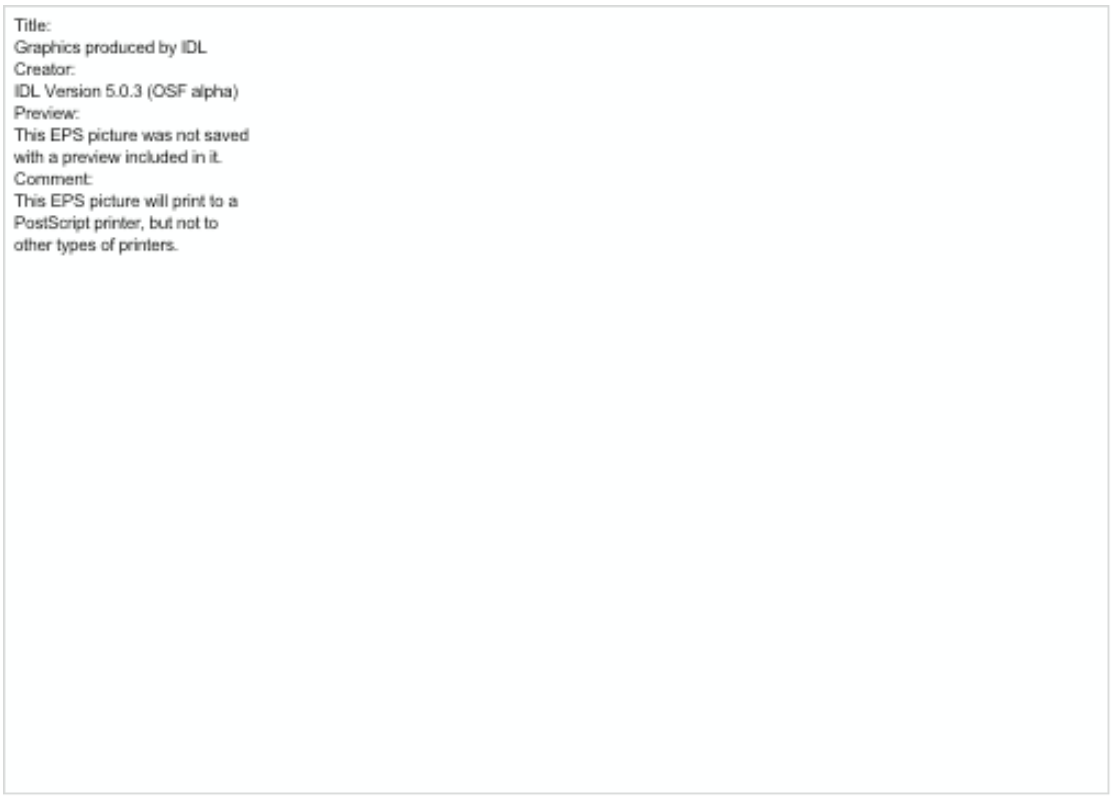
temperature (°C) precip (mm/day) precip (%)

model:	HadCM2
simulations:	Ga 1, Ga 2, Ga 3, Ga 4, Gd 1, Gd 2, Gd 3, Gd 4, s550, s750
base period:	the first 240 years of the control
anomalies:	temp (°C, Equation 4.1.1, **left**),
	prec (mm/day, Equation 4.1.2, **centre**),
	prec (%, Equation 4.1.3, **right**);
season:	JJA
temporal:	means of overlapping 30-year periods
spatial:	global-mean anomalies
x-axis:	period mid-year (**top**), period scaler (°C: **bottom**)

Figure 4.2

Contrast between global-mean and grid-box JJA precipitation.

global prec grid-box prec

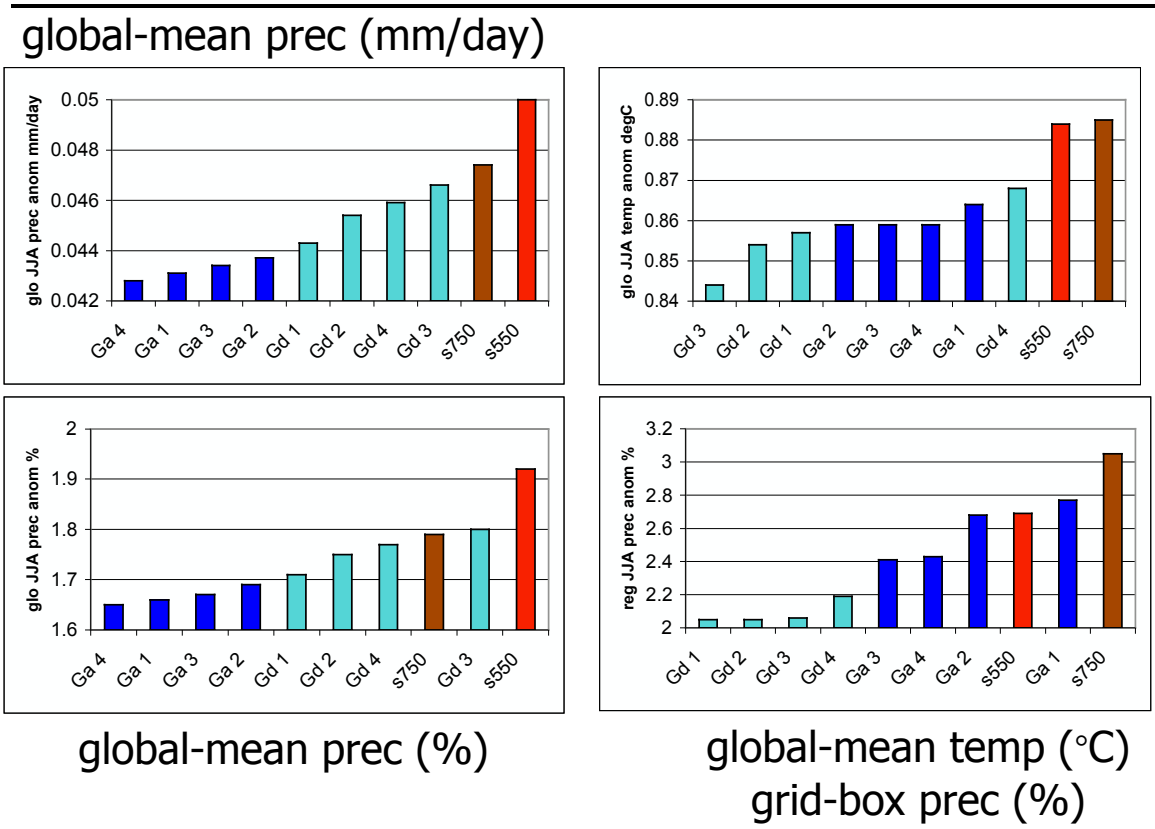


global precip grid-box prec

model:	HadCM2
simulations:	**Ga 1**, **Ga 2**, **Ga 3**, **Ga 4**, **Gd 1**, **Gd 2**, **Gd 3**, **Gd 4**, **s550**, **s750**
base period:	the first 240 years of the control
anomalies:	global-mean prec (%, Equation 4.1.3, **left**)
	average of all grid-box prec (%, Equation 4.1.4, **right**)
variables:	JJA precipitation anomalies (%)
temporal:	means of overlapping 30-year periods
x-axis:	period mid-year (**top**), period scaler (°C: **bottom**)
duplication:	the left plots are identical to the Figure 4.1 right plots

Figure 4.3

Best-fit gradients from Figures 4.1 and 4.2.

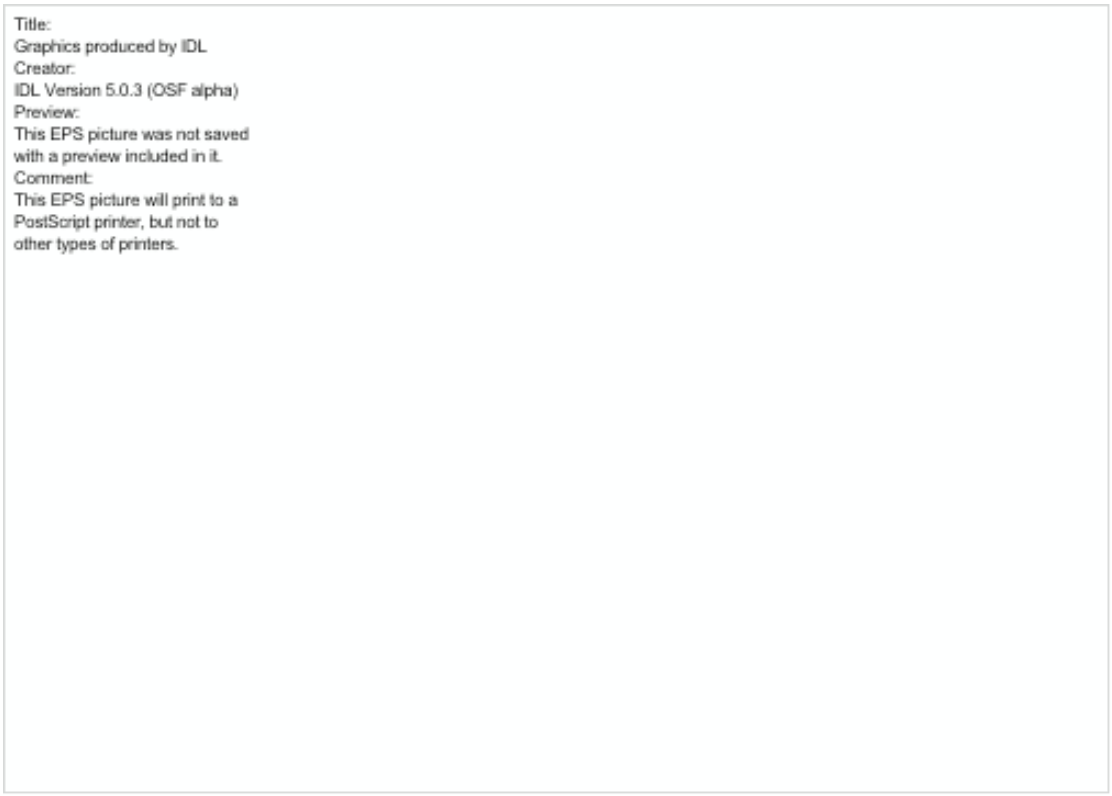


model:	HadCM2
simulations:	Ga 1, Ga 2, Ga 3, Ga 4, Gd 1, Gd 2, Gd 3, Gd 4, s550, s750
variable:	gradient (a) of best-fit line ($y=ax$) calc by least squares regression
y	scaler (annual global-mean temperature)
x	as specified below
top left:	global-mean JJA temperature (°C) from Figure 4.1 (bottom centre)
top right:	global-mean JJA precipitation (mm/day) from Figure 4.1 (bott. left)
bott. left:	global-mean JJA precipitation (%) from Figure 4.1 (bottom right)
bott. right:	grid-box JJA precipitation (%) from Figure 4.2 (bottom right)

Figure 4.4

Global JJA temperature and precipitation in terms of radiative forcing.

precip (mm/day) precip (%)

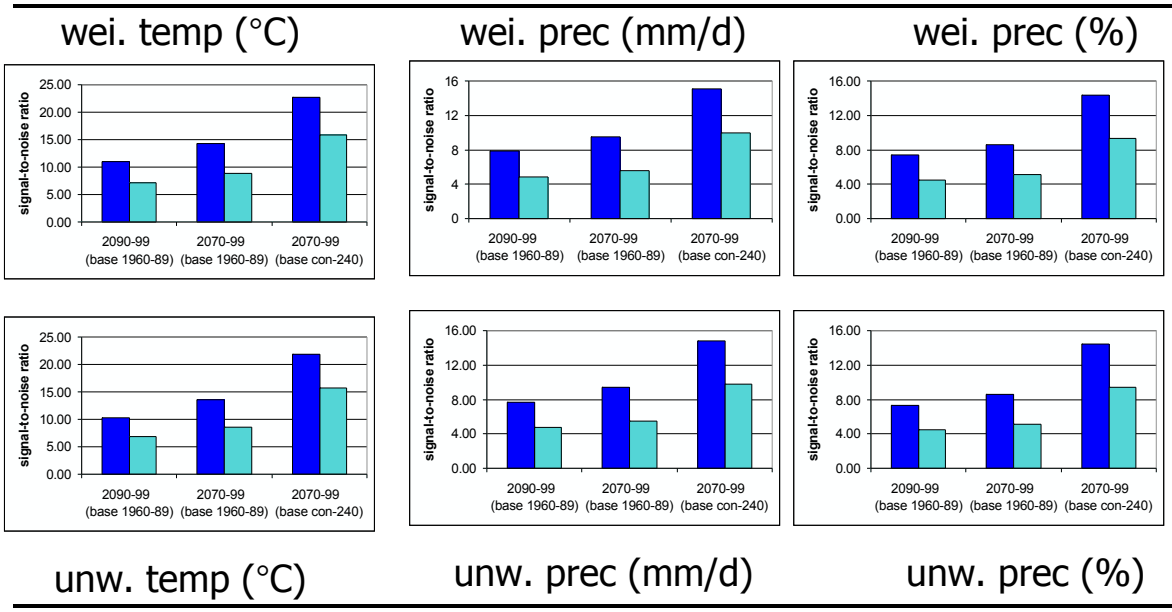


temp (°C)

model:	HadCM2
simulations:	Ga 1, Ga 2, Ga 3, Ga 4, Gd 1, Gd 2, Gd 3, Gd 4
anomalies:	JJA means of overlapping 30y periods, relative to first 240y of control
	global-mean temp (°C, Equation 4.1.1, **bottom left**),
	global-mean prec (mm/day, Equation 4.1.2, **top left**),
	global-mean prec (%, Equation 4.1.3, **top right**);
x-axis:	radiative forcing anomaly (Wm^{-2})

Figure 4.5

Pattern signal-to-noise ratios for annual means.

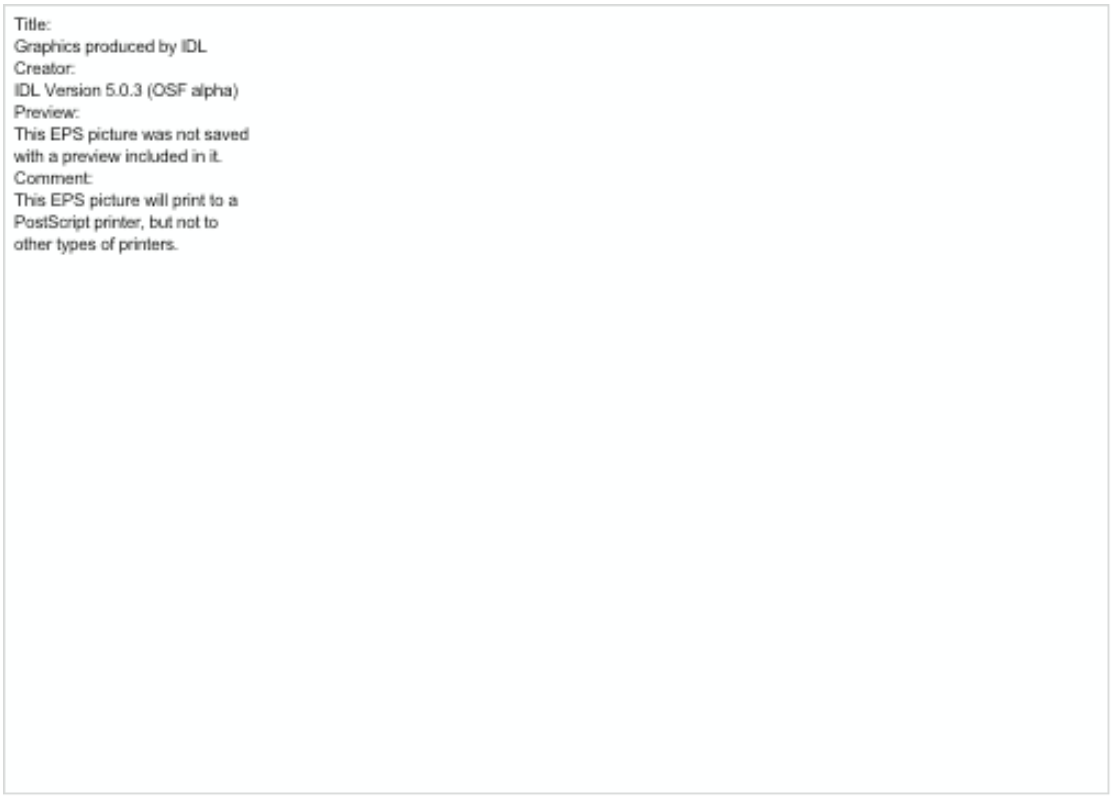


model:	HadCM2
simulations:	**Ga ensemble mean**, **Gd ensemble mean**
variable:	pattern signal-to-noise ratios for annual means (Equations 4.2)
weightings:	area-weighted (**top plots**), unweighted (**bottom plots**)
anomalies:	temp (°C, **left plot**), prec (mm/day, **centre plot**), prec (%, **right plot**)
signal:	2090-99 anomaly (**left bar**), 2070-99 anomaly (**centre & right bars**)
base:	1960-89 mean (**left & centre bars**), first 240y of con (**right bar**)
noise:	140y segments of control (**left & centre bars**)
	30y periods (10y gaps) in detrended final 1160y of control (**right bar**)

Figure 4.6

Time-series of pattern signal-to-noise ratios for individual simulations.

temp (°C) prec (mm/day) prec (%)

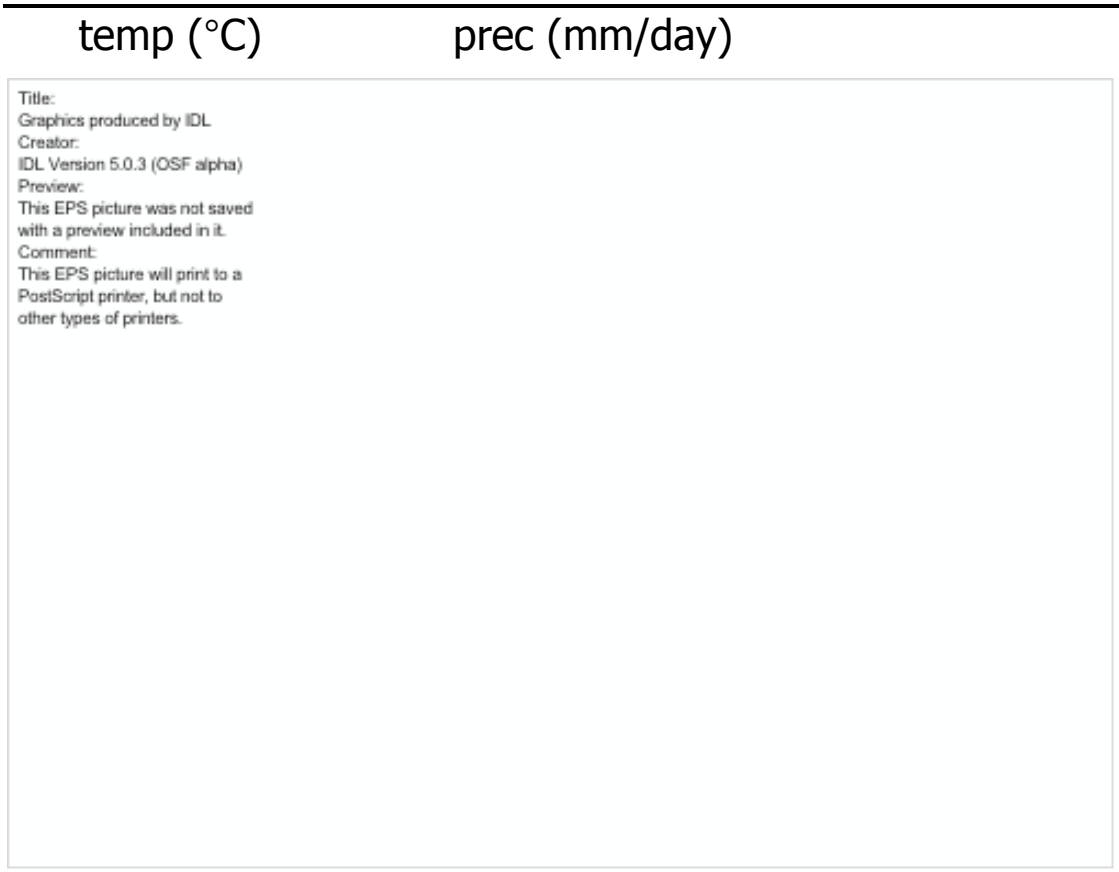


temp (°C) prec (mm/day) prec (%)

model:	HadCM2
simulations:	**Ga 1,2,3,4**; **Gd 1,2,3,4**; **s550**; **s750**
variable:	pattern signal-to-noise ratios (Equations 4.2)
signal:	JJA anomalies of overlapping 30y per., relative to first 240y of con.: temperature (°C, **left**), precip (mm/day, **centre**) precip (%, **right**)
noise:	from 30y periods (10y gaps) in detrended final 1160y of control
x-axis:	period mid-year (**top**), period scaler (°C, **bottom**)

Figure 4.7

Time-series of pattern signal-to-noise ratios for ensembles.

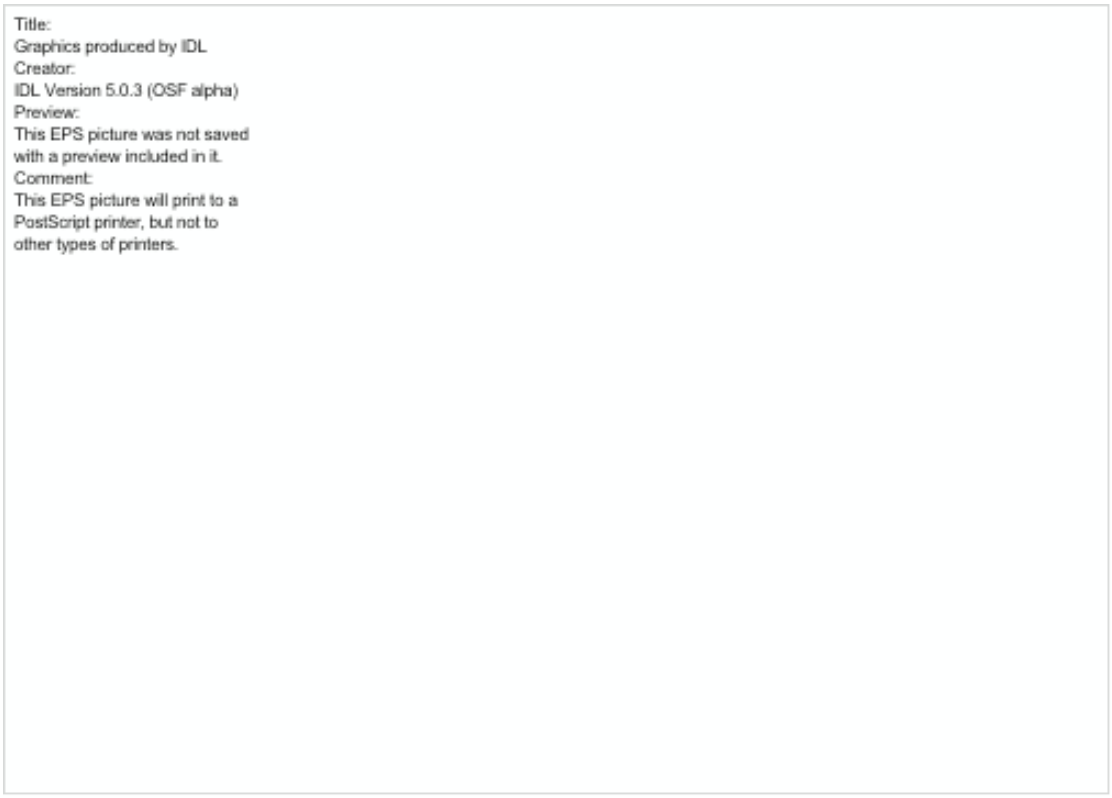


model:	HadCM2
simulations:	**Ga ensemble mean**; **Gd ensemble mean**
variable:	pattern signal-to-noise ratios (Equations 4.2)
signal:	JJA anomalies of overlapping 30y per., relative to first 240y of con.: temperature (°C, **left**), precip (mm/day, **centre**) precip (%, **right**)
noise:	from 30y periods (10y gaps) in detrended final 1160y of control
x-axis:	period mid-year (**top**), period scaler (°C, **bottom**)

Figure 4.8

Grid-box signal-to-noise ratios for ensemble mean temperatures.

Ga 2020s Ga 2080s

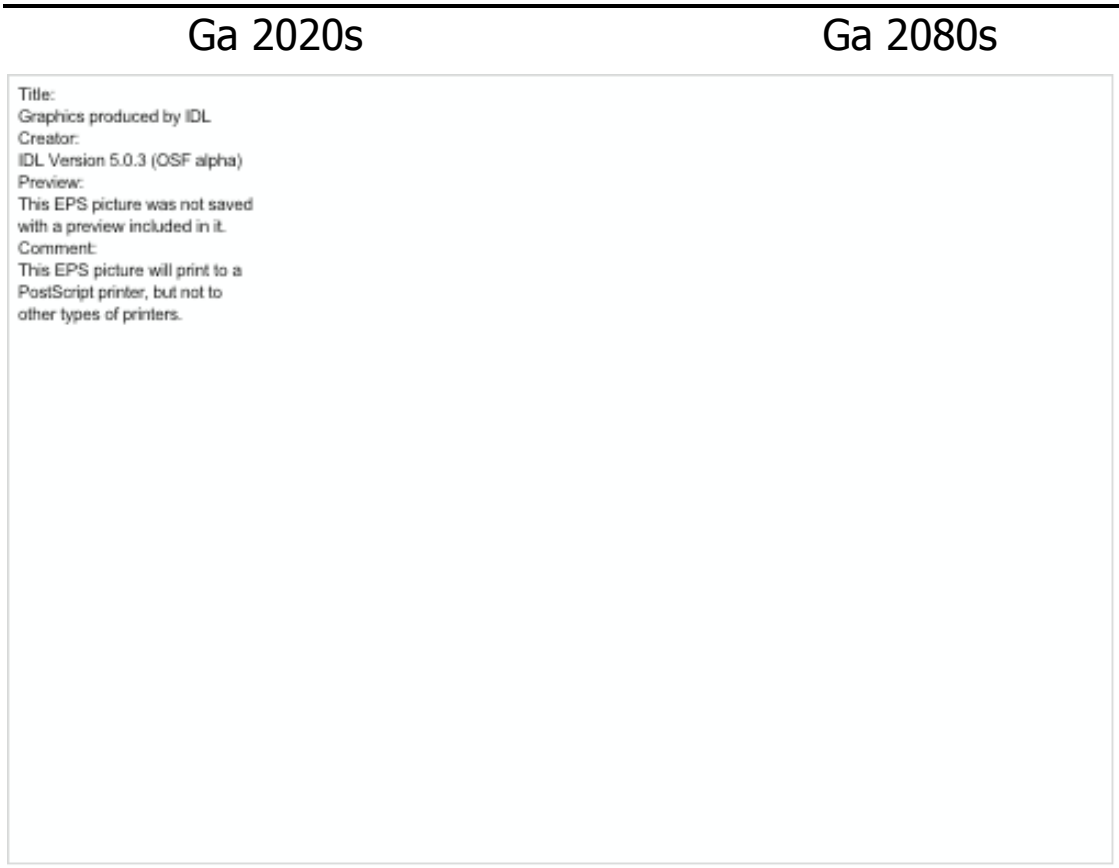


Gd 2020s Gd 2080s

model:	HadCM2
variable:	signal-to-noise ratios for grid-box temperatures (Equations 4.3): Ga ensemble mean (**top**) and Gd ensemble mean (**bottom**) 2020s (**left**) and 2080s (**right**)
signal:	JJA anomalies for 2020s and 2080s, relative to first 240y of control
noise:	from 30y periods (10y gaps) in detrended final 1160y of control

Figure 4.9

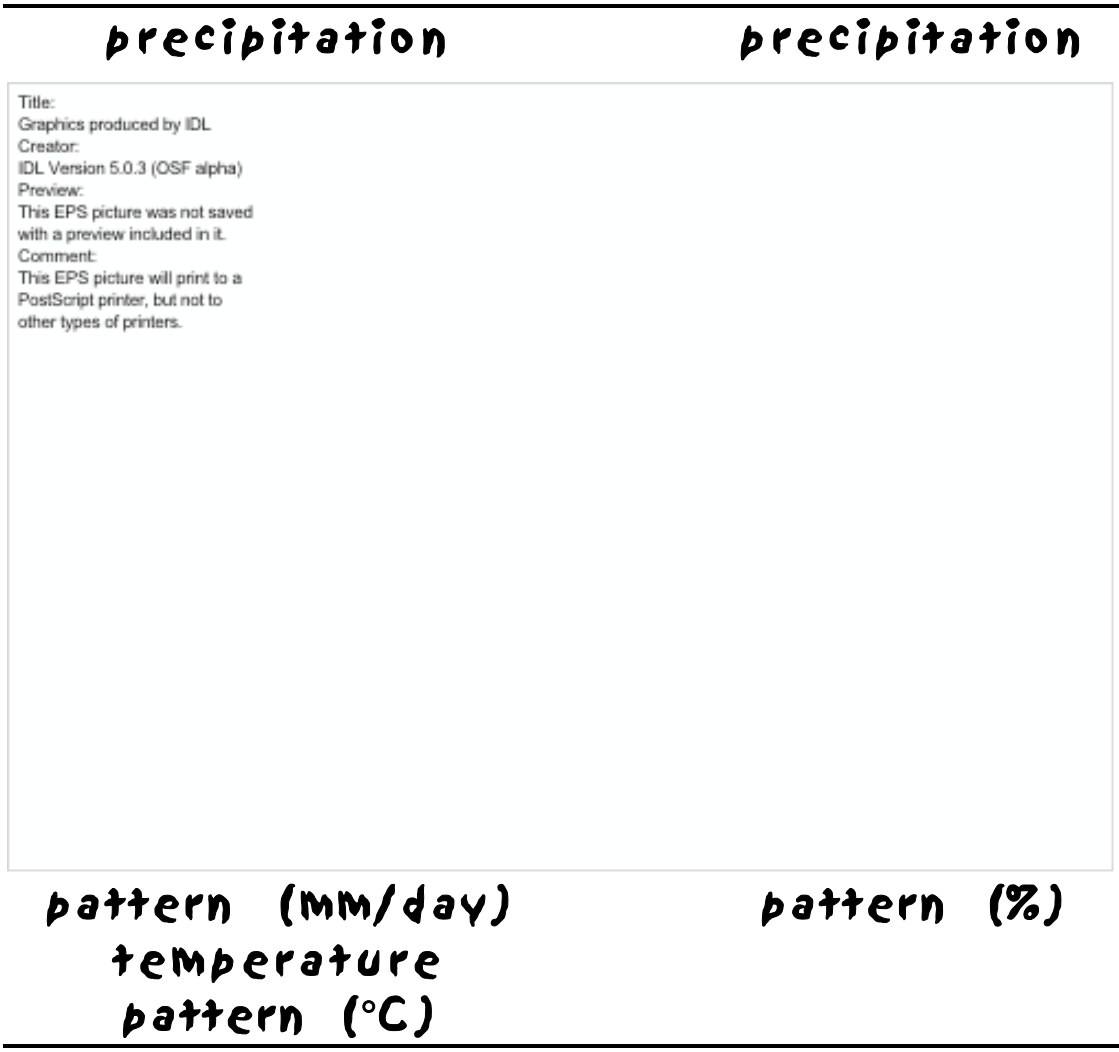
Grid-box signal-to-noise ratios for ensemble mean precipitation (mm/day).



model:	HadCM2
variable:	signal-to-noise ratios for grid-box precipitation (Equations 4.3): Ga ensemble mean (**top**) and Gd ensemble mean (**bottom**) 2020s (**left**) and 2080s (**right**)
signal:	JJA anomalies for 2020s and 2080s, relative to first 240y of control
noise:	from 30y periods (10y gaps) in detrended final 1160y of control

Figure 4.10

Response patterns from Ga for precipitation and temperature.

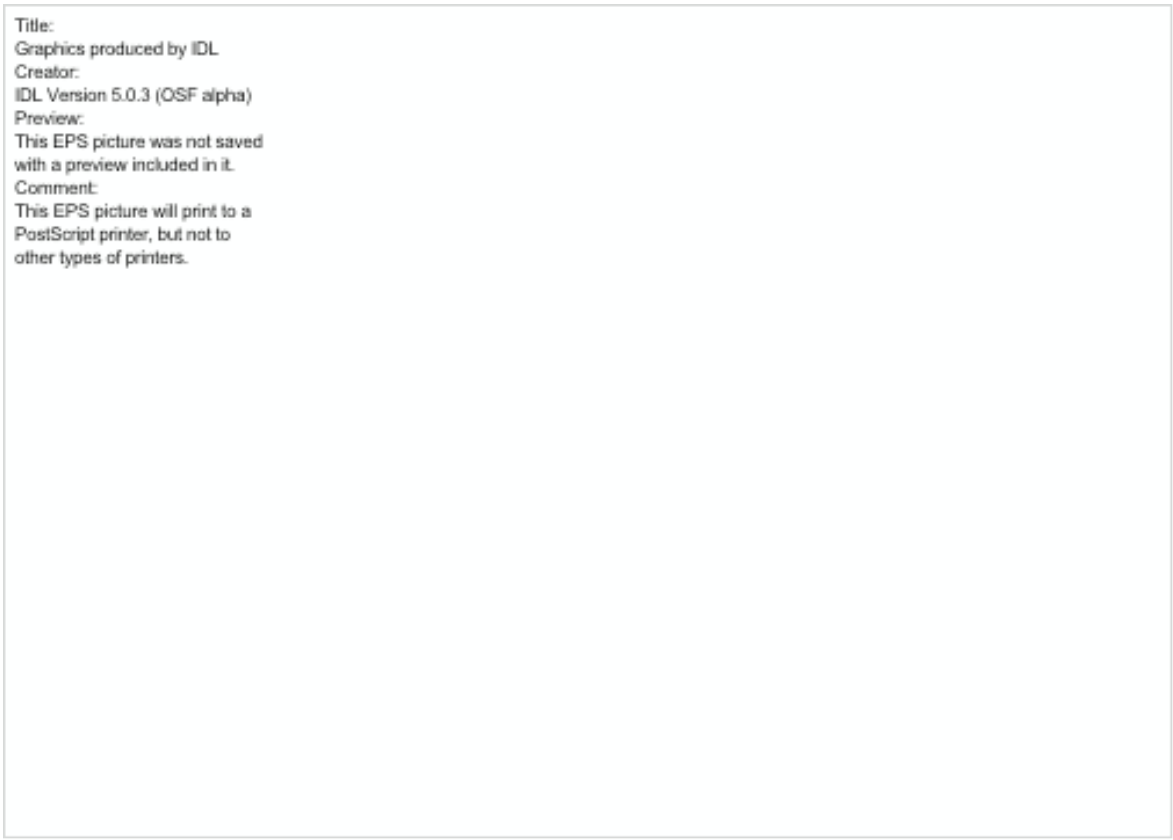


model:	HadCM2
simulations:	Ga 1+2+3+4
anomalies:	JJA for overlapping 30y periods, relative to first 240y of control
pattern:	a from best-fit line ($y=ax$) from anomalies (y) and scaler (x)
variables:	temp (°C, **bottom left**), prec (mm/day, **top left**), prec (%, **top right**)

Figure 4.11

Response patterns for temperature from two different periods in Ga.

2020s response pattern patterns: 2080s – 2020s



significance of top right strengthening of pattern

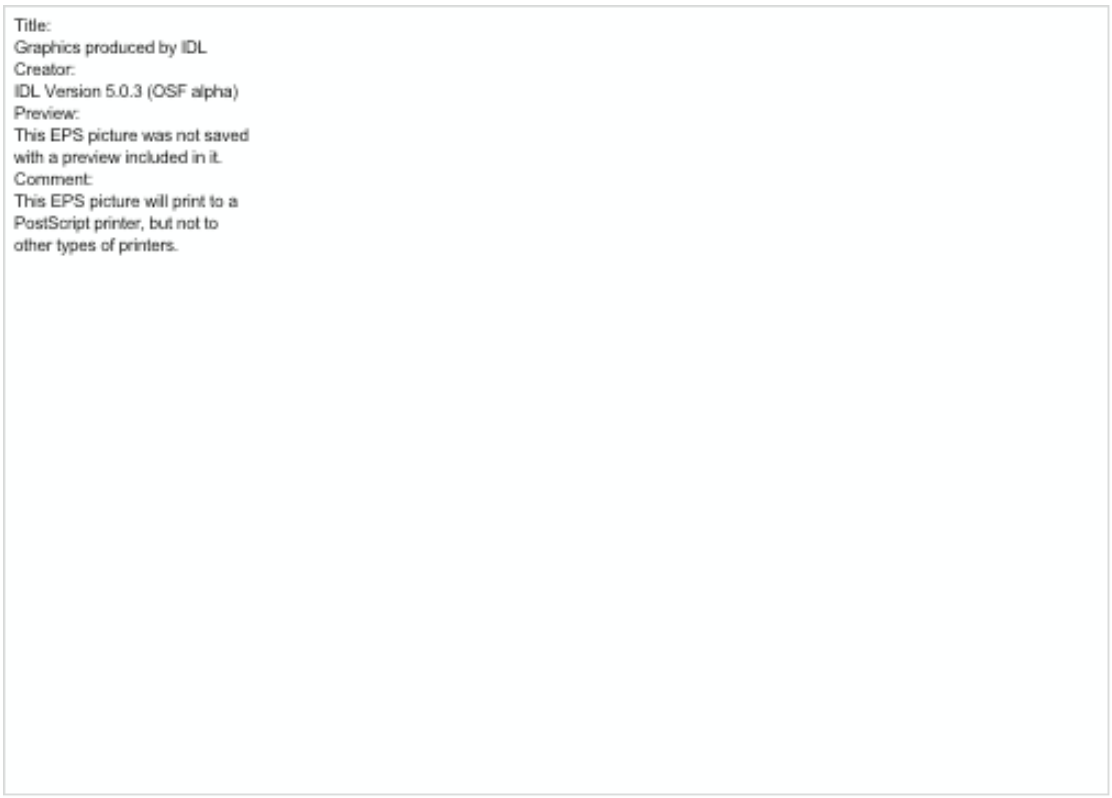
model:	HadCM2
simulations:	Ga 1,2,3,4
anomalies:	JJA temperature (°C), relative to first 240y of control
patterns:	30y anomalies divided by scaler: for 2020s (**top left**) and 2080s
difference:	2080s pattern – 2020s pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**, Eq. 4.4)
masking:	boxes selected where sig > 2□ or sig < -2□ (total: 2918 boxes = 42.8%)
	sig > 2□ in 1825 boxes (26.8%); sig < -2□ in 1093 boxes (16.0%)
adj. patt.:	2020s pattern – 1.0 for each grid box
ratio:	difference divided by adjusted pattern and masked (**bottom right**)
	ratio<-1 (6.9%), -1<ratio<0 (18.3%), 0<ratio (17.6%)
	ratio<-0.2 (19.1%), -0.2<ratio<0.2 (10.6%), 0.2<ratio(13.1%)

Figure 4.12

Differences between temperature response patterns from Ga and other scenarios.

patterns: s550 — Ga

patterns: s750 — Ga



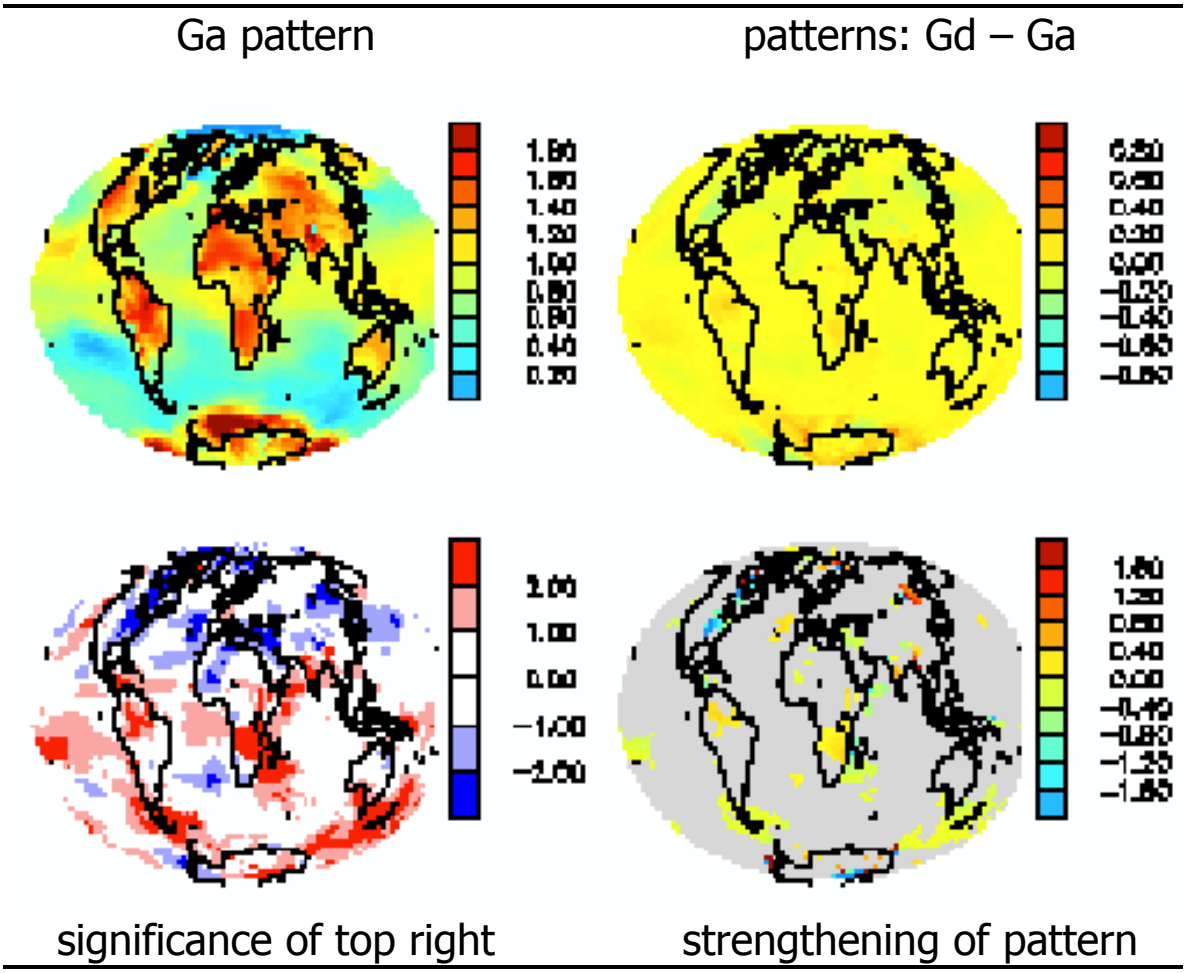
patterns: Gd — Ga

model:	HadCM2
simulations:	Ga 1+2+3+4, Gd 1+2+3+4, s550, s750
anomalies:	JJA temperature (°C), relative to first 240y of control
temporal:	30y periods with equivalent scalers for each comparison[176]
period patt.:	grid-box period mean divided by scaler (°C per °C)
mean patt.:	the mean of all four period patterns from a scenario
difference:	scenario mean pattern -- appropriate Ga pattern (**plotted**)

[176] For s550 (**top left**) we selected the 30-year periods (with global annual temperatures in °C in brackets) centred on 2120 (2.79), 2160 (3.06), 2200 (3.16), and 2240 (3.30). The equivalent periods in Ga were 2045 (2.81), 2052 (3.05), 2056 (3.18), and 2059 (3.29). For s750 (**top right**) we selected 2110 (3.19), 2150 (3.59), 2190 (4.11), and 2230 (4.23). The equivalent periods in Ga were 2056 (3.18), 2068 (3.61), 2082 (4.11), and 2085 (4.23). For Gd (**bottom**) we selected 2073 (2.78), and for Ga we selected 2045 (2.81).

Figure 4.13

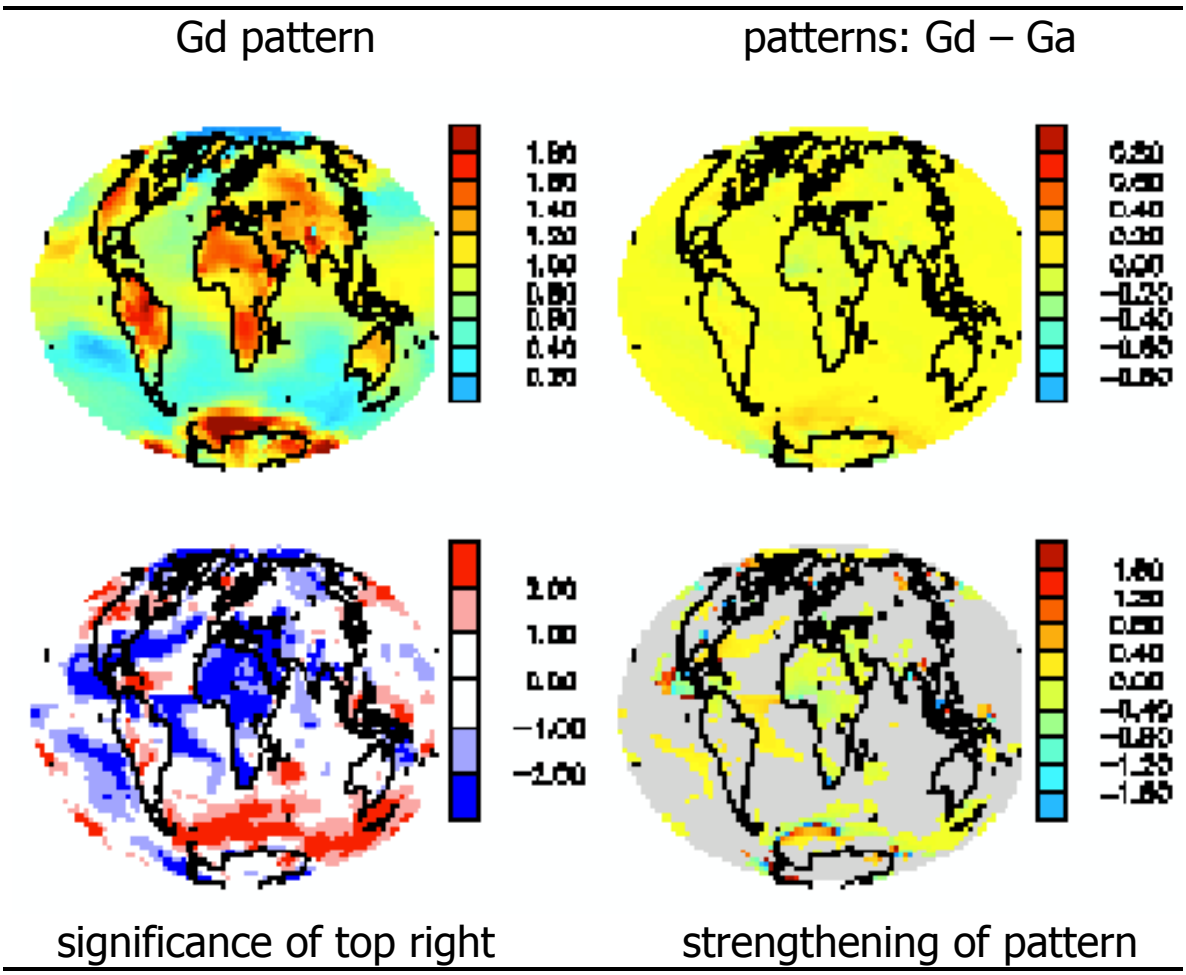
Response patterns for temperature from equivalent periods in Ga and Gd.



model:	HadCM2
simulations:	Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	JJA temperature (°C), relative to first 240y of control
patterns:	30y anomalies divided by scaler: for Ga 2045 (**top left**) and Gd 2073
difference:	Gd 2073 pattern – Ga 2045 pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**, Eq. 4.4)
masking:	boxes selected where sig > 2 or sig < -2 (total: 925 boxes = 13.6%)
	sig > 2 in 615 boxes (8.8%); sig < -2 in 310 boxes (4.5%)
adj. patt.:	Ga pattern – 1.0, for each grid box
ratio:	difference divided by adjusted pattern and masked (**bottom right**)
	ratio<-1 (2.0%), -1<ratio<0 (6.7%), 0<ratio (4.9%)
	ratio<-0.2 (4.6%), -0.2<ratio<0.2 (5.6%), 0.2<ratio(3.4%)

Figure 4.14

Response patterns for temperature from the entire Ga and Gd scenarios.

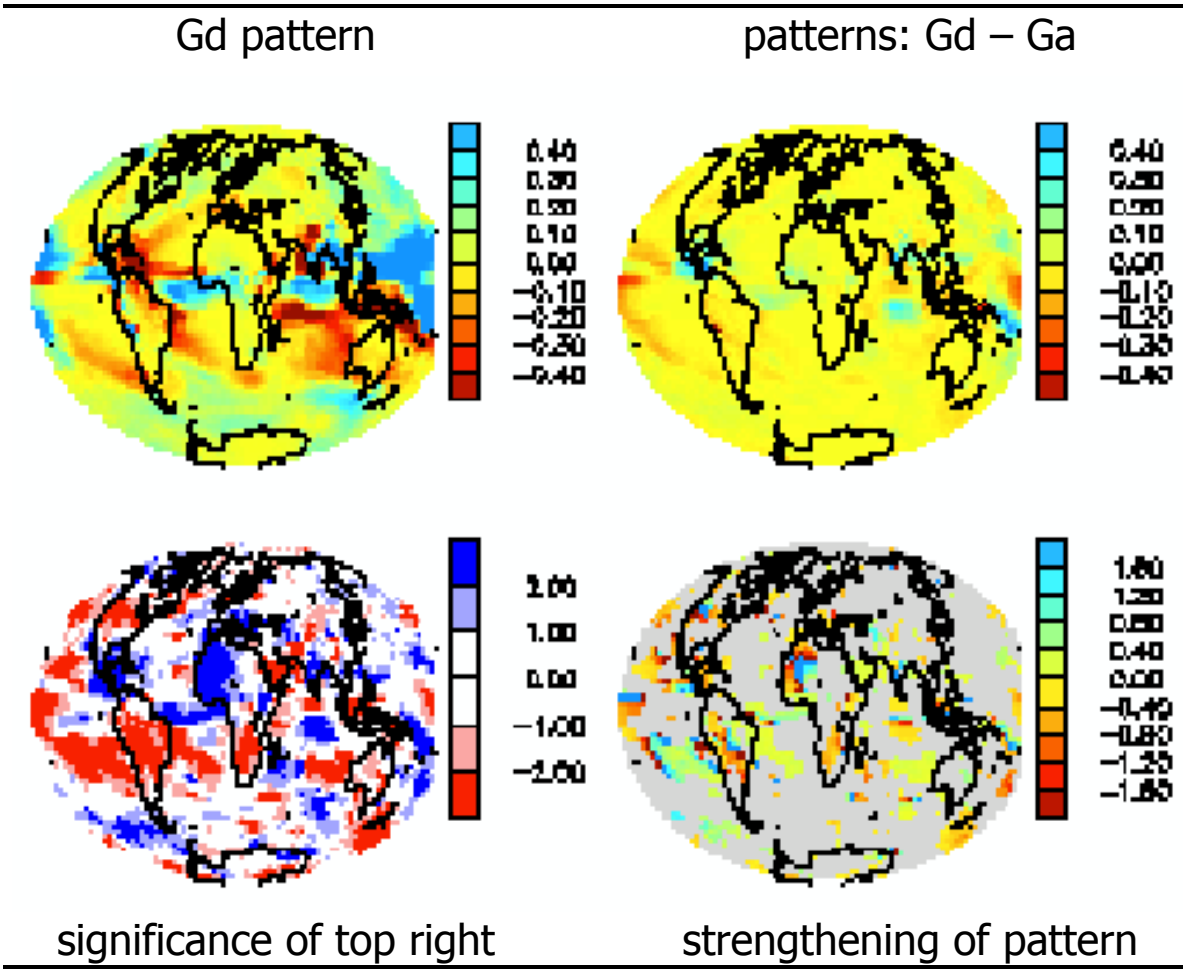


simulations:	HadCM2: Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
pattern:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x): for Ga and Gd (**top left**)
difference:	Gd pattern – Ga pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**)
masking:	boxes selected where sig > 2□ or sig < -2□ (total: 1542 boxes = 22.6%) sig > 2□ in 632 boxes (9.3%); sig < -2□ in 910 boxes (13.4%)
adj. patt.:	Ga pattern – 1.0, for each grid box
ratio:	difference divided by adjusted pattern and masked (**bottom right**)

ratio<-1 (1.5%),	-1<ratio<0 (11.3%),	0<ratio (9.8%)
ratio<-0.2 (6.3%),	-0.2<ratio<0.2 (11.9%),	0.2<ratio(4.3%)

Figure 4.15

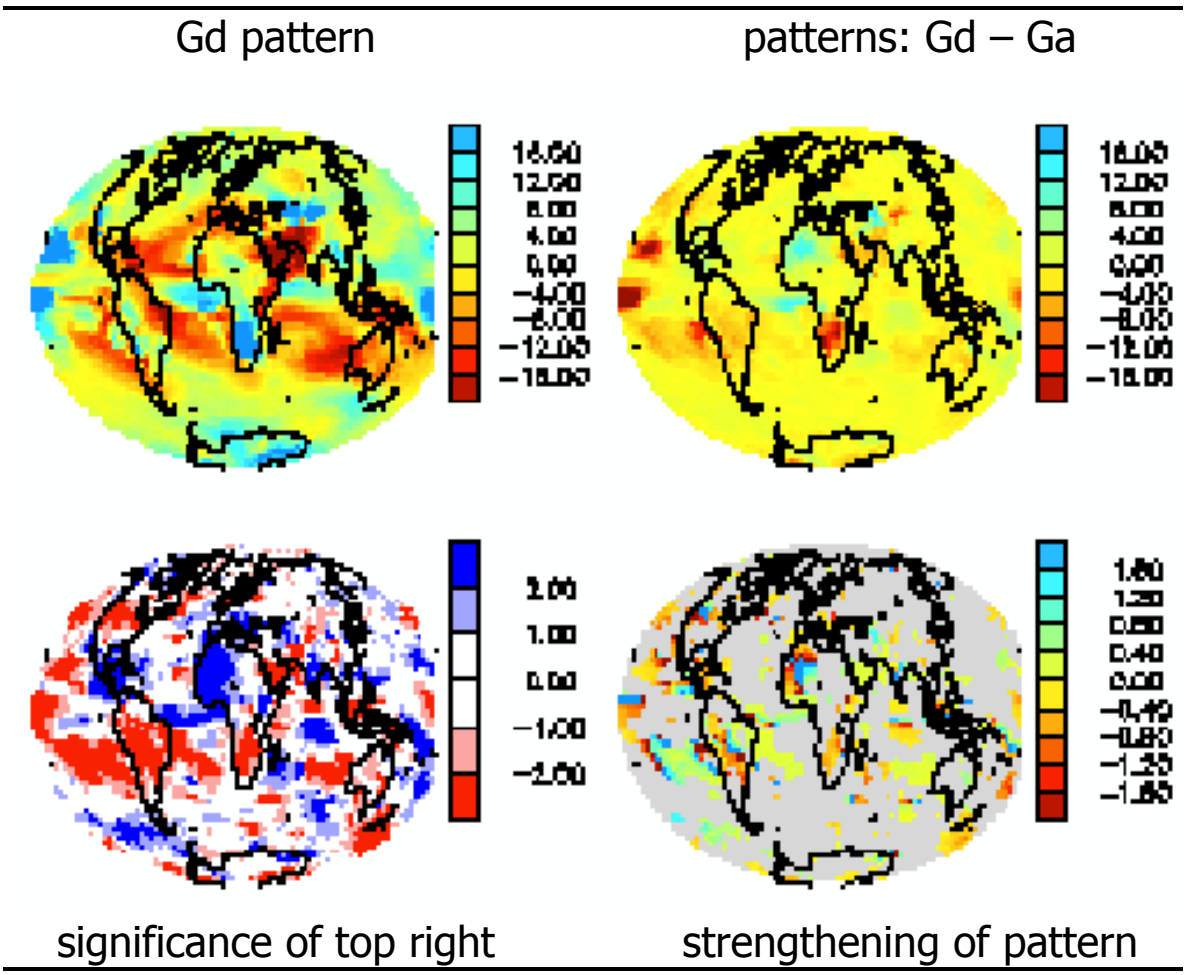
Response patterns for precip. (mm/day) from the entire Ga and Gd scenarios.



simulations:	HadCM2: Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	JJA prec (mm/day) for overlap. 30y per., relative to first 240y of con.
pattern:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x): for Ga and Gd (**top left**)
difference:	Gd pattern – Ga pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**)
masking:	boxes selected where sig > 2□ or sig < -2□ (total: 1545 boxes = 22.7%) sig > 2□ in 606 boxes (8.9%); sig < -2□ in 939 boxes (13.8%)
ratio:	difference divided by Ga pattern and masked (**bottom right**) ratio<-1 (2.3%), -1<ratio<0 (10.0%), 0<ratio (9.9%) ratio<-0.2 (10.4%), -0.2<ratio<0.2 (4.3%), 0.2<ratio(7.5%)

Figure 4.16

Response patterns for precip. (%) from the entire Ga and Gd scenarios.

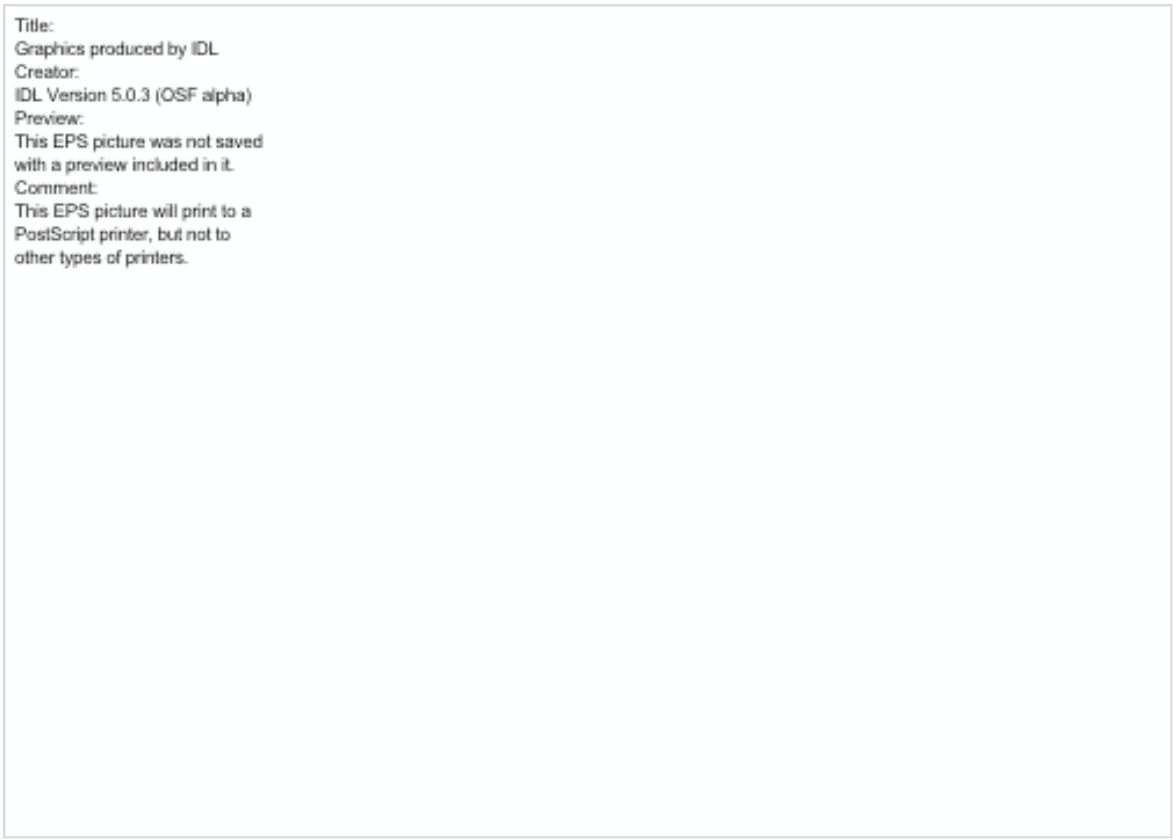


simulations:	HadCM2: Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	JJA prec (%) for overlapping 30y per., relative to first 240y of con.
pattern:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x): for Ga and Gd (**top left**)
difference:	Gd pattern – Ga pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**)
masking:	boxes selected where sig > 2□ or sig < -2□ (total: 1532 boxes = 22.5%) sig > 2□ in 604 boxes (8.9%); sig < -2□ in 928 boxes (13.6%)
ratio:	difference divided by Ga pattern and masked (**bottom right**) ratio<-1 (2.3%), -1<ratio<0 (10.1%), 0<ratio (10.0%) ratio<-0.2 (10.5%), -0.2<ratio<0.2 (4.5%), 0.2<ratio(7.5%)

Figure 4.17

Response patterns for precipitation from two different periods in Ga.

2020s response pattern patterns: 2080s – 2020s



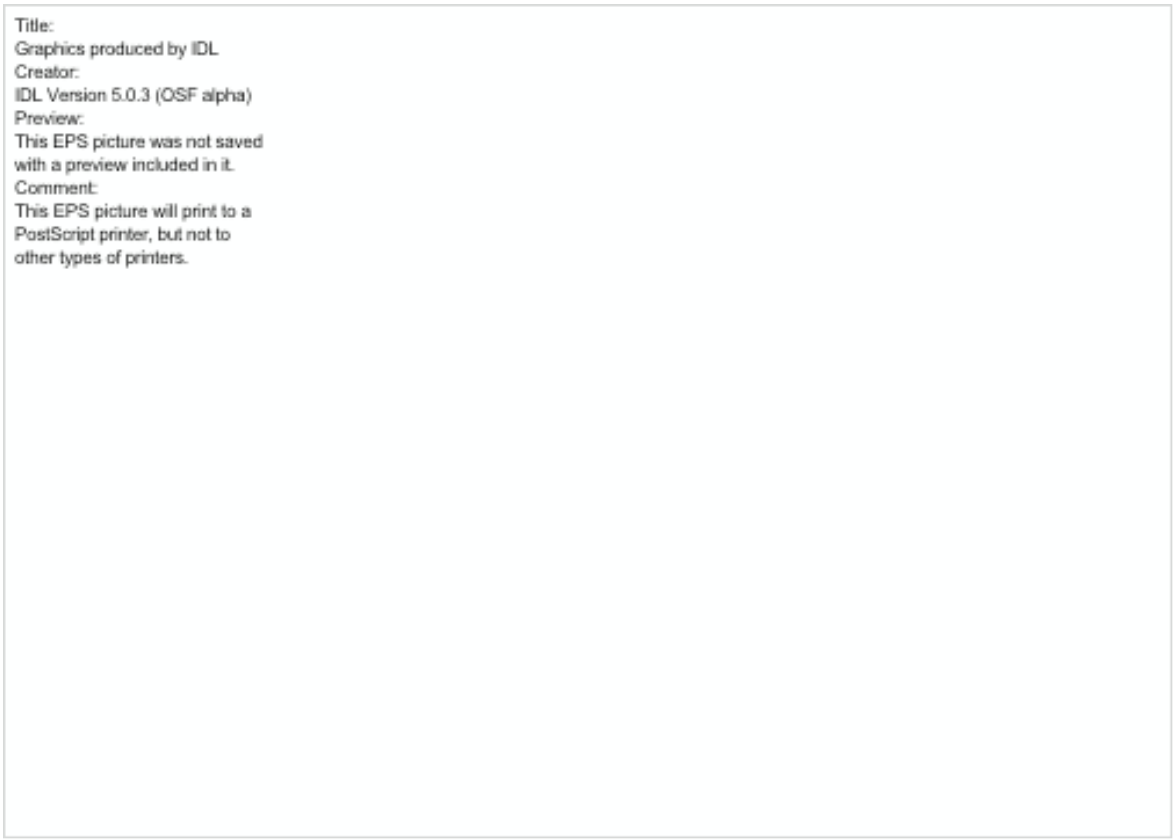
significance of top right strengthening of pattern

model:	HadCM2
simulations:	Ga 1,2,3,4
anomalies:	JJA precipitation (mm/day), relative to first 240y of control
patterns:	30y anomalies divided by scaler: for 2020s (**top left**) and 2080s
difference:	2080s pattern – 2020s pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**, Eq. 4.4)
masking:	boxes selected where sig > 2σ or sig < -2σ (total: 2316 boxes = 34.0%)
	sig > 2σ in 1427 boxes (20.9%); sig < -2σ in 889 boxes (13.0%)
ratio:	difference divided by 2085 pattern and masked (**bottom right**)
	ratio<-1 (7.4%), -1<ratio<0 (10.4%), 0<ratio (16.2%)
	ratio<-0.2 (17.0%), -0.2<ratio<0.2 (0.9%), 0.2<ratio(16.0%)

Figure 4.18

Response patterns for precipitation from two equivalent periods in Ga and Gd.

Ga response pattern — patterns: Gd – Ga

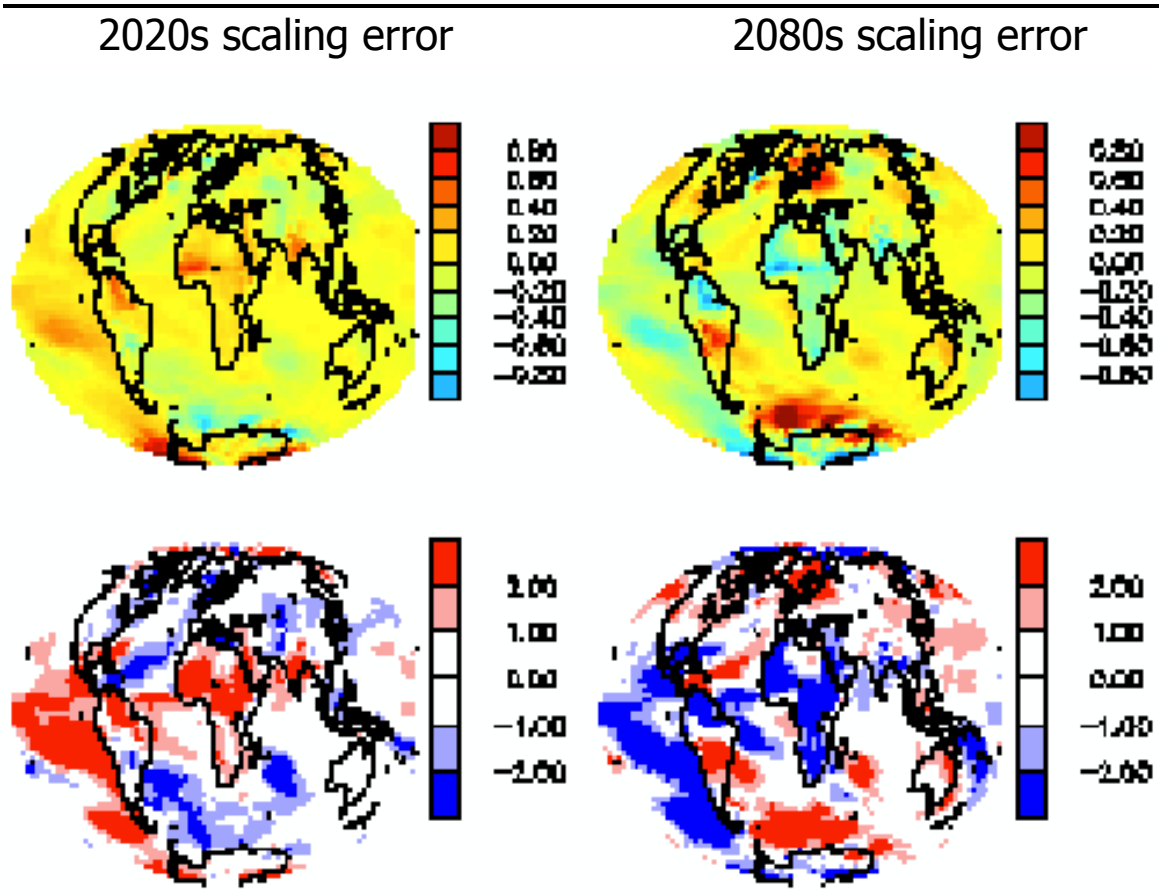


significance of top right — strengthening of pattern

model:	HadCM2
simulations:	Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	JJA temperature (°C), relative to first 240y of control
patterns:	30y anomalies divided by scaler: for Ga 2045 (**top left**) and Gd 2073
difference:	Gd 2073 pattern – Ga 2045 pattern (**top right**)
significance:	difference expressed in standard deviations (**bottom left**, Eq. 4.4)
masking:	boxes selected where sig > 2□ or sig < -2□ (total: 598 boxes = 8.8%)
	sig > 2□ in 435 boxes (6.4%); sig < -2□ in 163 boxes (2.4%)
ratio:	difference divided by Ga 2045 pattern and masked (**bottom right**)
	ratio<-1 (1.0%), -1<ratio<0 (2.5%), 0<ratio (5.3%)
	ratio<-0.2 (3.2%), -0.2<ratio<0.2 (0.7%), 0.2<ratio(4.8%)

Figure 4.19

Errors in estimating Ga temperature from scaling the Ga pattern.

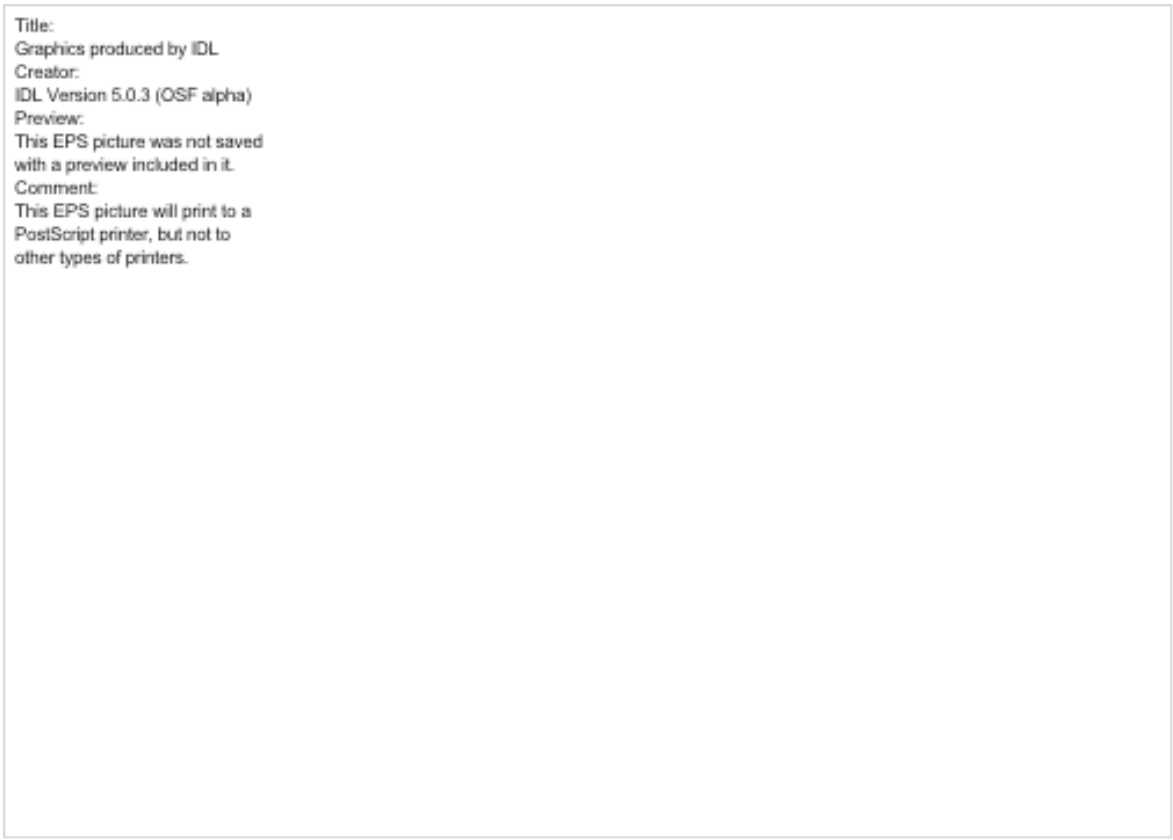


simulations:	HadCM2 Ga 1,2,3,4
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
pattern:	a from best-fit line ($y=ax$) from anomalies (y) and scaler (x)
estimate:	pattern * scaler for Ga 2020s and 2080s
error:	estimate – modelled value: 2020s (**top left**), 2080s (**top right**)
significance:	error expressed as □ (Equations 4.5, □$_M$ from control)
	2020s (**bottom left**): 1210 (17.8%) of boxes where mag. > 2□
	2080s (**bottom right**): 1876 (27.5%) of boxes where mag. > 2□

Figure 4.20

Errors in estimating 2080s Ga temperature from scaling the Ga pattern.

significance with control □ proportion (mask=top left)



significance with Ga □ proportion (mask=bott. left)

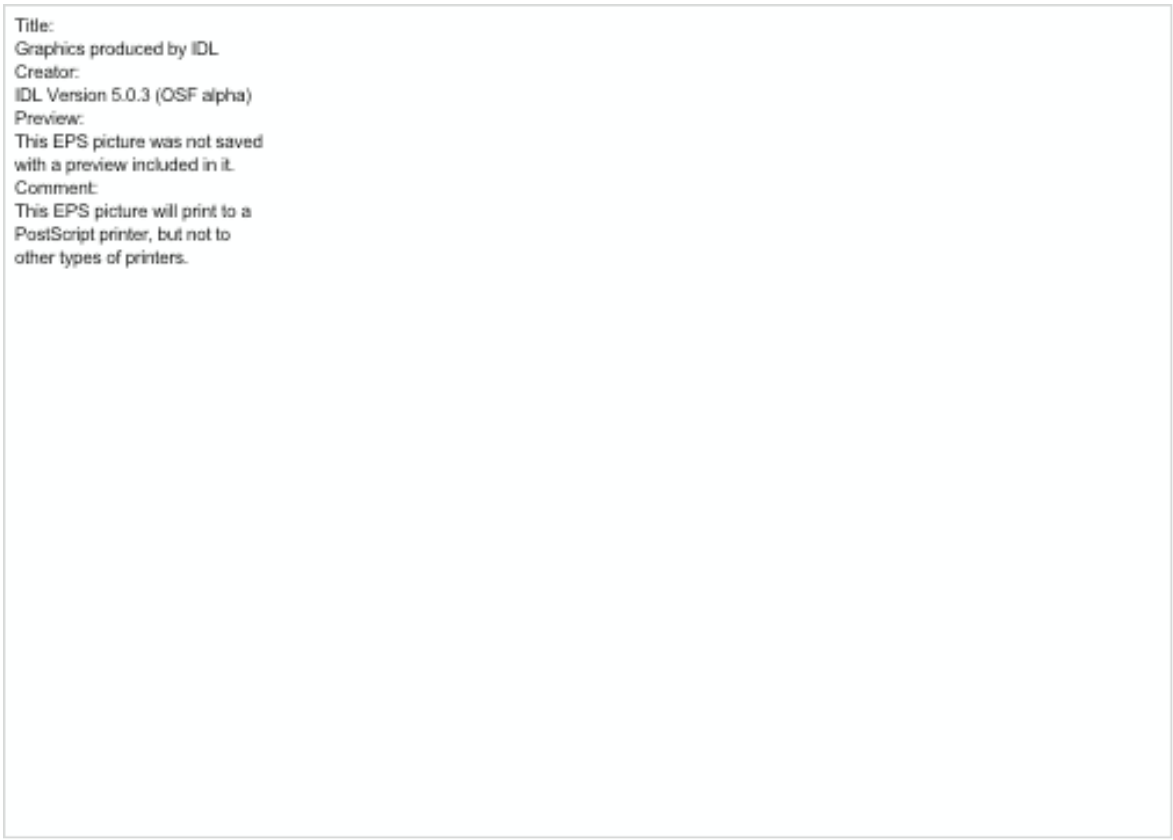
errors:	as calculated in Figure 4.19 (top right)		
significance:	error expressed as □ (Equations 4.5)		
	$□_M$ from control:	1876 (27.5%) of boxes >2□ or <-2□ (**top left**)[177]	
	$□_M$ from ensemble:	2123 (31.1%) of boxes >2□ or <-2□ (**bott. left**)	
ratio:	error / modelled: masked by left-hand plots (**right plots**)		
(**top right**)	ratio<-1 (0.2%),	-1<ratio<0 (16.3%),	0<ratio (11.0%)
	ratio<-0.2 (3.5%),	-0.2<ratio<0.2 (21.9%),	0.2<ratio(2.1%)
(**bott. right**)	ratio<-1 (0.3%),	-1<ratio<0 (17.0%),	0<ratio (13.9%)
	ratio<-0.2 (3.6%),	-0.2<ratio<0.2 (25.3%),	0.2<ratio(2.3%)

[177] This plot is identical to Figure 4.19 (bottom right).

Figure 4.21

Errors in estimating 2080s Ga precipitation from scaling the Ga pattern.

significance with control □ proportion (mask=top left)



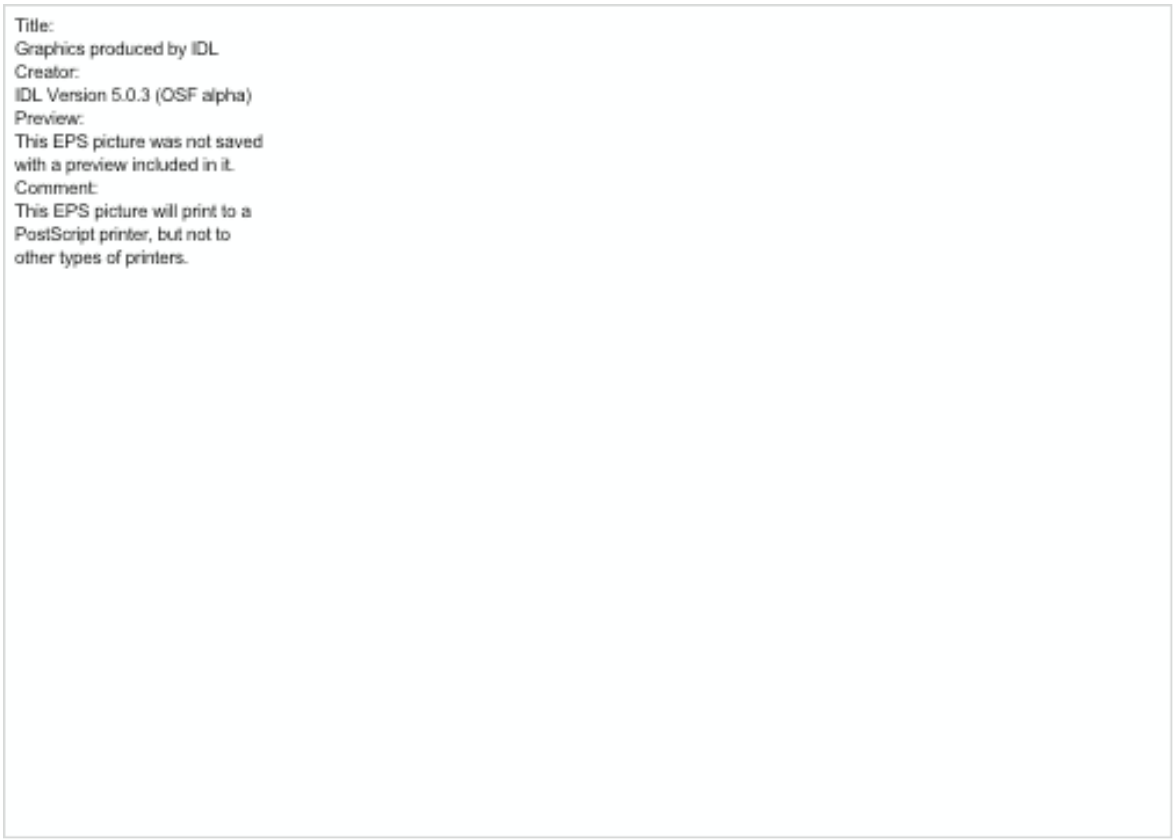
significance with Ga □ proportion (mask=bott. left)

errors:	as calculated in Figure 4.19 (top right), but for prec (mm/day)		
significance:	error expressed as □ (Equations 4.5)		
	$□_M$ from control:	1659 (24.3%) of boxes >2□ or <-2□ (**top left**)	
	$□_M$ from ensemble:	1780 (26.1%) of boxes >2□ or <-2□ (**bott. left**)	
ratio:	error / modelled: masked by left-hand plots (**right plots**)		
(**top right**)	ratio<-1 (2.4%),	-1<ratio<0 (14.5%),	0<ratio (7.5%)
	ratio<-0.2 (14.6%),	-0.2<ratio<0.2 (3.1%),	0.2<ratio(6.6%)
(**bott. right**)	ratio<-1 (2.4%),	-1<ratio<0 (14.8%),	0<ratio (9.0%)
	ratio<-0.2 (14.3%),	-0.2<ratio<0.2 (4.8%),	0.2<ratio(7.0%)

Figure 4.22

Errors in estimating 2080s Gd temperatures from scaling the Ga pattern.

modelled | estimate



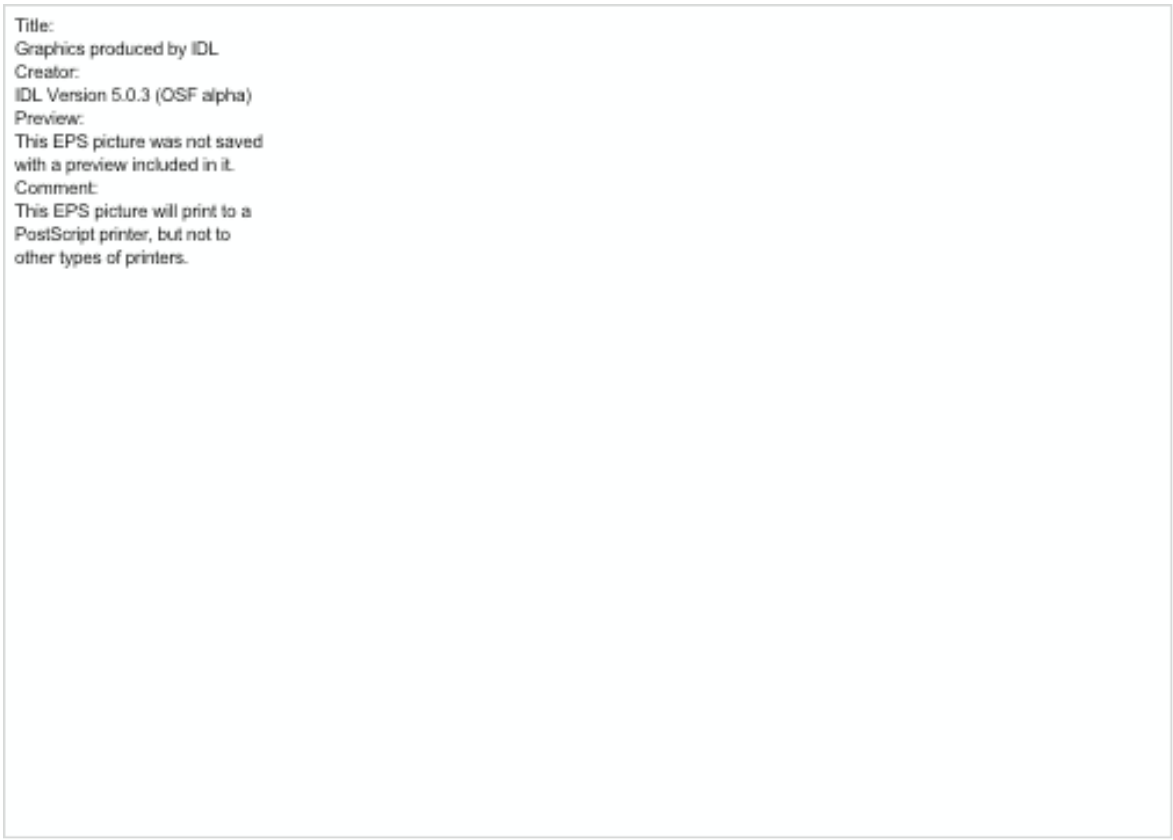
significance of error | error / modelled

simulations:	HadCM2: **Ga 1,2,3,4**; **Gd 1,2,3,4**
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
modelled:	as simulated, Gd in 2080s (**top left**)
pattern:	a from best-fit line ($y=ax$) from Ga anomalies (y) and scaler (x)
estimate:	Ga pattern * Gd 2080s scaler (**top right**)
error:	estimate – modelled
significance:	error expressed as □ (Equations 4.5, $\square_M$ from control) (**bottom left**)
ratio:	error / modelled, masked by error significance (**bottom right**)
	ratio<-1 (0.1%), -1<ratio<0 (7.8%), 0<ratio (5.9%)
	ratio<-0.2 (0.2%), -0.2<ratio<0.2 (12.0%), 0.2<ratio(1.6%)

Figure 4.23

Errors in estimating 2080s Gd precipitation from scaling the Ga pattern.

modelled | estimate



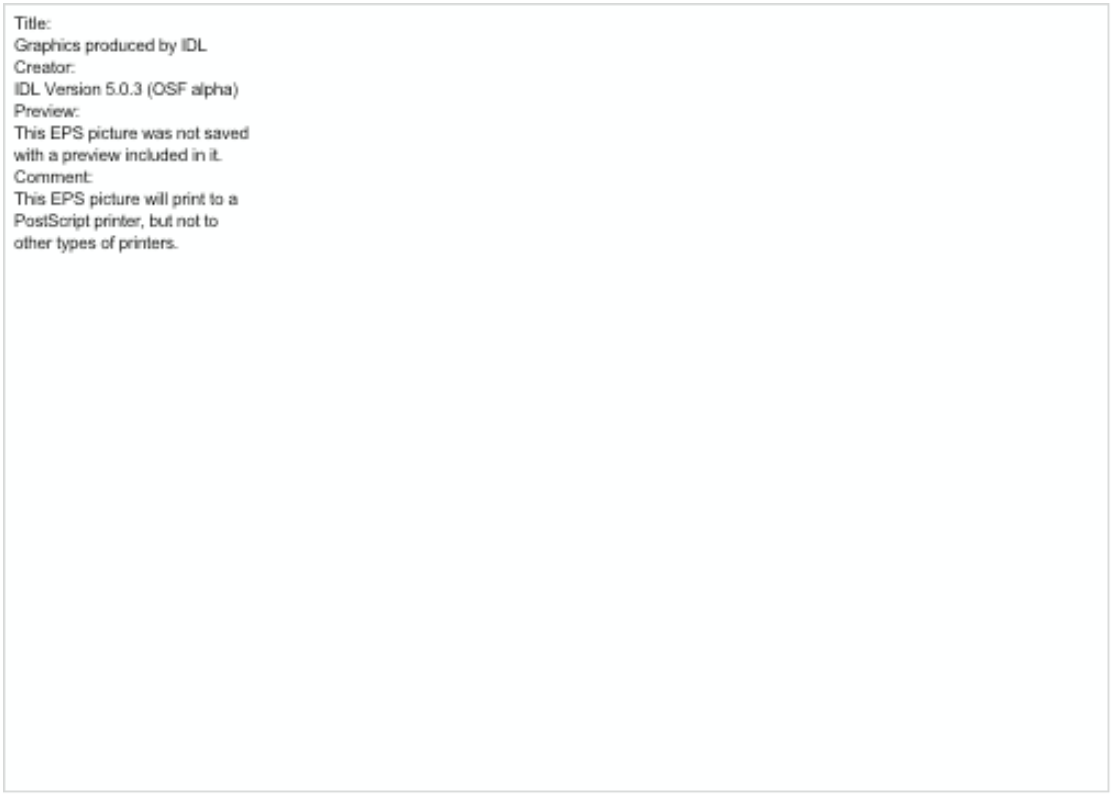
significance of error | error / modelled

simulations:	HadCM2: **Ga 1,2,3,4**; **Gd 1,2,3,4**
anomalies:	JJA prec (mm/day) for overlap. 30y per., relative to first 240y of con.
modelled:	as simulated, Gd in 2080s (**top left**)
pattern:	a from best-fit line ($y=ax$) from Ga anomalies (y) and scaler (x)
estimate:	Ga pattern * Gd 2080s scaler (**top right**)
error:	estimate – modelled
significance:	error expressed as □ (Equations 4.5, □$_M$ from control) (**bottom left**)
ratio:	error / modelled, masked by error significance (**bottom right**)
	ratio<-1 (1.5%), -1<ratio<0 (5.8%), 0<ratio (3.3%)
	ratio<-0.2 (5.6%), -0.2<ratio<0.2 (1.8%), 0.2<ratio(3.3%)

Figure 4.24

Changes in multi-decadal grid-box variability in Ga and Gd.

temperature sd precipitation sd precipitation cv



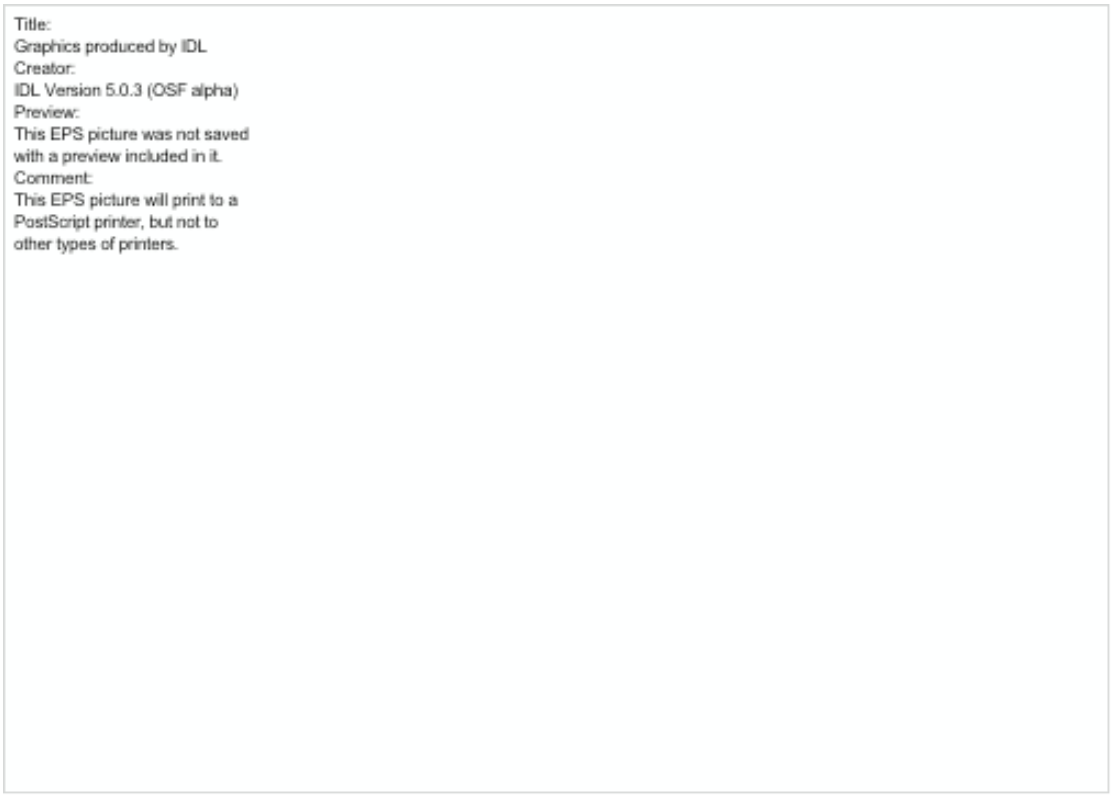
temperature sd precipitation sd precipitation cv

model:	HadCM2
simulations:	**1400y control**; Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	JJA means of overlapping 30-year per., relative to first 240y of con.
grid-box sd:	pop. standard deviation for each period in Ga 1,2,3,4, Gd 1,2,3,4
	pop. standard deviation for 30y periods (10y gaps) in **detrended con.**
global sd:	grid-box sd averaged for temp (°C, **left**) and prec (mm/day, **centre**)
grid-box cv:	grid-box sd / mean for each period in Ga 1,2,3,4, Gd 1,2,3,4
	grid-box sd / mean for 30y periods (10y gaps) in **detrended con.**
global cv:	grid-box cv averaged for prec (mm/day, **right**)

Figure 4.25

Errors in estimating Ga temperatures from scaling various patterns.

mean error | error sizes | significances



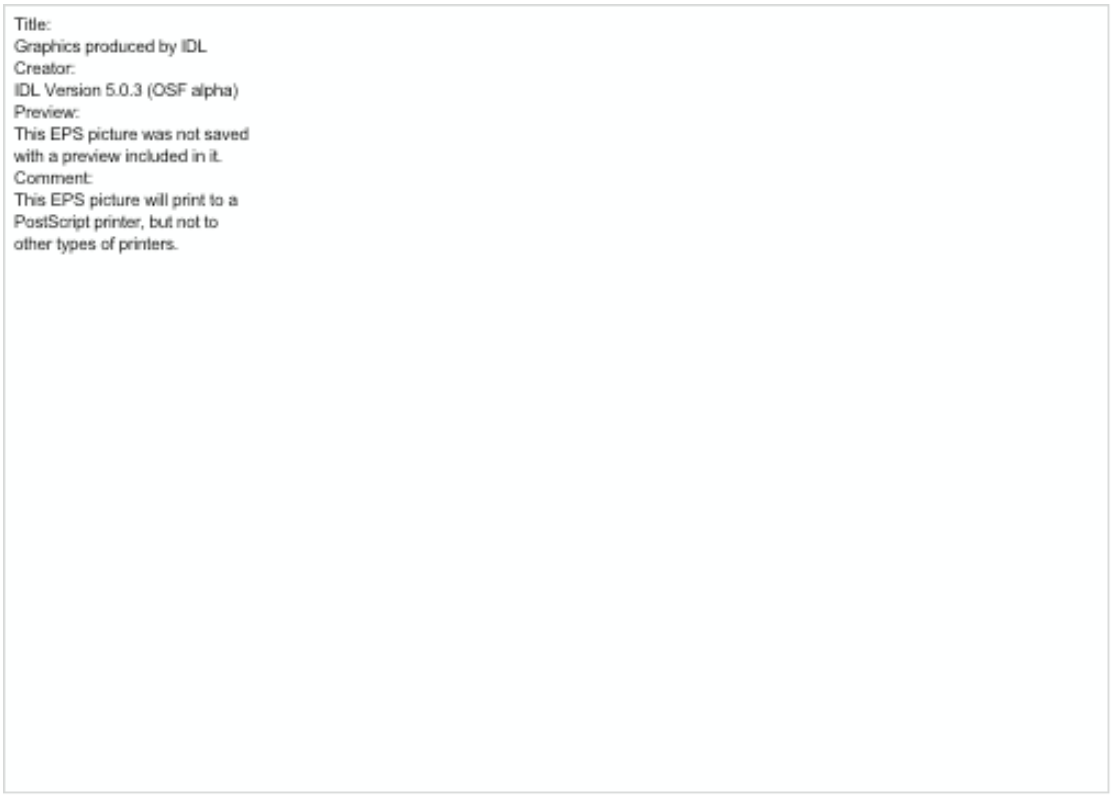
mean error | error sizes | significances

model:	HadCM2
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
pattern:	a from best-fit line (y=ax) from anomalies (y) and scaler (x): **Ga 1,2,3,4**; **Gd 1,2,3,4**; **s550**; **s750**
estimate:	for overlapping 30y periods in Ga: each pattern * Ga scaler
box error:	estimate – modelled value, for each box
mean error:	global-mean of box errors (**left**)
error sizes:	average magnitude of box errors (**centre**)
significance:	significance of each box error (Eq. 4.5, $\square_M$ from con.), averaged (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.26

Errors in estimating Gd temperatures from scaling various patterns.

mean error | error sizes | significances



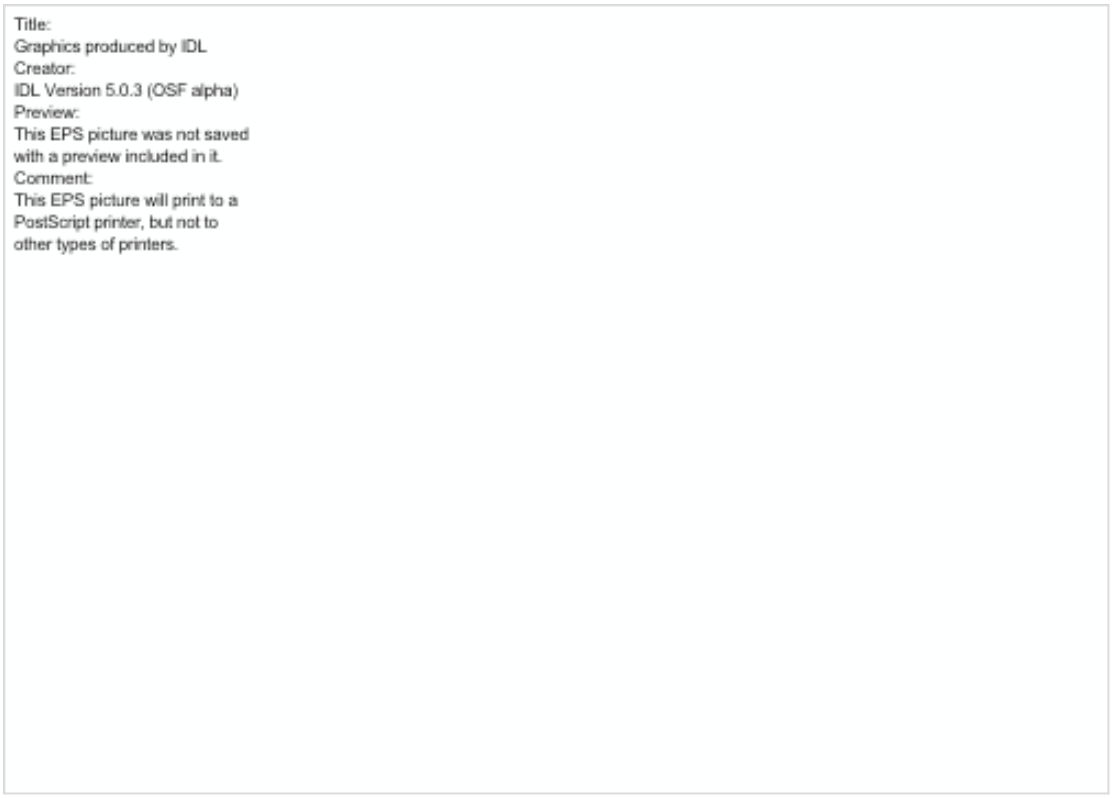
mean error | error sizes | significances

model:	HadCM2
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
pattern:	a from best-fit line (y=ax) from anomalies (y) and scaler (x): **Ga 1,2,3,4**; **Gd 1,2,3,4**; **s550**; **s750**
estimate:	for overlapping 30y periods in Gd: each pattern * Gd scaler
box error:	estimate – modelled value, for each box
mean error:	global-mean of box errors (**left**)
error sizes:	average magnitude of box errors (**centre**)
significance:	significance of each box error (Eq. 4.5, $\square_M$ from con.), averaged (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.27

Errors in estimating Ga precipitation from scaling various patterns.

mean error | error sizes | significances



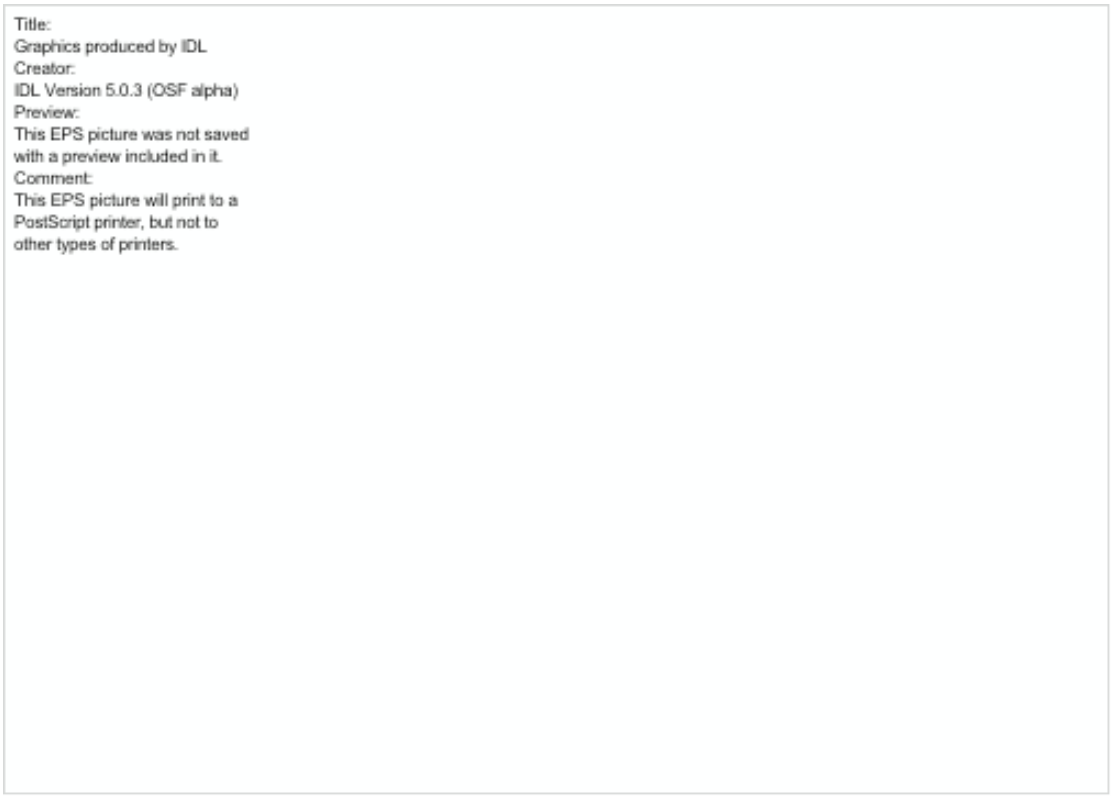
mean error | error sizes | significances

model:	HadCM2
anomalies:	JJA prec (mm/day) for overlap. 30y per., relative to first 240y of con.
pattern:	a from best-fit line (y=ax) from anomalies (y) and scaler (x): **Ga 1,2,3,4**; **Gd 1,2,3,4**; **s550**; **s750**
estimate:	for overlapping 30y periods in Ga: each pattern * Ga scaler
box error:	estimate – modelled value, for each box
mean error:	global-mean of box errors (**left**)
error sizes:	average magnitude of box errors (**centre**)
significance:	significance of each box error (Eq. 4.5, $\square_M$ from con.), averaged (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.28

Errors in estimating Gd precipitation from scaling various patterns.

mean error error sizes significances



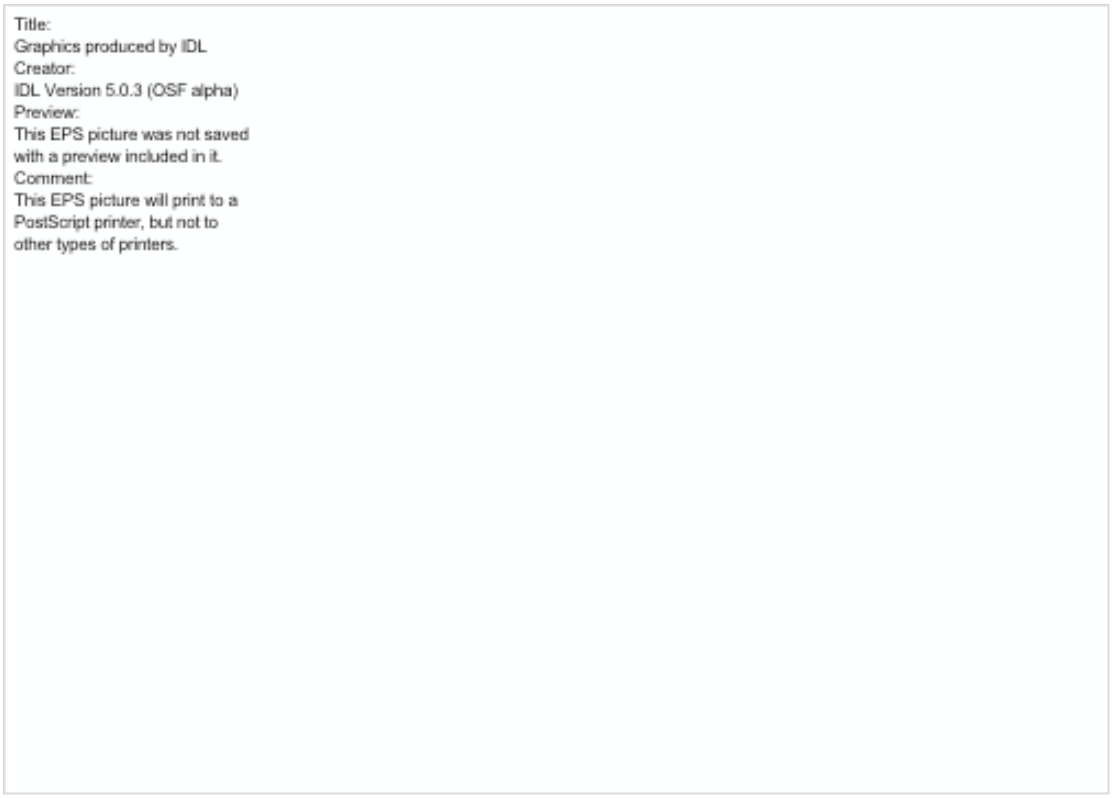
mean error error sizes significances

model:	HadCM2
anomalies:	JJA prec (mm/day) for overlap. 30y per., relative to first 240y of con.
pattern:	a from best-fit line (y=ax) from anomalies (y) and scaler (x): **Ga 1,2,3,4**; **Gd 1,2,3,4**; **s550**; **s750**
estimate:	for overlapping 30y periods in Gd: each pattern * Gd scaler
box error:	estimate – modelled value, for each box
mean error:	global-mean of box errors (**left**)
error sizes:	average magnitude of box errors (**centre**)
significance:	significance of each box error (Eq. 4.5, $\square_M$ from con.), averaged (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.29

Errors in estimating Ga and Gd temperatures under global-mean hypothesis.

estimating Ga: error sizes estimating Gd: error sizes



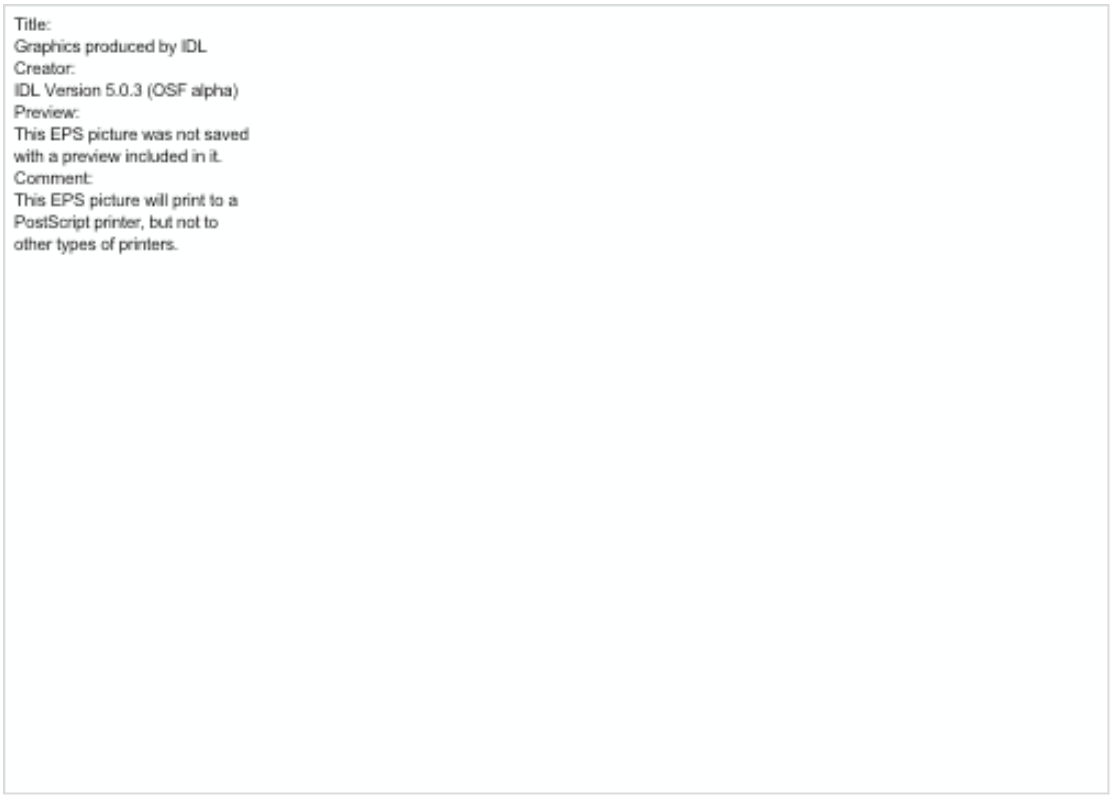
estimating Ga: error sizes estimating Gd: error sizes

model:	HadCM2
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
patterns:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x):
	Ga 1,2,3,4; **Gd 1,2,3,4**; **s550**; **s750**
	plus pattern where all boxes =0.86 (**global-mean**)
Ga estimate:	for overlapping 30y periods in Ga: each pattern * Ga scaler
Gd estimate:	for overlapping 30y periods in Gd: each pattern * Gd scaler
Ga box error:	Ga estimate – Ga modelled value, for each box
Gd box error:	Gd estimate – Gd modelled value, for each box
error sizes:	average magnitude of Ga box errors (**left**), Gd box errors (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.30

Errors in estimating Ga and Gd precipitation under global-mean hypothesis.

estimating Ga: error sizes estimating Gd: error sizes



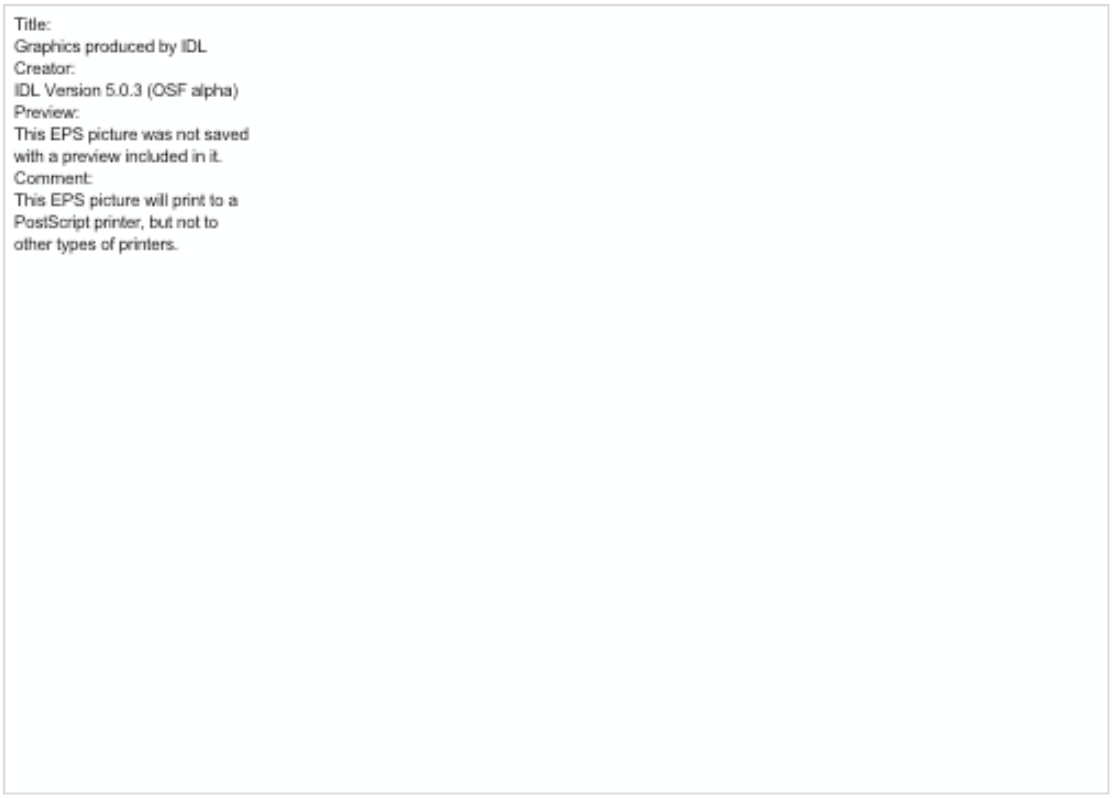
estimating Ga: error sizes estimating Gd: error sizes

model:	HadCM2
anomalies:	JJA prec (mm/day) for overlap. 30y per., relative to first 240y of con.
patterns:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x):
	Ga 1,2,3,4; **Gd 1,2,3,4**; **s550**; **s750**
	plus pattern where all boxes =0.043 (**global-mean**)
Ga estimate:	for overlapping 30y periods in Ga: each pattern * Ga scaler
Gd estimate:	for overlapping 30y periods in Gd: each pattern * Gd scaler
Ga box error:	Ga estimate – Ga modelled value, for each box
Gd box error:	Gd estimate – Gd modelled value, for each box
error sizes:	average magnitude of Ga box errors (**left**), Gd box errors (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.31

Errors in estimating Ga and Gd temperatures from using single simulations.

estimating Ga: error sizes estimating Gd: error sizes



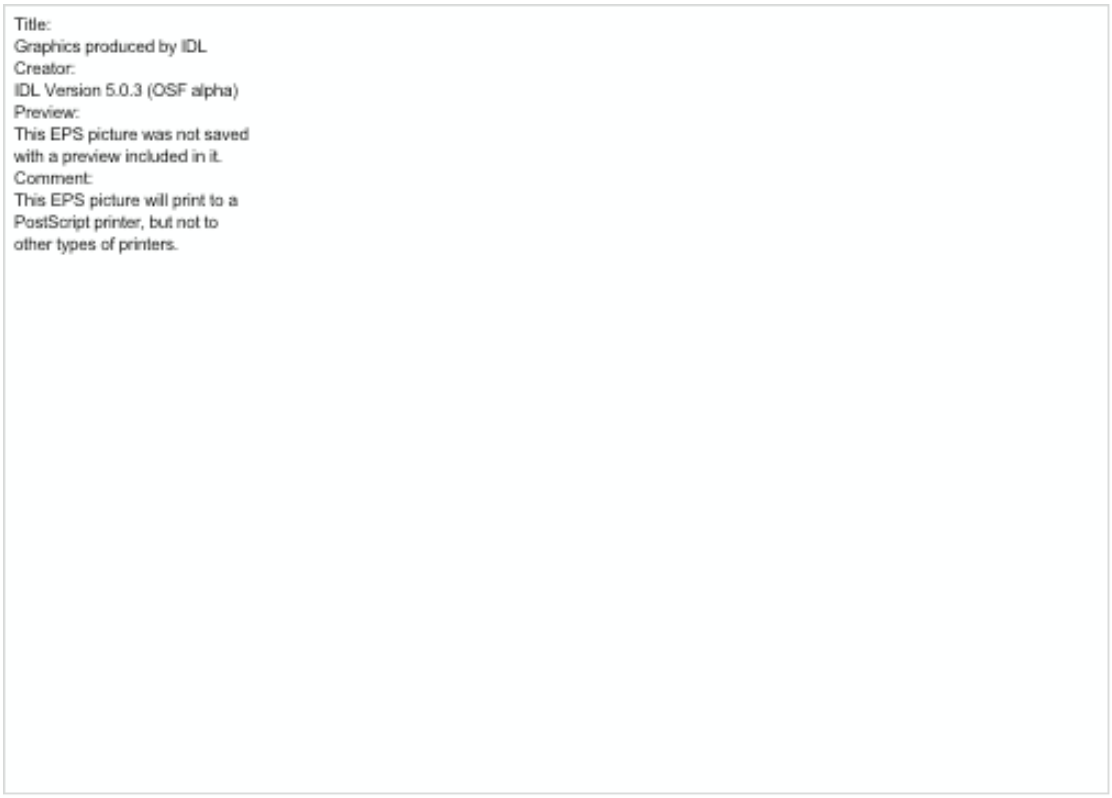
estimating Ga: error sizes estimating Gd: error sizes

model:	HadCM2
anomalies:	JJA temp (°C) for overlapping 30y per., relative to first 240y of con.
patterns:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x): Ga 1, Ga 2, Ga 3, Ga 4; Gd 1, Gd 2, Gd 3, Gd 4 (thin) **Ga 1,2,3,4**; **Gd 1,2,3,4** (thick)
Ga estimate:	for overlapping 30y periods in Ga: each pattern * Ga scaler
Gd estimate:	for overlapping 30y periods in Gd: each pattern * Gd scaler
Ga box error:	Ga estimate – Ga modelled value, for each box
Gd box error:	Gd estimate – Gd modelled value, for each box
error sizes:	average magnitude of Ga box errors (**left**), Gd box errors (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.32

Errors in estimating Ga and Gd precipitation from using single simulations.

estimating Ga: error sizes estimating Gd: error sizes

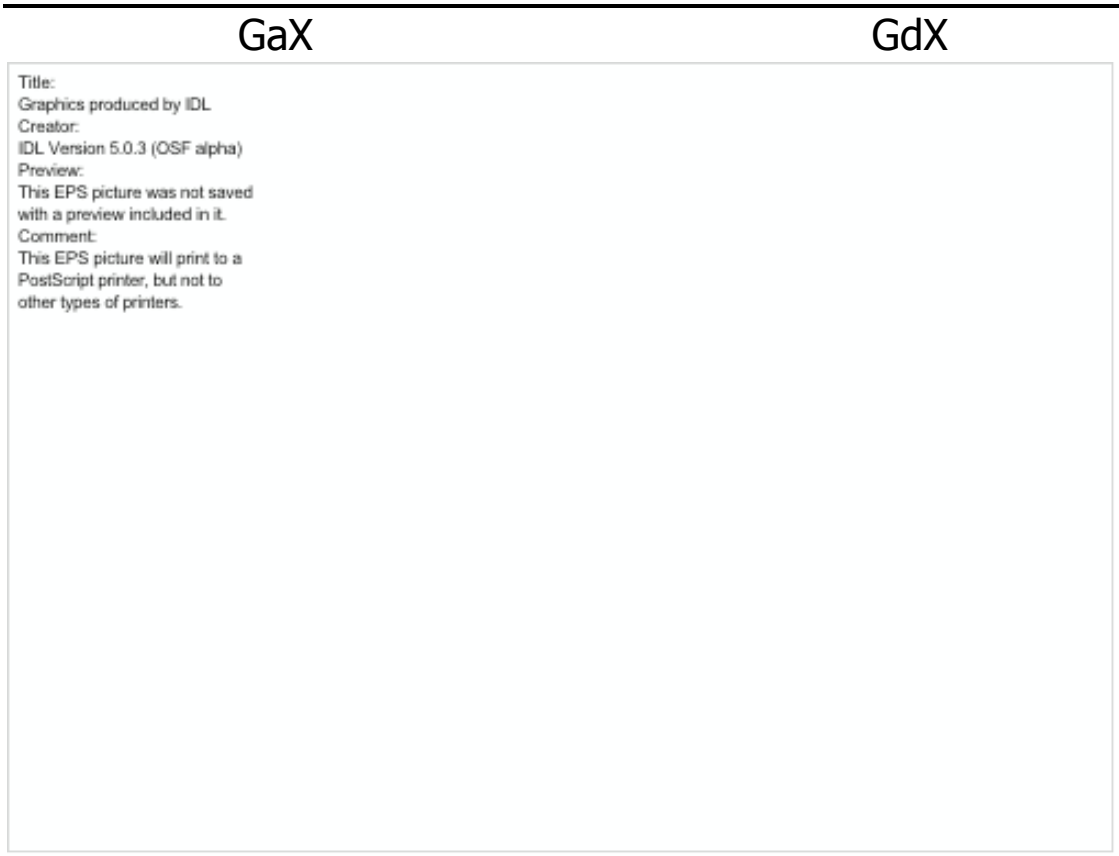


estimating Ga: error sizes estimating Gd: error sizes

model:	HadCM2
anomalies:	JJA prec (mm/day) for overlap. 30y per., relative to first 240y of con.
patterns:	*a* from best-fit line ($y=ax$) from anomalies (y) and scaler (x):
	Ga 1, Ga 2, Ga 3, Ga 4; Gd 1, Gd 2, Gd 3, Gd 4 (thin)
	Ga 1,2,3,4; **Gd 1,2,3,4** (thick)
Ga estimate:	for overlapping 30y periods in Ga: each pattern * Ga scaler
Gd estimate:	for overlapping 30y periods in Gd: each pattern * Gd scaler
Ga box error:	Ga estimate – Ga modelled value, for each box
Gd box error:	Gd estimate – Gd modelled value, for each box
error sizes:	average magnitude of Ga box errors (**left**), Gd box errors (**right**)
x-axis:	period mid-year (**top**), period scaler (**bottom**)

Figure 4.A2.1

Standard deviations of the temperature response patterns for Ga and Gd.

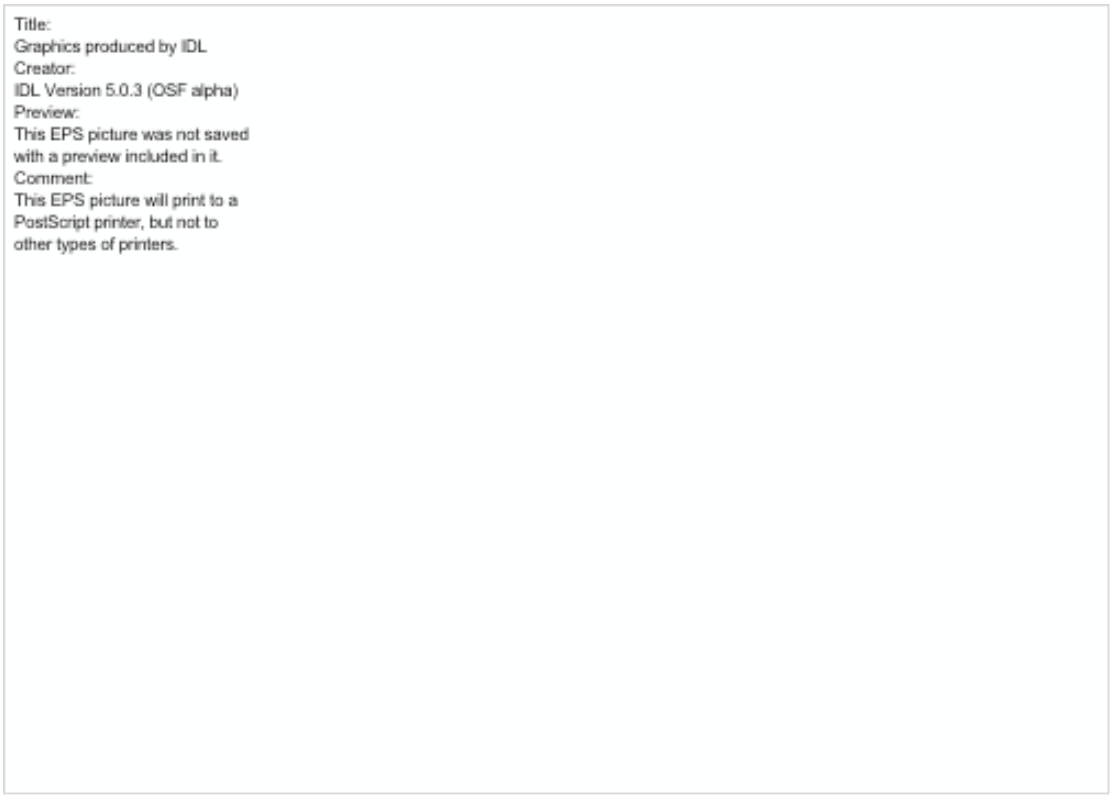


model:	HadCM2
simulations:	Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	temperature (°C) relative to first 240y of control
season:	JJA
temporal:	means of overlapping 30-year periods
patterns:	for each member, using all periods, in °C per °C
stdev:	standard dev. of Ga patterns (**top left**, global-mean: 0.0365)
	standard dev. of Gd patterns (**top right**, global-mean: 0.0472)
difference:	top right minus top left (**bottom left**),
	top right as a fraction of top left (**bottom right**), where:
	<0.5: 535 boxes (7.8%)
	>2.0: 1474 boxes (21.6%)
	both: 2009 boxes (29.5%)

Figure 4.A2.2

Standard deviations of the temperature response patterns for Ga and Gd.

GaX (2045) GdX (2073)



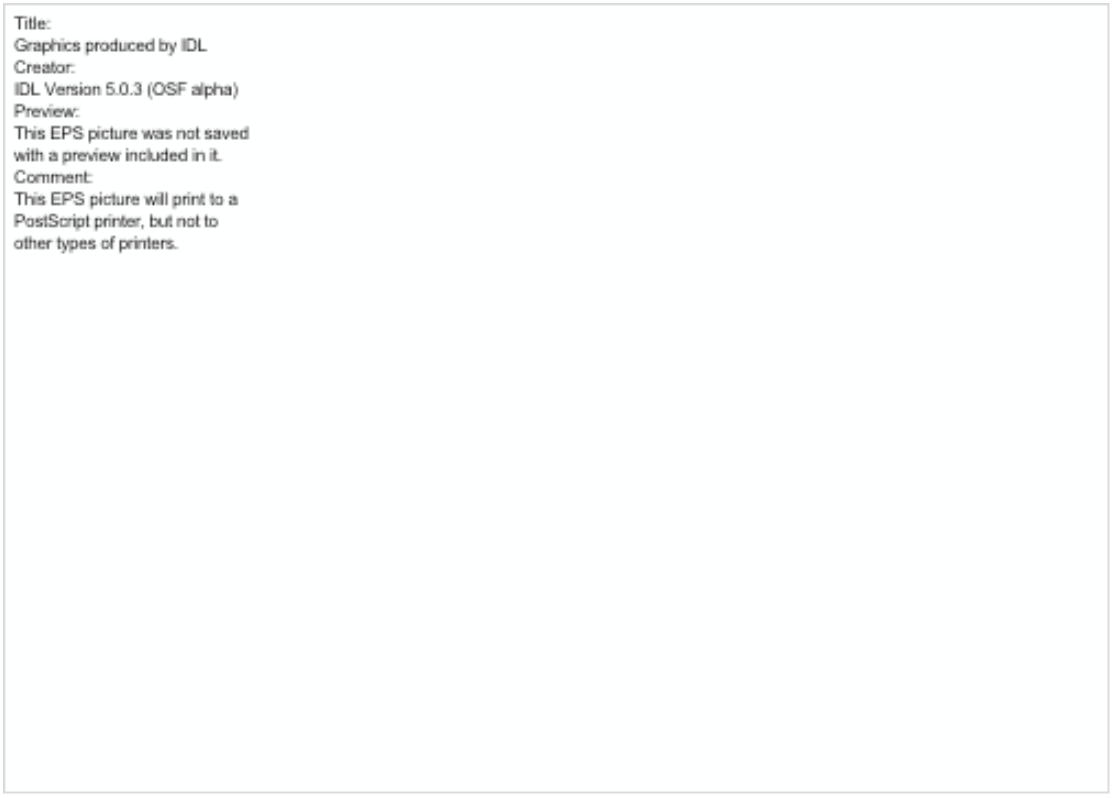
difference (GdX—GaX) fraction (GdX/GaX)

model:	HadCM2
simulations:	Ga 1,2,3,4; Gd 1,2,3,4
anomalies:	temperature (°C) relative to first 240y of control
season:	JJA
temporal:	means of 30-year periods centred on 2045 (Ga) and 2073 (Gd)
patterns:	each mem., anomaly divided by global annual mean, (°C per °C)
stdev:	standard dev. of Ga patterns (**top left**, global-mean: 0.0643)
	standard dev. of Gd patterns (**top right**, global-mean: 0.0652)
difference:	top right minus top left (**bottom left**),
	top right as a fraction of top left (**bottom right**), where:
	<0.5: 1242 boxes (18.2%)
	>2.0: 966 boxes (14.2%)
	both: 2208 boxes (32.4%)

Figure 4.A2.3

Effect of assumption ($\sigma_A^2 = \sigma_B^2$) on D and significances of pattern differences.

D stdev. diff. (ass.—no ass.) *D* stdev. fract. (ass.—no ass.)



sig. diff. (ass.—no ass.) sig. fract. (ass.—no ass.)

assumption:	Ga and Gd response pattern variances are equal ($\sigma_A^2 = \sigma_B^2$)
D :	Gd response pattern – Ga response pattern
σ_A , σ_B :	as in Figure 4.A2.1 (top)
D variance:	$\sigma_A^2/2$ (without ass.) and $\sigma_A^2/4 + \sigma_B^2/4$ (with ass.)
D stdev:	stdev. of D (with ass.) minus stdev. of D (without ass.) (**top left**)
	and as a fraction of stdev. of D (without ass.) (**top right**)
D sig:	sig. of D in stdev. from the mean (D divided by D stdev.)
	sig. of D (with ass.) minus sig. of D (without ass.) (**bottom left**)
	and as a fraction of sig. of D (without ass.) (**bottom right**)
errors in sig:	1806 boxes (26.5%) in error by at least 0.5
	963 boxes (14.1%) in error by at least 1.0
	787 boxes (11.5%) the sig. at least double the correct sig.

5. May we scale the distribution?

5.1 Introduction

In chapter 4 we investigated the possibility of using pattern scaling to estimate changes in grid-box means on 30-year time scales. Our interest in this chapter lies in the possibility of applying pattern scaling not merely to the mean, but to the entire probability distribution. Therefore we are interested in higher statistical moments, such as the variance, skewness, and kurtosis.

We might think of the means we estimated in chapter 4 from a single 30-year period in two equally valid ways:

i. as an estimate of a *multi-decadal* mean from a sample of one;
ii. as an estimate of an *inter-annual* mean from a sample of thirty.

From the viewpoint of chapter 4 this choice does not matter. However, it becomes relevant when we consider not merely the mean of the probability distribution, but also higher order statistics such as the variance:

i. for a *multi-decadal* variance we would calculate the means of corresponding 30-year periods from an ensemble (size n), and calculate the variance between them with a sample size of n;
ii. for an *inter-annual* variance we would calculate the variance among the individual seasonal means in a 30-year period,[178] giving a sample size of 30 for a single simulation, and n^* 30 for an ensemble (size n).[179]

Since our ensembles only have four members, we are likely to find it difficult to obtain stable estimates of changes in *multi-decadal* probability distributions. Therefore in this chapter we discuss the estimation of changes in *inter-annual* probability distributions. We examine seasonal (JJA) changes in temperature (°C) and precipitation (mm/day) at the grid-box scale, using the same HadCM2 simulations as in chapter 4.

We begin by calculating globally-averaged time series for the variance, skewness, and kurtosis (section 5.2). The magnitude of the internal variability suggests that

[178] In order to remove any influence from the changes in mean during a 30-year period, it would be necessary to detrend the 30-year period before using it to estimate the inter-annual variability.

[179] In practice we estimate higher-order statistics from an ensemble by averaging between the inter-annual statistics estimated (using the sample size of 30) from each of the ensemble members.

we need ensembles in order to calculate the higher statistical moments with an acceptable level of stability, so we limit our analysis in this chapter to the Ga and Gd ensembles. In section 5.3 we calculate response patterns that approximate the behaviour of the higher order statistics. The failure of the response patterns to explain much of the variability in skewness and kurtosis leads us to limit the analyses that follow to the inter-annual variance. In section 5.4 we estimate the changes in inter-annual variance under increased radiative forcing by scaling the response patterns by annual global-mean temperature (the 'scaler'). In section 5.5 we draw together the results from chapters 4 and 5 by presenting the grid-boxes where the errors introduced through estimating means and variances by pattern scaling are neither statistically or practically significant. If we assume normality then pattern scaling provides an accurate method for representing changes in the entire probability distribution for such grid-boxes. We draw together our conclusions in section 5.6.

5.2 Globally-averaged moments

The motive for our investigation is the possibility of using pattern scaling to estimate changes not merely in the *mean*, but in the *entire distribution*. Therefore we build upon our discussion of the mean in chapter 4 by considering the variance (Equations 5.1), Pearson's measure of skewness (Equations 5.2), and a standard measure of kurtosis (Equations 5.3). Taking each Ga ensemble member in turn, we calculate these statistics for each grid-box and each overlapping 30-year period. We then average the grid-box statistics across the globe to obtain the time-series of individual simulations plotted in Figure 5.1, together with the average of those four time-series. Thus each value plotted for an individual simulation is itself the average of the variance (or skewness or kurtosis) calculated at each grid-box from the 30 seasonal (JJA) means in that period, after detrending that period.[180]

The most striking difference between these time-series and the equivalent time-series plotted for the mean (Figure 4.1) lies in the much larger magnitude of internal variability relative to any response to increased radiative forcing. The magnitude of internal variability suggests that we need an ensemble, not merely a single simulation, in order to provide stable estimates of the higher statistical

[180] The detrending is a vital step. If we do not detrend, changes in the long-term mean will interfere with our estimate of the inter-annual variance. It is not even sufficient to detrend the entire time-series at a grid-box at once, since any non-linearities in the response to radiative forcing will hinder the effectiveness of the detrending. The only effective method is to individually detrend each 30-year period at each grid-box, relative to the scaler (not time).

moments. Therefore we restrict our analysis in this chapter to the Ga and Gd ensembles; we do not consider the stabilisation scenarios.

In Figure 5.2 (blue lines) we replot the ensemble means from Figure 5.1, and add the equivalent statistics from the Gd ensemble (green lines). We repeat the same method for precipitation (Figure 5.3). The level of internal variability, even in the ensemble means, is such that it is difficult to identify any responses to radiative forcing in the temperature statistics with any certainty; however, it appears that there may be slight increases in inter-annual variance and skewness over the globe as a whole (Figure 5.2). The small changes in Figure 5.2 do not imply that there are no changes in the temperature statistics for pattern scaling to estimate, because changes of one sign in one region may be cancelled out in these global averages by changes of the opposite sign in another region.

By way of contrast with temperature, the increases in all the inter-annual statistics for precipitation are clearly identifiable from internal variability (Figure 5.3). This is not surprising; the heteroscedasticity of precipitation is such that changes in higher order statistics accompany any changes in mean precipitation.

The non-Gaussian values of the skewness and kurtosis parameters for precipitation – and the changes in them – are such that we cannot assume that the inter-annual distribution of precipitation is Gaussian. This introduces two complications:

1. Routine statistical analyses (such as for hypothesis testing) that assume that the distribution is normal will produce results that are erroneous to the extent that the distribution is not normal.
2. If the mean and the variance of the modelled distribution are estimated by scaling, and a probability distribution is estimated from those two statistics, the estimated distribution will be erroneous to the extent that the modelled distribution is not normal. However, it is still of interest to estimate higher statistical moments for precipitation, not least because it may prove possible to describe inter-annual precipitation in terms of another probability distribution.

5.3 Response patterns

In chapter 4 (section 4.5) we calculated spatial patterns that described the response of temperature and precipitation means to increased radiative forcing. Here we calculate spatial patterns that describe the responses for inter-annual variance, skewness, and kurtosis.

We begin by describing the inter-annual variance of Ga temperatures. For each grid-box and each Ga simulation we use least squares regression to calculate a best-fit line between the scaler (x) and the inter-annual variance derived from overlapping (detrended) 30-year periods (y). We fit to an equation of the form $y=ax+b$ (where a and b are constants). We also calculate r, the product-moment correlation coefficient. We then calculate the ensemble mean of the four values of a, b, r obtained from the four simulations. Thus we obtain three patterns (Figure 5.4):

- the 'response' pattern consists of the value of a in each grid-box, and represents the rate of change in variance as a function of the scaler (top left);
- the 'initial' pattern consists of the value of b in each grid-box, and represents the variance under pre-industrial radiative forcing (top right);
- the 'variance explained' pattern consists of the value of r^2 in each grid-box (bottom left).

The initial pattern (top right) shows a marked land-sea contrast: the variance is greatest over the northern continents and smallest in the Arctic, but there is also a region of high variance over the Southern Ocean associated with sea-ice. The response pattern (top left) also shows a marked land-sea contrast; the largest rates of increase in variance under increased radiative forcing generally occur in the regions where the variance is already large (top right). In the parts of the Southern Ocean associated with the largest increases in temperatures (Figure 4.10, bottom left) the sea ice retreats, leading to a substantial decrease in variance in these regions (Figure 5.4, top left). In half (50.4%) of the grid-boxes the response to increased radiative forcing explains at least half of the variability in the variance (Figure 5.4, bottom left). Moreover, the regions with the largest rates of change in variance (Figure 5.4, top left) are among the regions where the response to increased radiative forcing explains the largest proportion of the variability in the variance (Figure 5.4, bottom left). Therefore it appears reasonable to attempt to estimate the inter-annual variance under increased radiative forcing by scaling this pattern.

We also calculated an equivalent set of patterns for the Gd ensemble (Figure 5.5). We plot the Gd response pattern (top right) next to the Ga response pattern (top left). The similarity between the different forcing scenarios may be seen by visual inspection. However, although the spatial patterns are similar, they have different

magnitudes.[181] The Gd response is weaker than the Ga response (bottom left), and explains less of the variability.[182] The differences between the forcing scenarios are significant in 20.3% of the grid boxes (bottom right). The slight increase in inter-annual variability over the oceans in Ga is less in Gd, and the substantial increase over the continents in Ga is much less in Gd. The Southern Ocean sea-ice retreats much less in Gd than in Ga, so the inter-annual variance also decreases much less in Gd than in Ga. Moreover, the spatial coherence of the differences suggests that the differences are not merely artefacts of internal variability. Therefore it seems likely that the magnitude of the response of inter-annual temperature variability depends to some extent on the *rate* of increase in radiative forcing.

We apply the same method to precipitation. We calculate an equivalent set of patterns for Ga precipitation (Figure 5.6) and compare the Ga and Gd response patterns (Figure 5.7). The initial pattern (Figure 5.6, top right) shows that the regions where the inter-annual variance is relatively large are generally regions where precipitation amounts are also relatively great.[183] This is not merely the case in the tropics, but also in extratropical mountainous regions, such as the Andes, Alps, and Scandinavia. The effects of ENSO on the inter-annual variance in precipitation do not appear to be particularly large, judging from the regions with the largest variance, despite the strength of the model ENSO (Tett *et al.*, 1997). The reason is probably that the effects of ENSO are not dominant in this particular season (JJA).[184] As with temperature, in half (51.8%) of the grid-boxes the response to increased radiative forcing explains at least half of the variability in the variance, and it is generally where the responses are largest (Figure 5.6, top left) that they explain the largest proportion of the variability in the variance (bottom left).

The Ga and Gd patterns are broadly similar (Figure 5.7)[185] and the differences between patterns are significant in only 9.4% of the grid-boxes. We introduced

---

[181] The average magnitudes are 0.156 (top left) and 0.147 (top right) $°C^2$ per °C.

[182] In Gd the response to increased radiative forcing explains at least half of the variability in the variance in only 30.2% of the grid-boxes.

[183] The climatological (JJA) precipitation from the HadCM2 control is displayed in Johns *et al.* (1997) Figure 19b.

[184] The seasons when, and regions where, precipitation is most strongly related to ENSO are documented in Ropelewski and Halpert (1987, 1989). Their findings are summarised in their Figures 21 (1987) and 18 (1989).

[185] The average magnitudes are 0.185 (top left) and 0.174 (top right) $(mm/day)^2$ per °C.

heteroscedasticity[186] in section 4.6, and we see it here in the close relationship between the precipitation mean and variance.[187]

Once again we have concentrated on the differences between patterns rather than their similarities. Nonetheless the patterns have strong similarities, particularly spatially. Therefore we expect the estimates to be reasonably accurate if we scale a response pattern from a different forcing scenario.

Finally, we apply the same methods to skewness and kurtosis. We summarise the results in Figure 5.8 by plotting the proportion of the variability explained by the skewness (left) and kurtosis (right) patterns for temperature (top) and precipitation (bottom). The very low proportions obtained suggest that in most grid-boxes there is little or no recognisable response to increased radiative forcing that might prove amenable to pattern scaling. The precipitation plots (Figure 5.8, bottom) are slightly more promising than the temperature plots. However, even so most of the response pattern for the skewness (Figure 5.9, top left) is noise, and noise strongly influences most of the grid-boxes in the kurtosis response pattern (Figure 5.9, bottom left). Therefore we do not attempt to apply pattern scaling to the skewness and kurtosis, but limit our analysis to the inter-annual variance.

5.4 Estimating the variance

In this section we scale the response patterns calculated in section 5.3 to estimate changes in inter-annual variance, first of temperature and then of precipitation. The 'changes' are from the 'pre-industrial' variance, which we define as the initial patterns in Figures 5.4 (temperature) and 5.6 (precipitation).

5.4a Estimating temperatures

We present our estimates of the changes in the inter-annual variance of temperature under Ga in Figures 5.10 (the 2020s) and 5.11 (the 2080s).

- The modelled anomaly (top left) is the mean inter-annual variance among the four ensemble members for that period, minus the pre-industrial variance.
- The estimated anomaly (top right) is the response pattern from Ga multiplied by the Ga scaler, minus the pre-industrial variance.

[186] A variable is heteroscedastic when its standard deviation varies in proportion to its mean.

[187] Compare Figure 5.6 (top left) with Figure 4.10 (top left). Also consider the difference plots: compare Figure 5.7 (bottom left) with Figure 4.15 (bottom left). Furthermore, consider the relative weakness (compared to Ga) of the mean Gd precipitation response to radiative forcing (Figure 4.15, top right), and the relative weakness of the variance response (Figure 5.7, bottom left).

- The error (bottom left) is the estimated anomaly minus the modelled anomaly.
- The error significance (bottom right) follows Equations 4.5.[188]

A visual inspection of the modelled (top left) and estimated (top right) inter-annual variances in Figures 5.10 and 5.11 suggests that the scaling method is broadly accurate in estimating temperatures throughout the 21st century. However, pattern scaling does introduce some errors (bottom left), and a visual inspection suggests that many of the errors change sign during the 21st century; as when estimating the means in chapter 4, we suggest that slight non-linearities in the response to increased radiative forcing are responsible. These errors are larger as a proportion of the estimated values than was the case in chapter 4.[189] However, the internal variability in the inter-annual variance is very large, so the errors are only statistically significant in a small proportion (about 8%) of the grid-boxes.

We extend our analysis to the Gd ensemble, estimating inter-annual variances of temperatures as follows:

- for Ga using patterns from Gd (Figure 5.12),
- for Gd using patterns from Ga (Figure 5.13),
- for Gd using patterns from Gd (Figure 5.14).

In these figures we plot the error (top) and its significance (bottom) for the 2020s (left) and 2080s (right).

When we again estimate changes using a pattern drawn from the same scenario (Gd, Figure 5.14) the errors from pattern scaling again arise from non-linearities in the response, and are statistically significant in a similar proportions of grid-boxes to the Ga case (Figures 5.10, 5.11). However, when we estimate changes in one forcing scenario using the patterns in another (Figures 5.12, 5.13) there are increases both in the errors themselves, and in the proportion of grid-boxes where these errors are statistically significant. The reason is – as we noted in section 5.3 – that the response to increased radiative forcing in Gd is weaker than in Ga, particularly over the continents (Figure 5.5). Therefore when we extrapolate,[190] the estimated variance is generally too small (Figure 5.12). Although the errors are larger over land (top), the errors are most significant over the oceans (bottom),

[188] In Equations 4.5 the random variables representing statistics from a single simulation (M_i and E_i) represent multi-decadal means, but here they represent inter-annual variances. We estimate $\square_M$ from the same period in the ensemble.

[189] We demonstrate this in Figures 5.24-5.27, which are described below.

[190] We estimate the Ga inter-annual variance by scaling the Gd response pattern.

presumably because the background level of inter-annual variance is generally smaller over the oceans (Figure 5.4, top right). When we interpolate,[191] the errors are smaller because the smaller Gd scaler does not magnify the difference between the Ga and Gd response patterns to the same extent (Figure 5.13). Thus in the 2080s the error from interpolation is statistically significant in only 11.5% of the grid-boxes, whereas the proportion from extrapolation is 18.5%.

We present time-series of the errors in estimating Ga and Gd in Figures 5.15 and 5.16 respectively.[192] We estimate the average error introduced by pattern scaling in the absence of increased radiative forcing by inspecting the first century displayed in these figures (top left); it amounts to ±0.08 °C^2. When we estimate the variance under increased radiative forcing by scaling the response pattern from the same scenario,[193] the average error increases slightly to ±0.10 °C^2, reflecting either slight non-linearities in the global response or enhanced internal variability. The average magnitude of the errors remains approximately constant at 0.12-0.17° C^2 (centre), and the average significance amounts to less than a standard deviation (right).

When we estimate the variance by scaling the response pattern from a different scenario there are larger increases in the error magnitudes and significances.[194] When we extrapolate we obtain smaller estimates of the Ga variance (Figure 5.15, left, green lines) because of the weaker response in Gd compared to Ga. Similarly, when we interpolate we obtain larger estimates of the Gd variance (Figure 5.16, left, blue lines). The errors from extrapolating are much larger than from interpolating: the average magnitude of the error only increases to 0.2 °C^2 when interpolating for the 2080s, whereas when we extrapolate it doubles to 0.3 °C^2.

5.4b Estimating precipitation

We present equivalent results for precipitation in Figures 5.17-5.23. The results are broadly similar to those we obtained for temperature. We summarise the results from the spatial plots (Figures 5.17-5.21) as follows:

- Pattern scaling gives broadly accurate estimates of changes in inter-annual variance.[195]

---

[191] We estimate the Gd inter-annual variance by scaling the Ga response pattern.

[192] The equivalent figures for estimating multi-decadal means (4.25 and 4.26) were discussed in section 4.7.

[193] See the blue lines in Figure 5.15 and green lines in Figure 5.16.

[194] See the green lines in Figure 5.15 and blue lines in Figure 5.16.

[195] See Figures 5.17 and 5.18 (top).

- The errors when estimating changes by scaling a response pattern from the same forcing scenario are statistically significant in a few more grid-boxes than was the case for temperature.[196]
- There is a change in the sign of the error in many regions during the 21st century that is similar to that obtained for temperature.[197]
- When we scale a response pattern from one forcing scenario to estimate changes under another forcing scenario the errors introduced by pattern scaling increase, and are significant in 10-20% of the grid-boxes.[198]
- In the 2080s, the errors introduced by pattern scaling are smaller when interpolating than when extrapolating.[199]

We summarise the results from the time series (Figures 5.22, 5.23) as follows:

- The error in the absence of radiative forcing amounts to ±0.06 $(mm/day)^2$ when averaged over the globe, with an average magnitude of 0.08-0.12 $(mm/day)^2$.
- When we estimate changes by scaling a response pattern from the same forcing scenario the average error increases in Ga (not Gd) to ±0.09 $(mm/day)^2$. The large non-linearities in the response to increased radiative forcing in individual regions are such that the error magnitudes increase, doubling in Ga by the end of the 21st century. However, these errors remain small in relation to internal variability.
- When we estimate changes by scaling a response pattern from a different forcing scenario the errors increase further, more so when extrapolating (green lines in Figure 5.22) than when interpolating (blue lines in Figure 5.23).

5.4b Estimating changes in the inter-annual variance

The evidence presented in this section suggests that it is feasible to scale patterns of change in the inter-annual variance from one forcing scenario to another for a large proportion of HadCM2 grid-boxes. The error is never statistically significant in more than a fifth of the grid-boxes. Moreover, if one interpolates rather than extrapolates (*e.g.* estimating for Gd using Ga patterns rather than *vice versa*), this proportion improves considerably. Nonetheless unless ensemble sizes are increased, the inter-annual variance cannot be estimated with the same precision as the multi-decadal mean because of the large internal variability.

---

[196] See Figures 5.17 and 5.18 (bottom right), and 5.21 (bottom).

[197] Compare bottom left in Figures 5.17 and 5.18.

[198] See Figures 5.19 and 5.20 (bottom).

[199] Compare bottom right in Figures 5.19 and 5.20.

5.5 Estimating changes in the distribution

We may also infer from the evidence presented in the previous section that it may be possible to estimate changes in inter-annual probability distributions, particularly of temperature, at the grid-box scale. If we may estimate both the mean and the inter-annual variance with reasonable accuracy,[200] then under the assumption of normality we may estimate the distribution itself. The values that we obtained for skewness and kurtosis parameters in section 5.3 suggest that this assumption may hold good for temperature, but perhaps not for precipitation. Nonetheless, it may prove possible to identify a theoretical probability distribution that gives a good fit for inter-annual precipitation.[201] In this case the accuracy with which we may estimate the precipitation variance suggests that we might obtain similar accuracy for the parameters of this other probability distribution.

In this section we summarise the extent to which it is possible to simultaneously estimate changes in both mean and variance, by drawing together some of the results from chapters 4 and 5. As we noted in chapter 4, although it is possible to give a strict definition of the *statistical* significance of any error, the *practical* significance is context-dependent. Therefore we adopt an arbitrary threshold for practical significance: an error that amounts to 20% of the estimated value is said to be practically significant.

In chapter 4 we presented the estimates for the 2080s of changes in Gd mean temperature (Figure 4.22) and precipitation (Figure 4.23) that we obtained by scaling the Ga response patterns (*i.e.* the interpolation case). In particular we presented plots (bottom right) showing the ratio of error to modelled value for the grid-boxes where the error was statistically significant. These plots enable us to estimate the proportion of grid-boxes with errors from pattern scaling that are both statistically and practically significant. We repeat this method not just for the 2080s, but also for the 2020s, and not just for the mean, but also for the inter-annual variance. We present our results in Figures 5.24-5.27:

- temperature (5.24 and 5.25) and precipitation (5.26 and 5.27),
- 2020s (5.24 and 5.26) and 2080s (5.25 and 5.27),
- mean (top left) and inter-annual variance (top right).

[200] The former condition was addressed in chapter 4, the latter in this chapter.

[201] One candidate is the gamma distribution.

We go one step further in Figures 5.24-5.27 by combining these results. For each variable, time period, and grid box we compare the error (as a proportion of the estimated value) in estimating the mean with that from estimating the inter-annual variance. We plot the larger magnitude (bottom left) and then mask out all grid-boxes where the larger magnitude does not exceed 0.2 (bottom right). Thus we identify for each variable and time period the grid-boxes where we are able to estimate changes in both the mean and variance without any errors that are both statistically and practically significant; these are the grey boxes in the bottom right plots. The grid-boxes where there are statistically and practically significant errors are in the minority, up to a fifth for temperature and up to a third for precipitation.

In most cases the estimates of the mean (top left) and the variance (top right) contribute approximately equal numbers of grid-boxes with statistically significant errors. However, a much larger proportion of the statistically significant errors in estimating the variance are also practically significant than is the case when estimating the mean. The estimates of temperature and precipitation also have approximately equal numbers of grid-boxes with statistically significant errors. Similarly, the difference between temperature and precipitation lies in the larger proportion of the statistically significant errors in precipitation that are also practically significant.

We conclude that to a large extent it is possible to simultaneously estimate changes in both mean and variance at the grid-box scale. In the majority of grid-boxes pattern scaling does not introduce errors that are both statistically and practically significant, whether in estimating the mean or the inter-annual variance. The proportions where there are such errors are limited to a fifth (temperature) or a third (precipitation). On this basis we suggest that changes in inter-annual probability distributions, particularly of temperature, may be estimated at the grid-box scale.

5.6 Conclusions

We recognise that much of the value in climate scenarios may lie in changes in distributional parameters other than the mean. In this chapter we have examined whether or not we may scale patterns of higher-order statistics in a similar manner to the method for means examined in chapter 4, with the aim in mind of applying pattern scaling to changes in probability distributions. To our knowledge this has not been attempted before. Our analyses have been conducted on seasonal (JJA) inter-annual probability distributions of temperature (°C) and precipitation

(mm/day). Our initial investigations suggested that a single simulation was insufficient to extract stable estimates of higher-order statistics, so we restricted our analysis to the perturbed ensembles. Our investigation of the skewness and kurtosis parameters suggested that the responses to increased radiative forcing explained a very small proportion of the variability, so we concentrated our attention on changes in inter-annual variance.

Our results suggest that it is indeed practical to scale a response pattern of inter-annual variance from one scenario to estimate changes in inter-annual variance under another scenario. Although the errors are much larger (in proportion to the modelled value) than was the case when estimating means, they are never significant in more than a fifth of the grid-boxes. In general, the precipitation estimates are at least as accurate as the temperature estimates. For both precipitation and temperature the response to increased radiative forcing is weaker in Gd than in Ga, so errors are introduced when estimating one scenario using patterns from the other. These errors are smaller when interpolating (estimating Gd using the Ga patterns) than when extrapolating (estimating Ga using the Gd patterns).

When we combined our estimates of the mean and inter-annual variance for the interpolation case we found that the errors introduced by pattern scaling were both statistically and practically significant in only a minority of grid-boxes: up to a fifth (temperature) or up to a third (precipitation). On this basis we suggest that changes in inter-annual probability distributions, particularly of temperature, may be estimated at the grid-box scale. To do this for precipitation requires further consideration to be given to the distributional form that is most appropriate to inter-annual precipitation. However, the accuracy of estimating changes in the variance from pattern scaling suggests that other distributional parameters might also be accurately estimated.

If it is possible to estimate changes in a probability distribution using pattern scaling, then we may exploit this ability when constructing climate scenarios:

(a) We may present estimated changes in probability distributions for forcing scenarios not simulated by GCMs.

(b) We may apply Monte Carlo techniques to the generation of climate 'projections' at regional scales. We may combine probabilities in different elements of the 'cascade' of uncertainty[202] using Bayesian statistics, and exploit

[202] See section 2.3 for our discussion of the 'cascade' of uncertainty.

the highly complex integrals by random sampling.[203] However, rather than terminating our evaluation of the cascade at global-mean temperature, as we might otherwise be forced to do by the lack of large samples of regional climate changes from GCMs, through pattern scaling we may extend our evaluation of the cascade to regional climate changes.[204]

Some further directions in which our work might be extended are common to the work described in chapter 4:

- we might examine seasons other than JJA;
- we might examine variables other than precipitation and temperature;
- we might examine time scales other than the inter-annual;
- if ensembles were available for stabilisation scenarios we might examine the accuracy of estimates made for them.

As in chapter 4, we point out that we have restricted our analysis to scenarios of greenhouse gases alone, and that further complexity is introduced by the introduction of spatially-varying radiative forcings such as that from sulphate aerosols.[205] We also recognise that pattern scaling possesses some of the limitations of the GCMs it represents:

i. the credibility of the results from pattern scaling depends on the credibility of the GCM, particularly at regional scales;

ii. if the GCM does not represent the full range of conceivable changes – including abrupt changes – then neither will the results from pattern scaling.

However, the credibility of the models on regional scales – limitation (i) – depends partly on the definition of 'regions' that we adopt. Therefore in chapter 6 we examine the extent to which the accuracy of scaling is dependent on the spatial scale and on the aggregation of grid-boxes into regions.[206] Limitation (ii) may be partly addressed by including a number of GCMs in any assessment of future climate.[207] Therefore in chapter 7 we examine whether we may apply pattern scaling to models other than HadCM2.

---

[203] See section 2.4 for our discussion of the roles of Bayesian statistics and Monte Carlo techniques.

[204] Huntingford and Cox (2000) have already taken a step in this direction by constructing an analogue model to HadCM3 that relates radiative forcing to land surface variables.

[205] See section 2.8 for our discussion of the importance of sulphate aerosols to pattern scaling.

[206] See section 2.10 for our discussion of the spatial choices that influence our results.

[207] See section 2.11 for our discussion of the value of using multiple GCMs.

Equations 5.1

1. The interannual variance (σ^2) of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for a single simulation (m):

$$\sigma_{mij}^2 = \frac{\sum_{y=1}^{n_y} a_{mijy}^2}{n_y} - \left(\frac{\sum_{y=1}^{n_y} a_{mijy}}{n_y} \right)^2$$

2. The interannual variance (σ^2) of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for an ensemble (M, size n_m):

$$\sigma_{Mij}^2 = \frac{\sum_{m=1}^{n_m} \sum_{y=1}^{n_y} a_{mijy}^2}{n_m n_y} - \left(\frac{\sum_{m=1}^{n_m} \sum_{y=1}^{n_y} a_{mijy}}{n_m n_y} \right)^2$$

Equations 5.2

Following Francis (1979) section 2.7.2

1. Pearson's measure of skewness is given as follows:

   skewness = (mean – mode) / standard deviation

   For moderate skewness the following relation approximately holds:

   mode = mean – 3 (mean – median)

   Therefore since we are calculating for a set of values from a continuous distribution, to make the skewness more stable we estimate it in terms of the median rather than the mode:

   skewness = 3 (mean – median) / standard deviation

   This is a dimensionless statistic, evaluating to:

   - zero for a Gaussian distribution;
   - negative for a left-skewed distribution;
   - positive for a right-skewed distribution.

2. The interannual skewness of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for a single simulation (m):

   Equation 5.2.1;

   the mean and standard deviation are as expressed in Equation 5.1.1;

   the median is the mid-value when the values (n_j) are ranked.

3. The interannual skewness of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for an ensemble (M, size n_m):

   mean of the skewness values from the individual members of the ensemble, as calculated in Equation 5.2.2

Equations 5.3

Following Francis (1979) section 2.7.2

1. The standard measure of kurtosis is given by:

   kurtosis = $M_4 / (M_2)^2$

   For a Gaussian distribution this expression evaluates to three, so we calculate our measure of kurtosis as follows:

   kurtosis = $(M_4 / (M_2)^2) - 3$

   This is a dimensionless statistic, evaluating to:

   - zero for a Gaussian distribution;
   - negative values for a platykurtic (relatively less peaked) distribution;
   - positive values for a leptokurtic (relatively peaked) distribution.

2. The $\square$th statistical moment about zero ($M'_{\square}$) of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for a single simulation (m):

$$M'_{\square} = \frac{1}{n_y} \sum_{y=1}^{n_y} a_{mijy}^{\square}$$

3. The 2nd and 4th statistical moments about the *mean* (M_2, M_4) are given by:

   $M_2 = M'_2 - (M'_1)^2$

   $M_4 = M'_4 - 4\,M'_1 M'_3 + 6(M'_1)^2\,M'_2 - 3(M'_1)^4$

4. The interannual kurtosis of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for a single simulation (m):

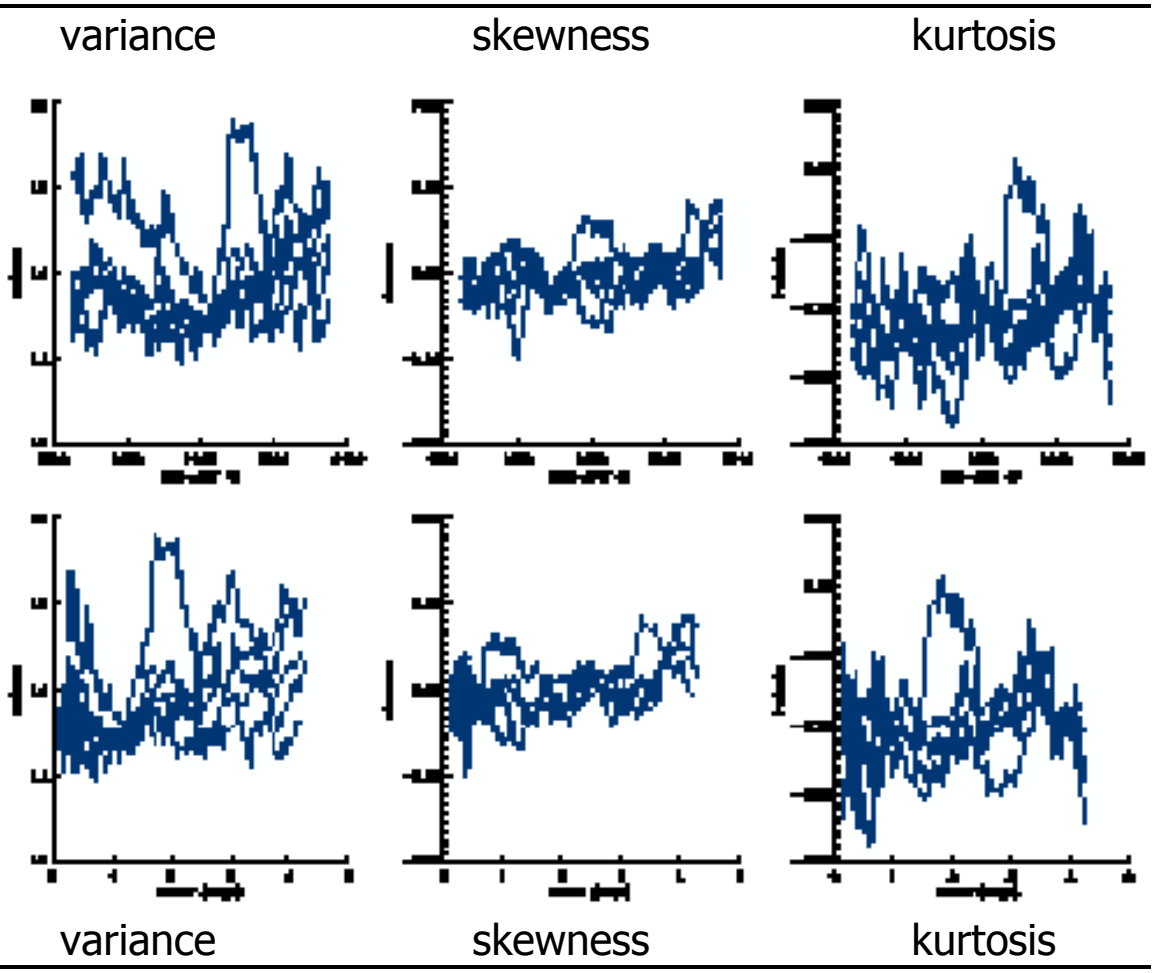
   Equations 5.3.1- 5.3.3

5. The interannual kurtosis of the seasonal anomalies (a) for a particular time slice (j, length in years n_y) and grid-box (i) for an ensemble (M, size n_m):

   mean of the kurtosis values from the individual members of the ensemble, as calculated in Equation 5.3.4

Figure 5.1

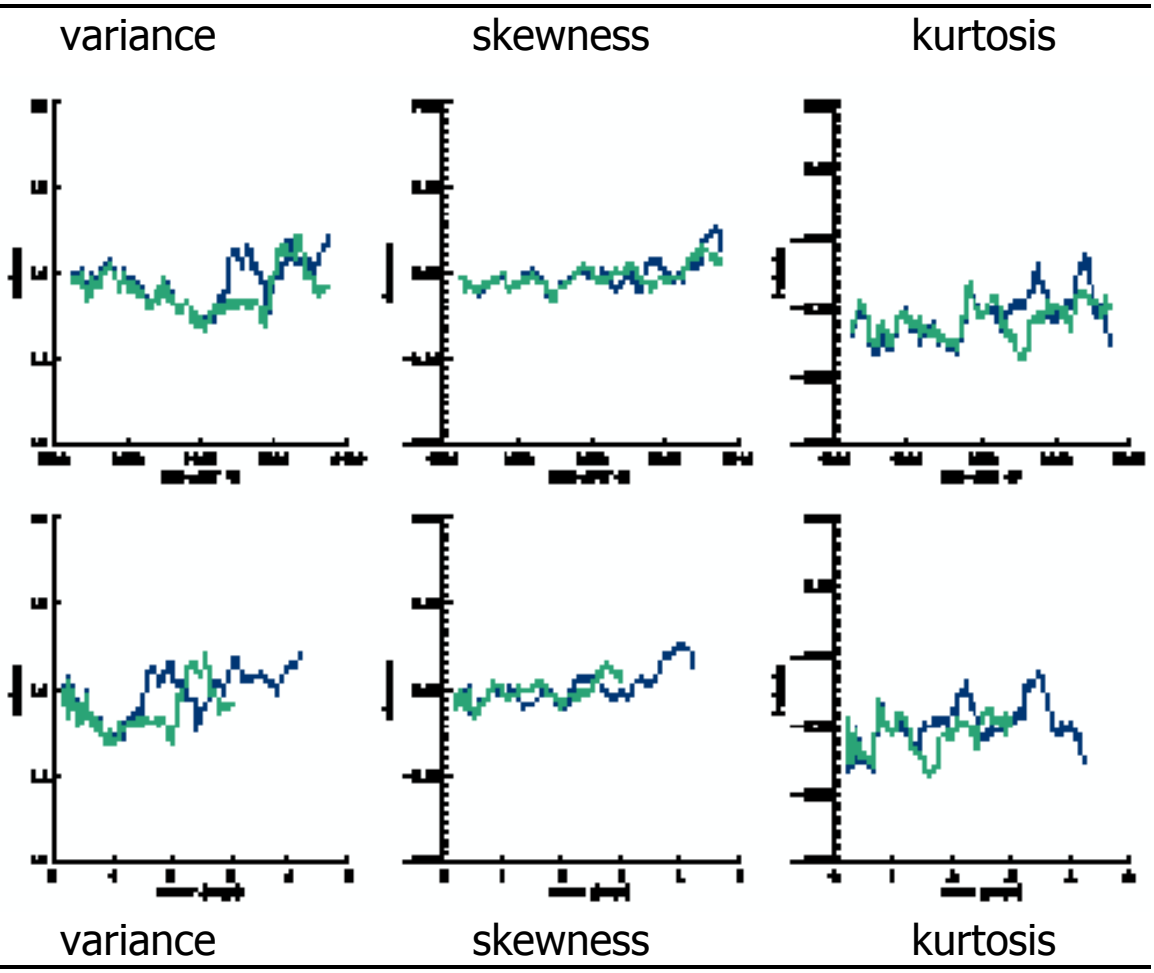
Statistics for 30-year slices of seasonal temperatures from Ga.



simulations:	HadCM2 Ga 1, Ga 2, Ga 3, Ga 4, **mean of Ga statistics**
diagnostic:	grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
statistics:	interannual variance (Equations 5.1, **left**) from 30 values, interannual skewness (Equations 5.2, **centre**) from 30 values, interannual kurtosis (Equations 5.3, **right**) from 30 values
x-axis:	period mid-year (**top**), period global annual temp (°C: **bottom**)

Figure 5.2

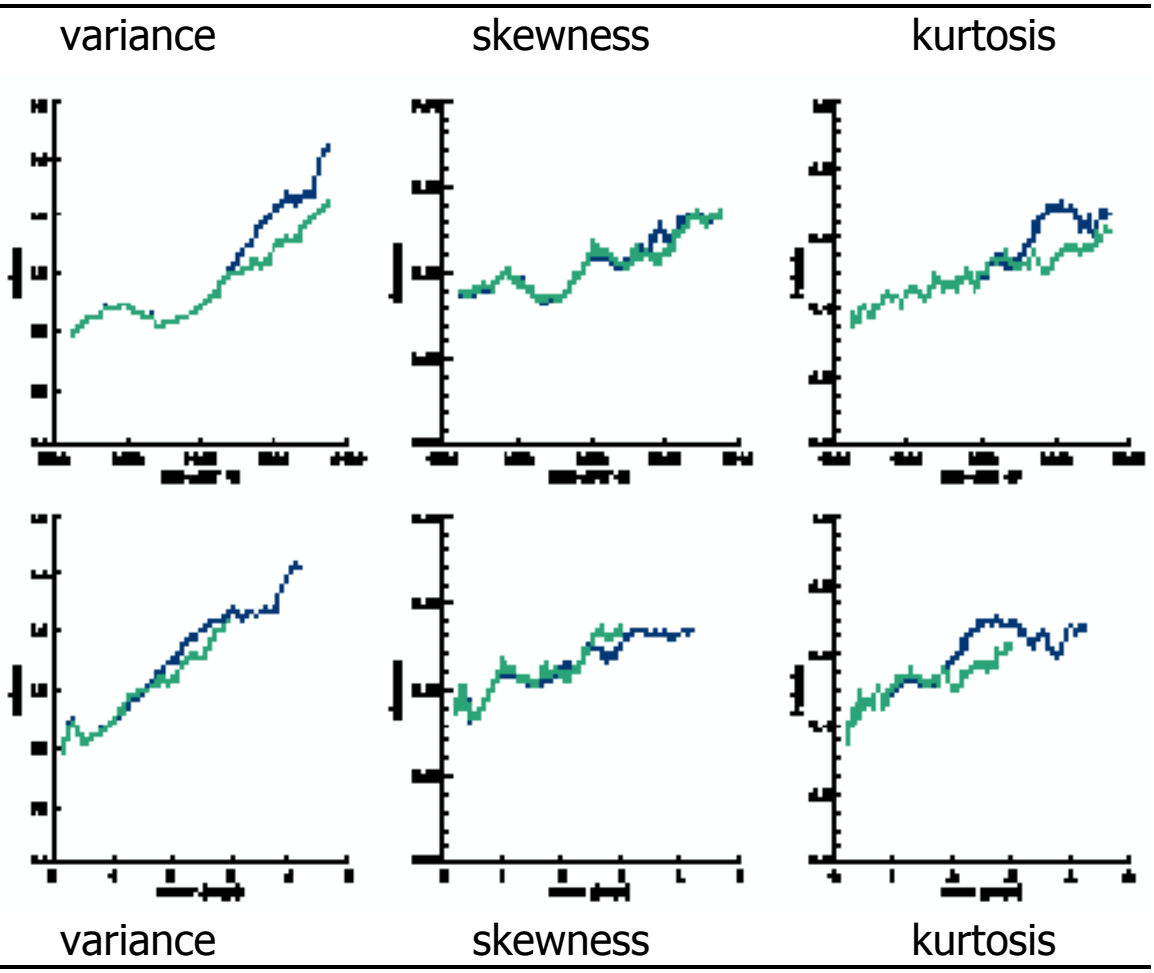
Statistics for 30-year slices of seasonal temperatures from Ga and Gd.



simulations:	HadCM2 **mean of Ga statistics**, **mean of Gd statistics**
diagnostic:	grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
statistics:	interannual variance (Equations 5.1, **left**) from 30 values, interannual skewness (Equations 5.2, **centre**) from 30 values, interannual kurtosis (Equations 5.3, **right**) from 30 values
x-axis:	period mid-year (**top**), period global annual temp (°C: **bottom**)

Figure 5.3

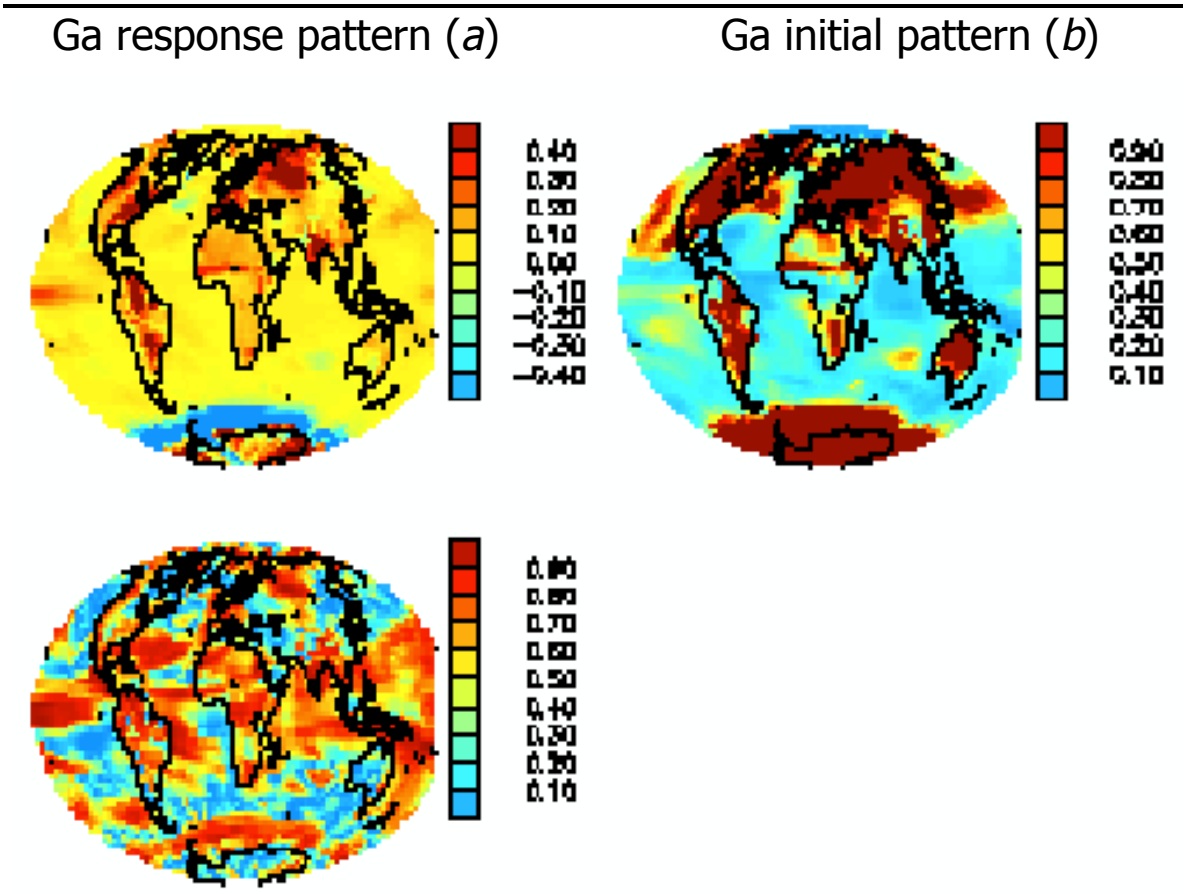
Statistics for 30-year slices of seasonal precipitation from Ga and Gd.



simulations:	HadCM2 **mean of Ga statistics**, **mean of Gd statistics**
diagnostic:	grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
statistics:	interannual variance (Equations 5.1, **left**) from 30 values, interannual skewness (Equations 5.2, **centre**) from 30 values, interannual kurtosis (Equations 5.3, **right**) from 30 values
x-axis:	period mid-year (**top**), period global annual temp (°C: **bottom**)

Figure 5.4

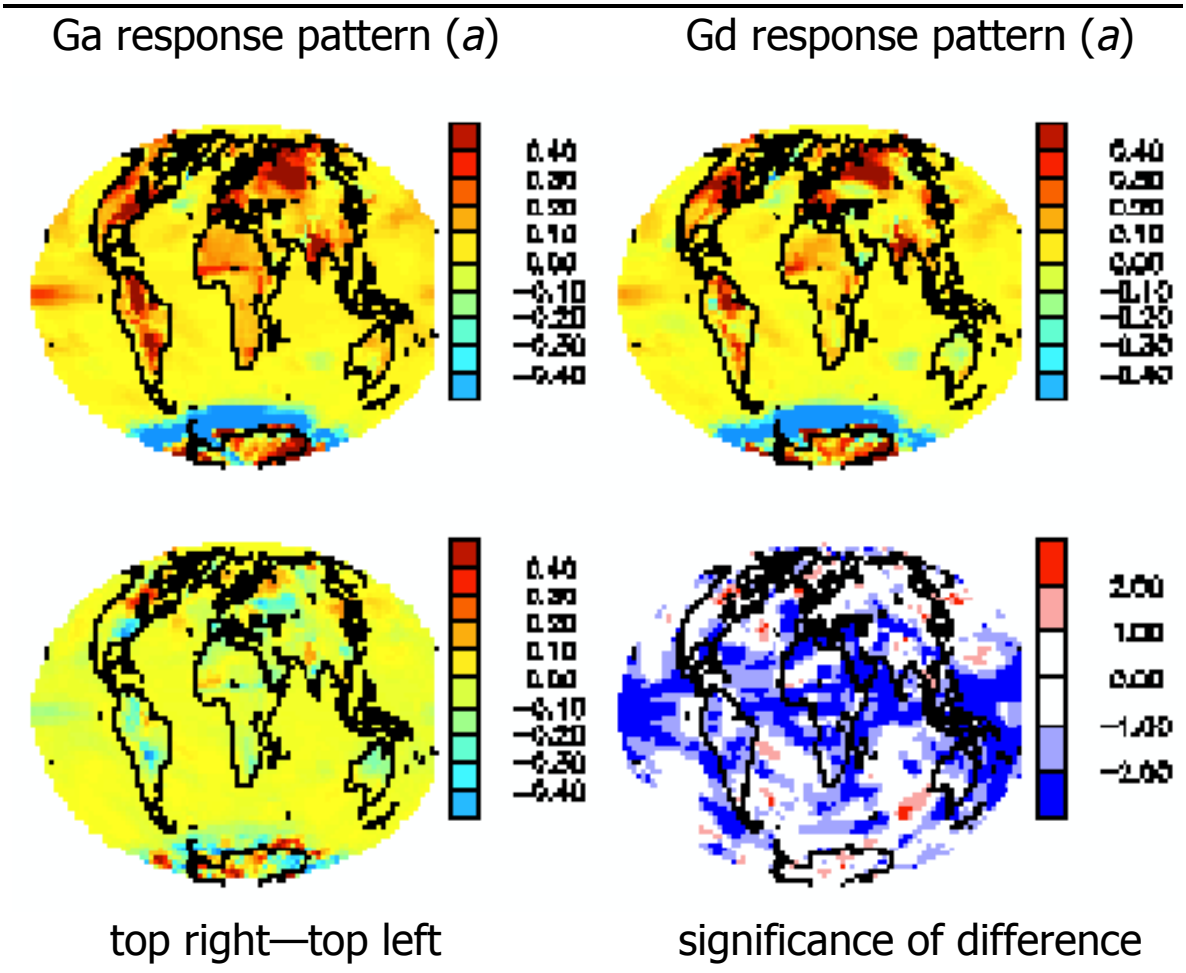
The patterns of temperature interannual variance from the Ga ensemble.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
patterns:	a (**top left**), b (**top right**), r^2 (**bottom left**)
$r^2 > 0.5$:	proportion of grid-boxes is 50.4% (Ga, **bottom left**), 30.2% (Gd)

Figure 5.5

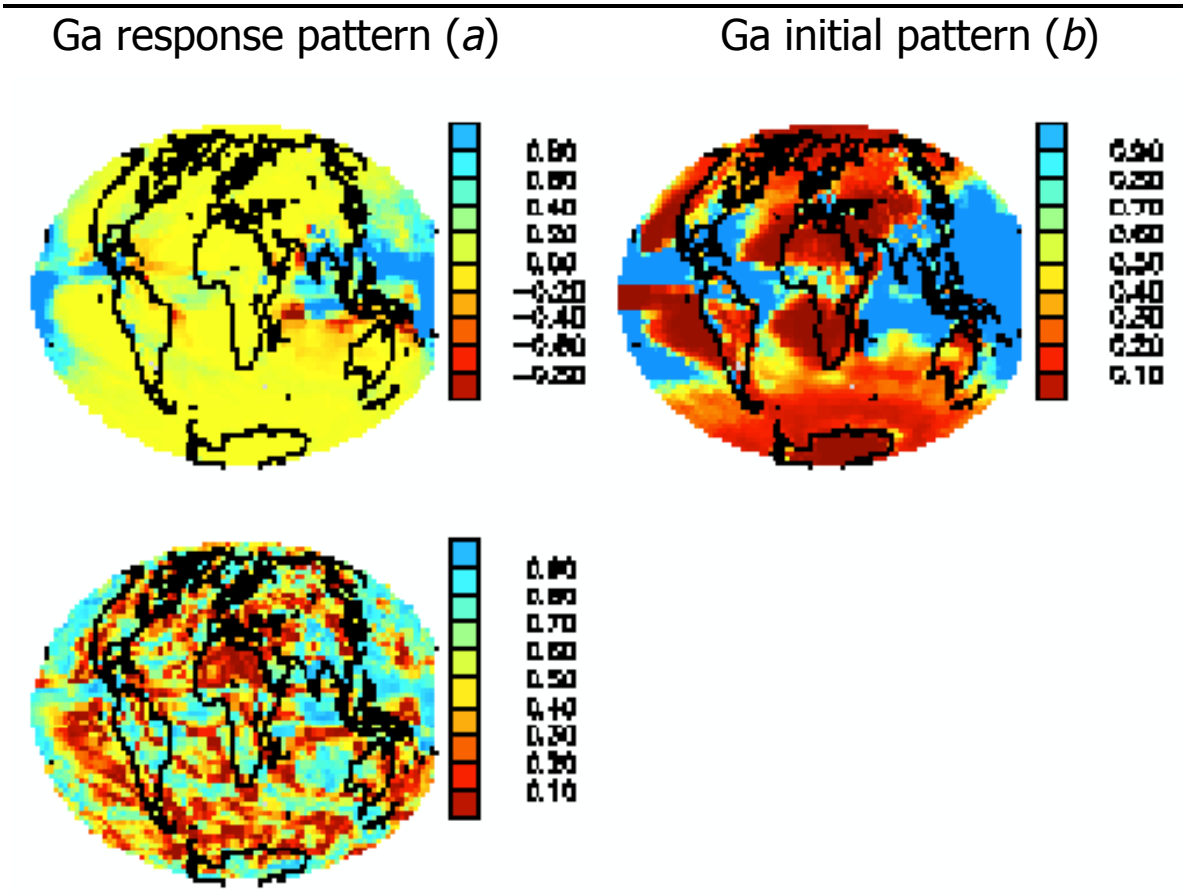
The response of temperature interannual variance in Gd compared with Ga.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
responses:	a for the Ga ensemble (**top left**) and Gd ensemble (**top right**)
	average magnitudes are 0.156°C (Ga) and 0.147°C (Gd)
difference:	a from Gd 1+2+3+4 minus a from Ga 1+2+3+4 (**bottom left**)
diff sig:	difference expressed in standard deviations (**bott. right**, Eq. 4.4)

Figure 5.6

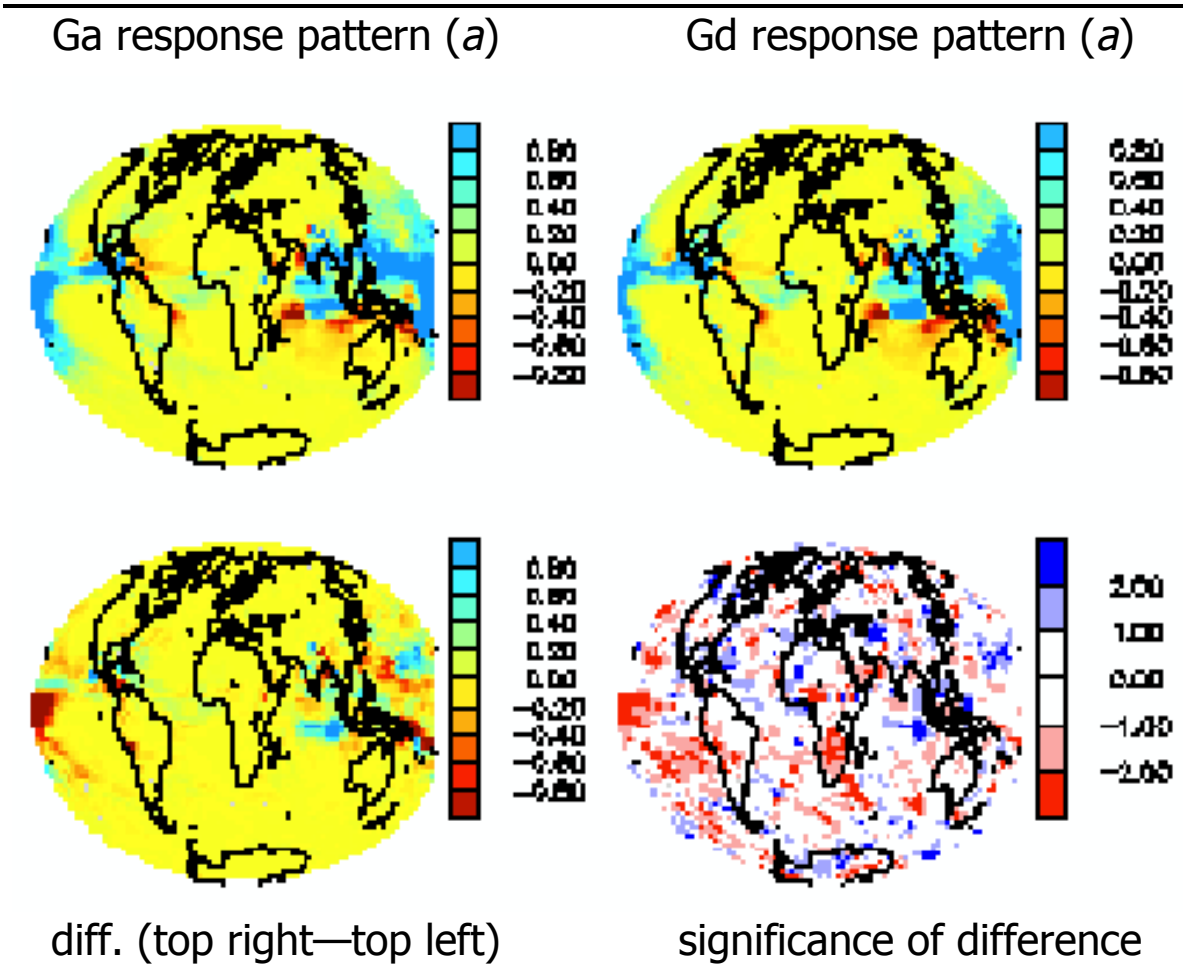
The response of precipitation interannual variance in Ga.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
patterns:	a (**top left**), b (**top right**), r^2 (**bottom left**)
$r^2 > 0.5$:	proportion of grid-boxes is 51.8% (Ga, **bottom left**), 40.8% (Gd)

Figure 5.7

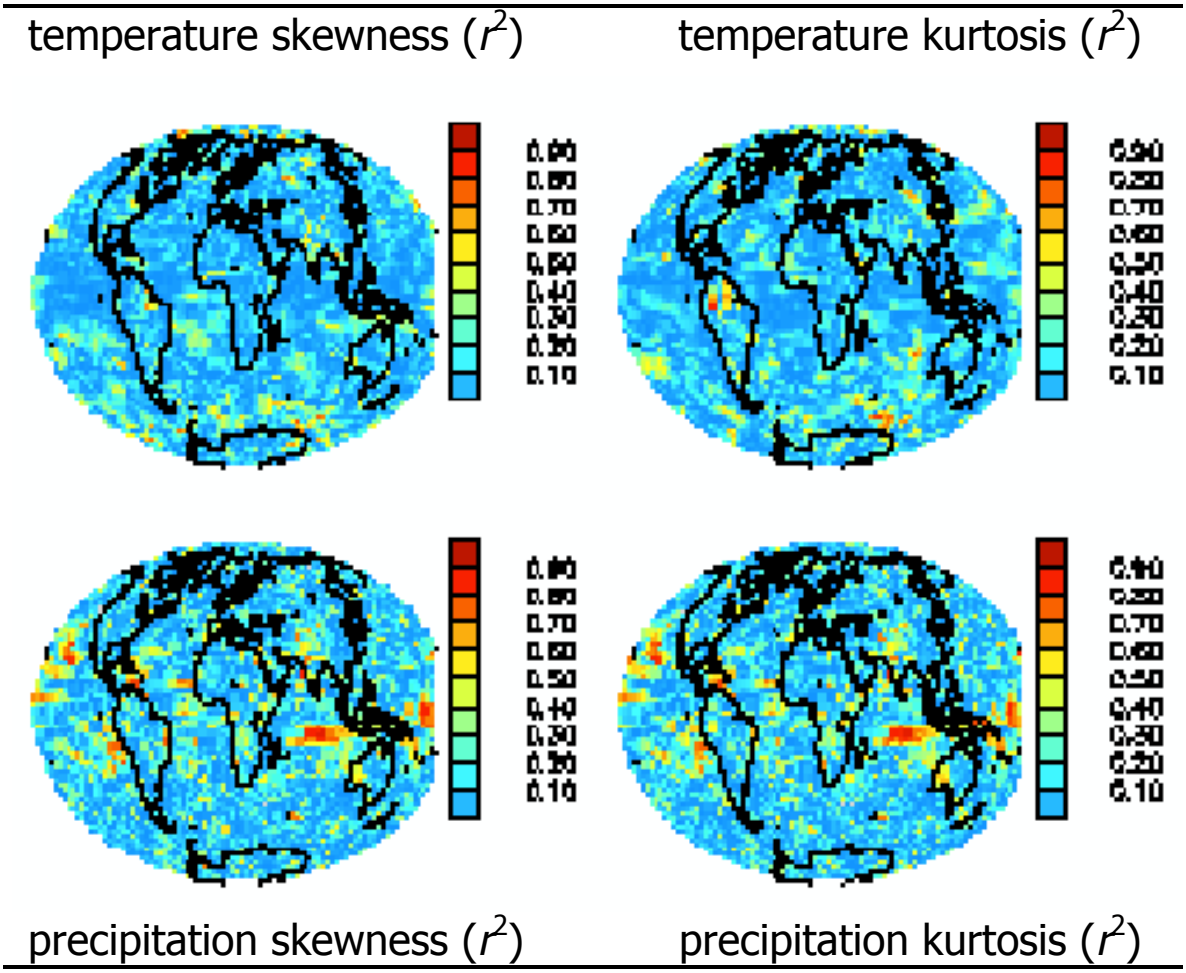
The response of precipitation interannual variance in Gd compared with Ga.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
responses:	a for the Ga ensemble (**top left**) and Gd ensemble (**top right**)
difference:	a from Gd 1+2+3+4 minus a from Ga 1+2+3+4 (**bottom left**)
diff sig:	difference expressed in standard deviations (**bott. right**, Eq. 4.4)

Figure 5.8

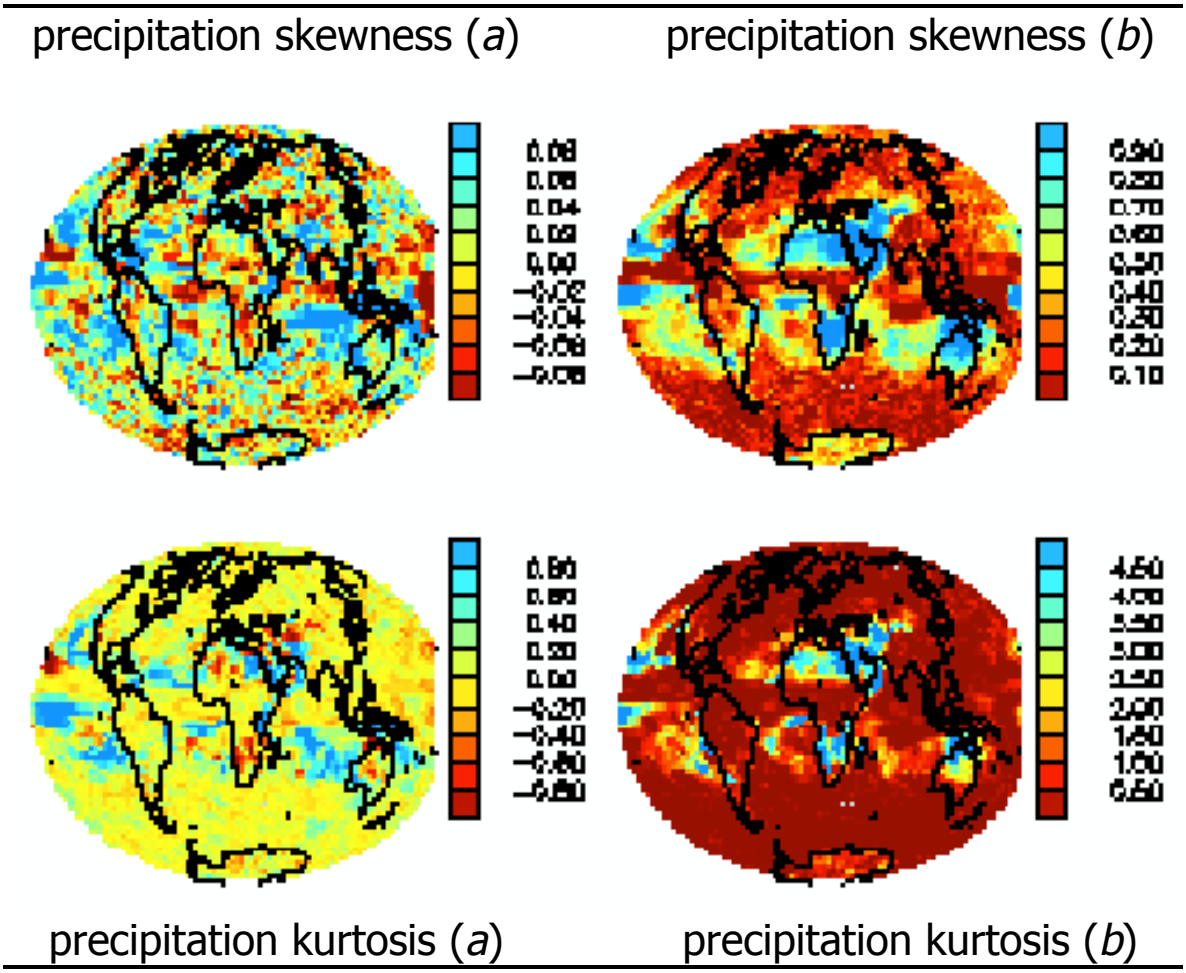
The variance explained by the temperature and precipitation response patterns.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	grid-box JJA temperature (**top**, °C) and prec (mm/day, **bottom**)
original:	overlapping 30y periods, each detrended
statistics:	interannual skewness (Equations 5.2)
	interannual kurtosis (Equations 5.3)
patterns:	$y=ax+b$ where y is original statistic and x is the scaler
plots:	r^2 of skewness (**left**) and kurtosis (**right**)

Figure 5.9

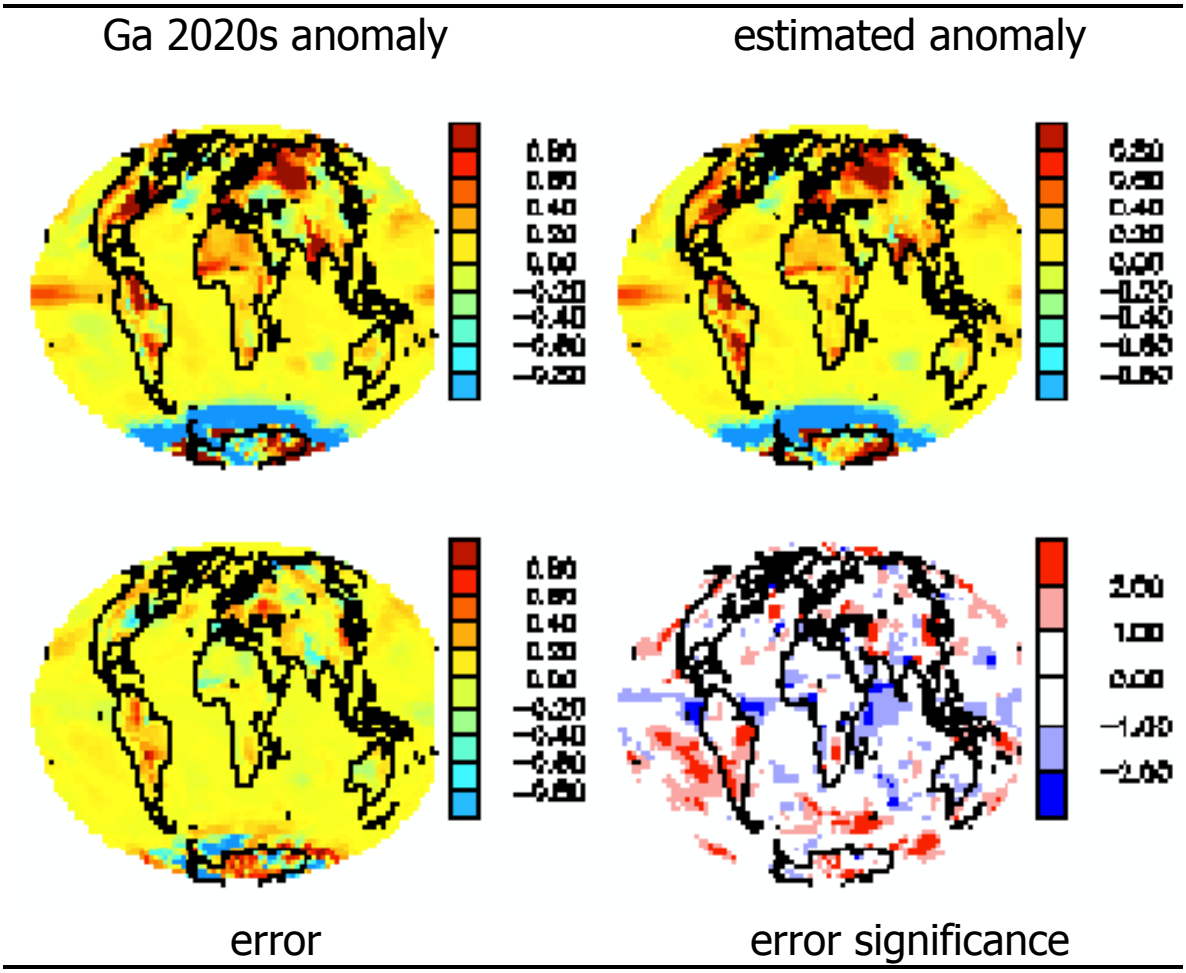
The precipitation response patterns from Ga.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
statistics:	interannual skewness (Equations 5.2) (**top**)
	interannual kurtosis (Equations 5.3) (**bottom**)
patterns:	$y=ax+b$ where y is original statistic and x is the scaler
plots:	a (**left**) and b (**right**)

Figure 5.10

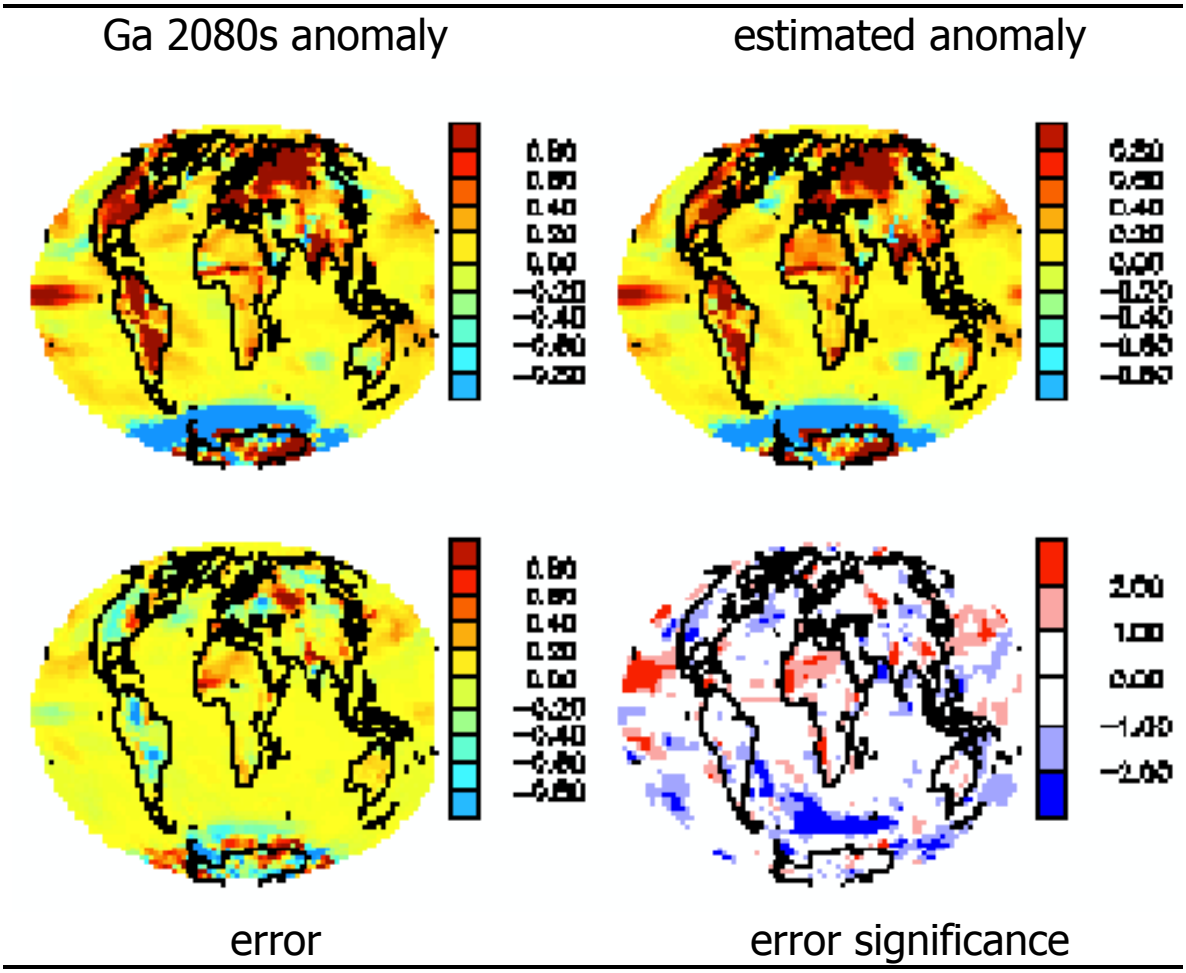
The estimated temperature interannual variance in Ga in the 2020s.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
modelled:	2020s variance – pre-industrial (**top left**)
estimated:	b + (a * Ga scaler for 2020s) – pre-industrial (**top right**)
error:	estimated – modelled (**bottom left**)
error sig:	error in standard deviations (**bott. right**, Eq. 4.5)
	114 boxes (1.7%) <-2□ ; 412 boxes (6.0%) >2□

Figure 5.11

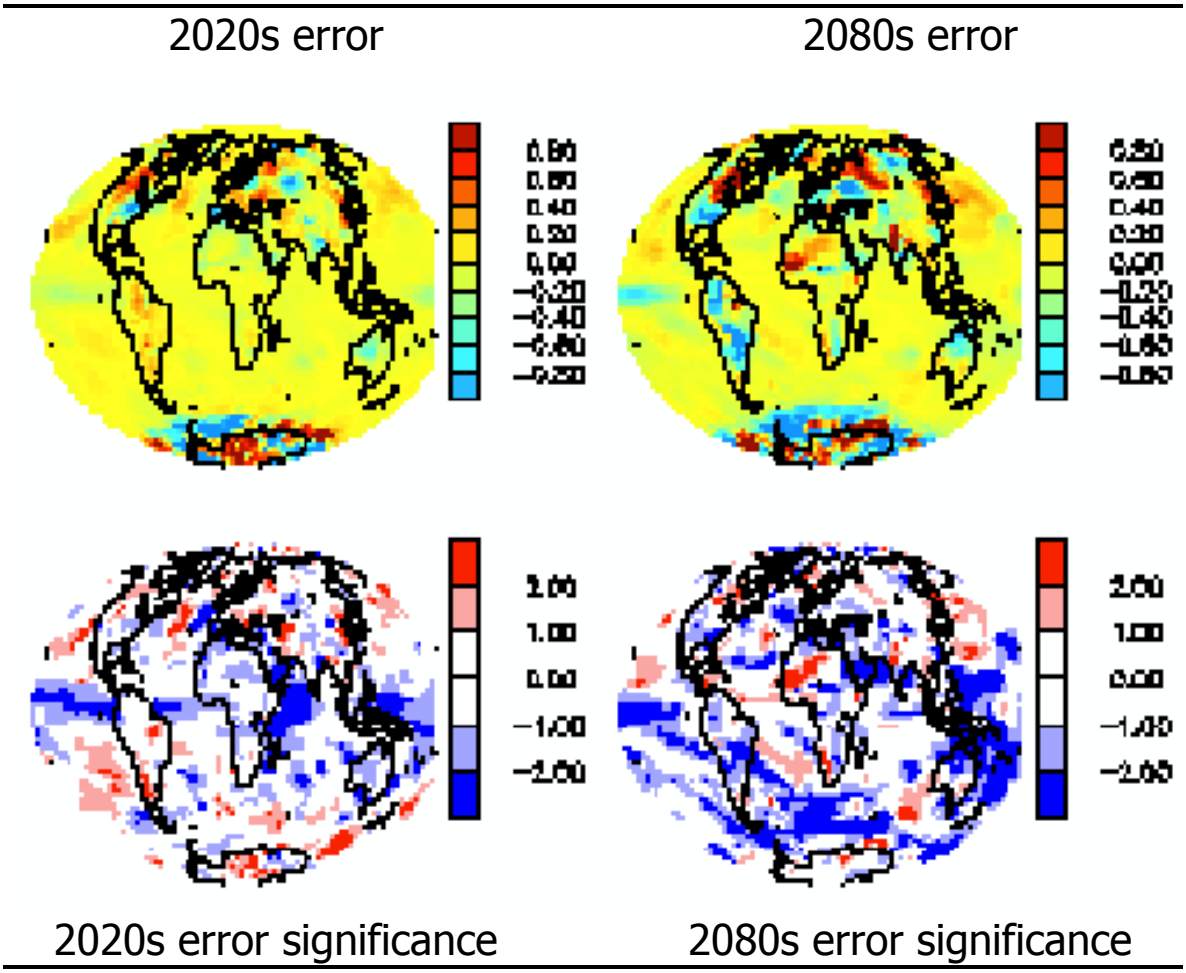
The estimated temperature interannual variance in Ga in the 2080s.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
modelled:	2080s variance – pre-industrial (**top left**)
estimated:	b + (a * Ga scaler for 2080s) – pre-industrial (**top right**)
error:	estimated – modelled (**bottom left**)
error sig:	error in standard deviations (**bott. right**, Eq. 4.5)
	325 boxes (4.8%) <-2□ ; 198 boxes (2.9%) >2□

Figure 5.12

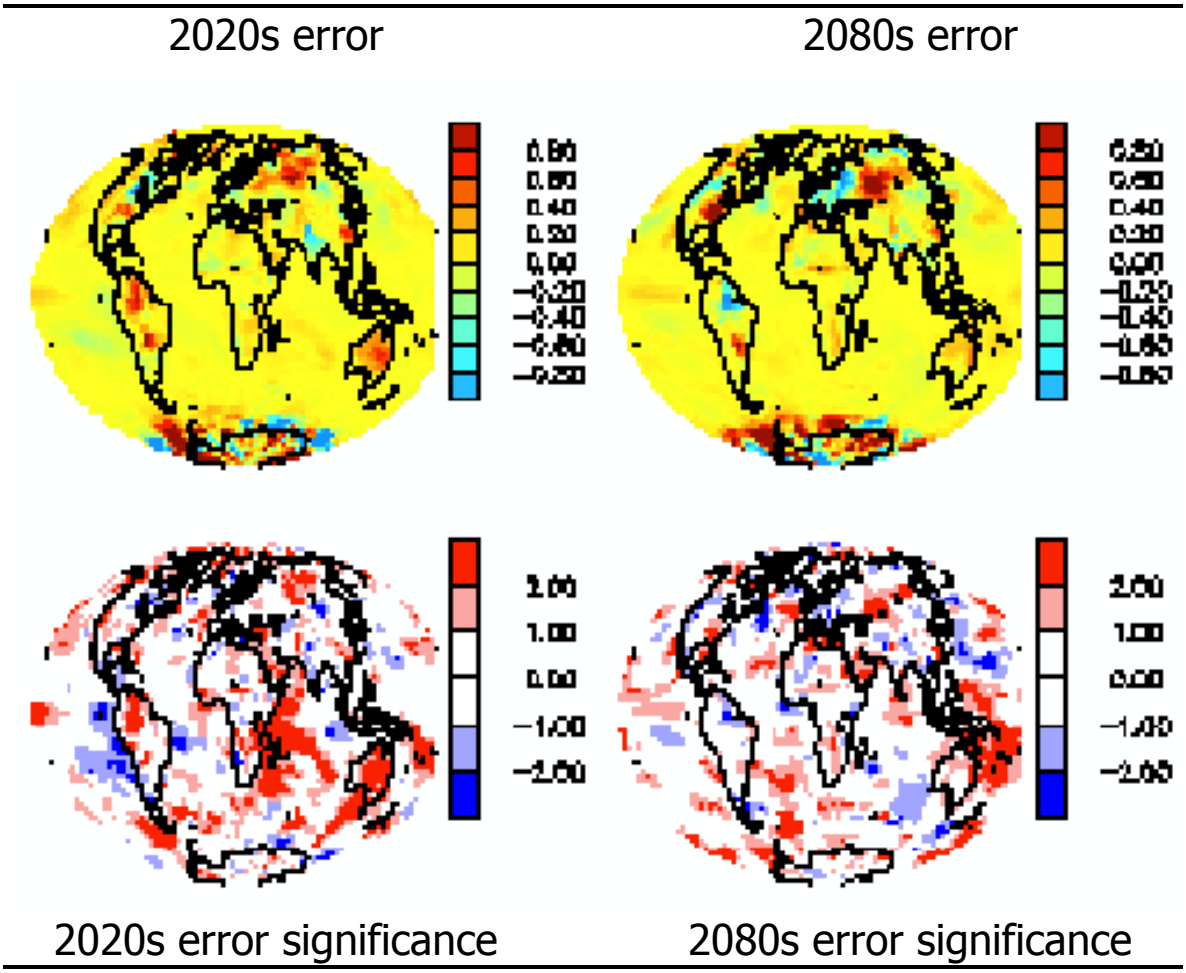
The estimated temperature interannual variance in Ga by scaling Gd pattern.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler, for Gd
modelled:	2020s [or 2080s] variance – pre-industrial
estimated:	b + (a * Ga scaler for 2020s [or 2080s]) – pre-industrial
error:	estimated – modelled for 2020s (**top left**) and 2080s (**top right**)
error sig:	error in standard deviations (**bottom**, Eq. 4.5)
	2020s (**bott. left**): 427 boxes (6.3%) <-2□ ; 282 boxes (4.1%) >2□
	2080s (**bott. right**): 1073 boxes (15.7%) <-2□ ; 189 boxes (2.8%) >2□

Figure 5.13

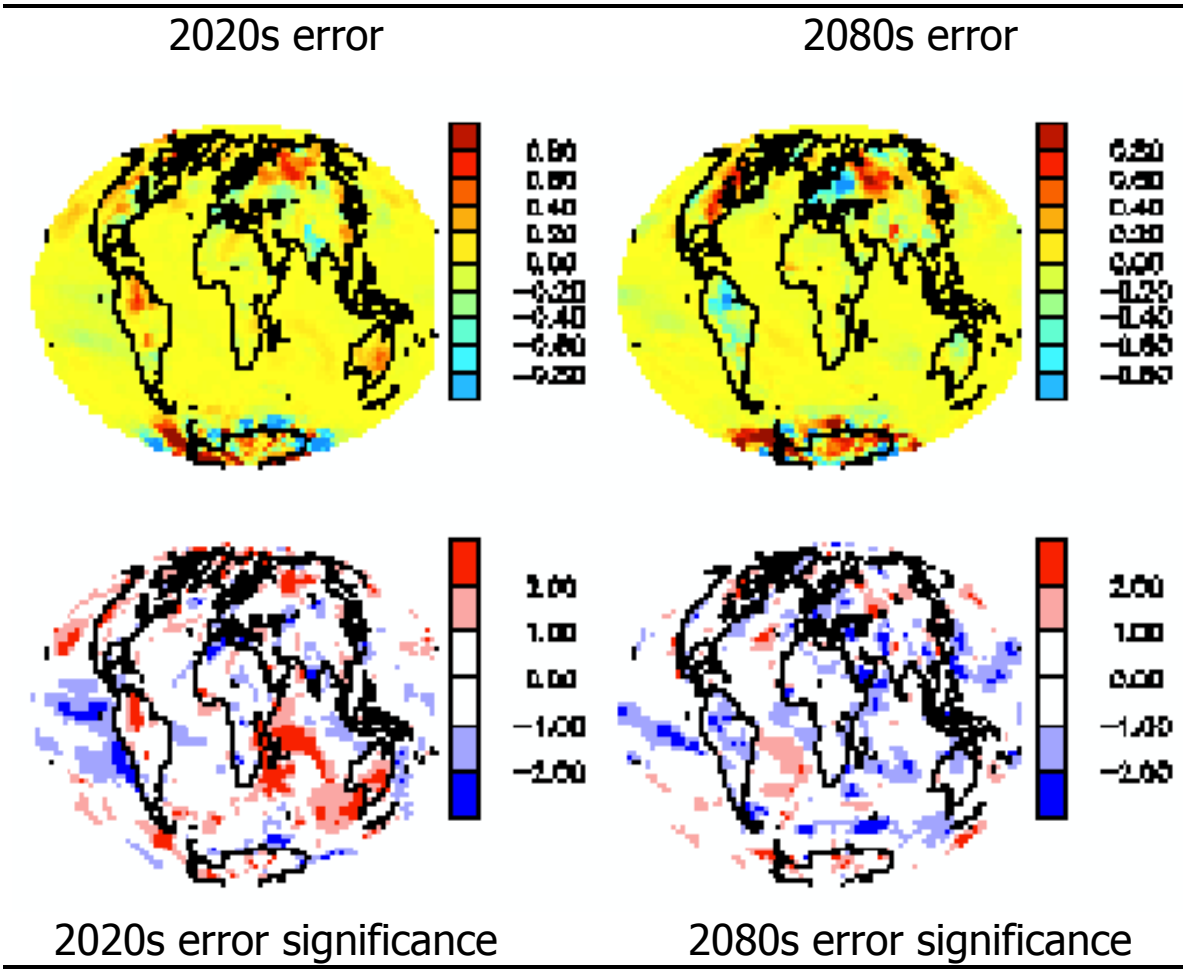
The estimated temperature interannual variance in Gd by scaling Ga pattern.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler, for Ga
modelled:	2020s [or 2080s] variance – pre-industrial
estimated:	b + (a * Gd scaler for 2020s [or 2080s]) – pre-industrial
error:	estimated – modelled for 2020s (**top left**) and 2080s (**top right**)
error sig:	error in standard deviations (**bottom**, Eq. 4.5)
	2020s (**bott. left**): 114 boxes (1.7%) <-2□ ; 862 boxes (12.6%) >2□
	2080s (**bott. right**): 145 boxes (2.1%) <-2□ ; 644 boxes (9.4%) >2□

Figure 5.14

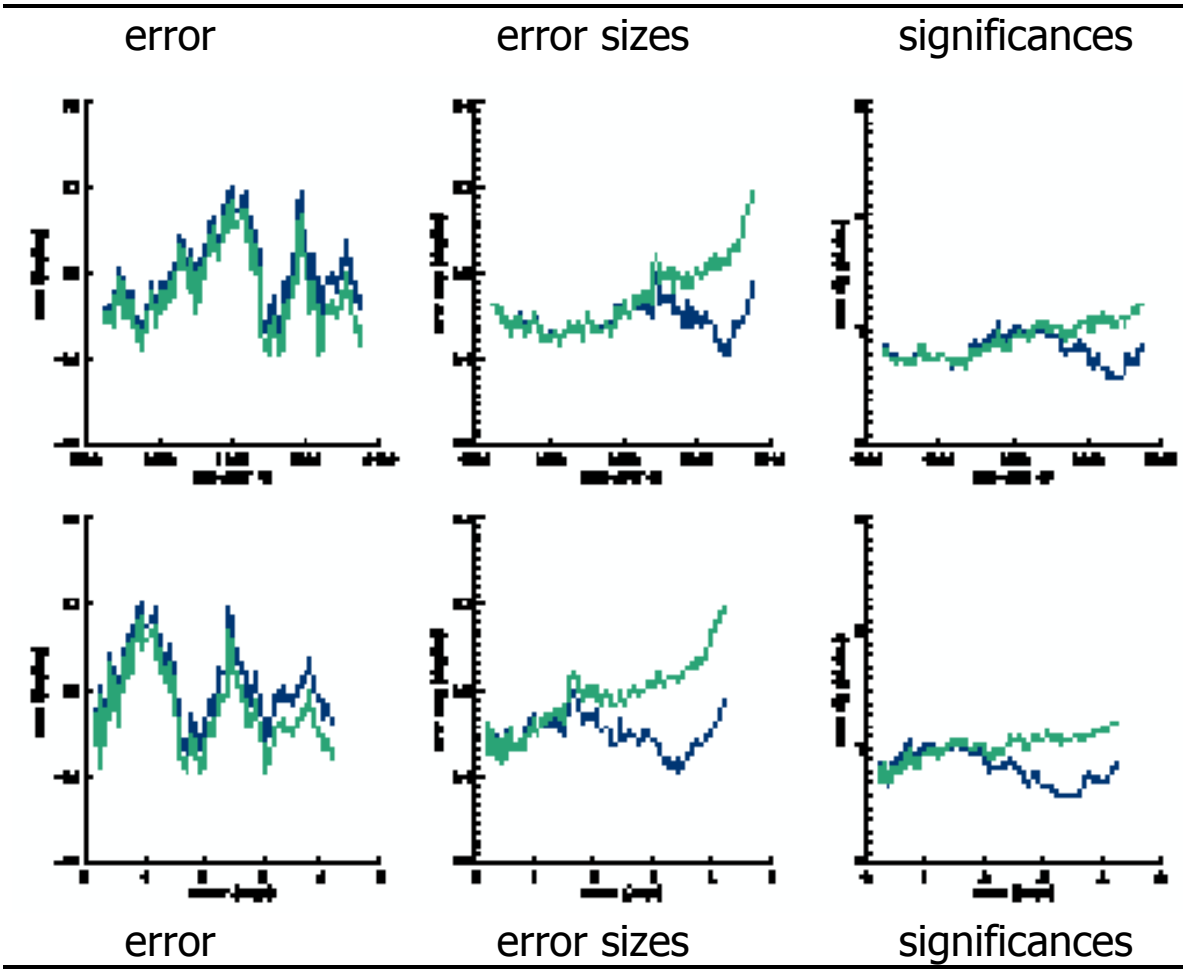
The estimated temperature interannual variance in Gd by scaling Gd pattern.



simulations:	HadCM2 Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler, for Gd
modelled:	2020s [or 2080s] variance – pre-industrial
estimated:	b + (a * Gd scaler for 2020s [or 2080s]) – pre-industrial
error:	estimated – modelled for 2020s (**top left**) and 2080s (**top right**)
error sig:	error in standard deviations (**bottom**, Eq. 4.5)
	2020s (**bott. left**): 131 boxes (1.9%) <-2□ ; 457 boxes (6.7%) >2□
	2080s (**bott. right**): 207 boxes (3.0%) <-2□ ; 127 boxes (1.9%) >2□

Figure 5.15

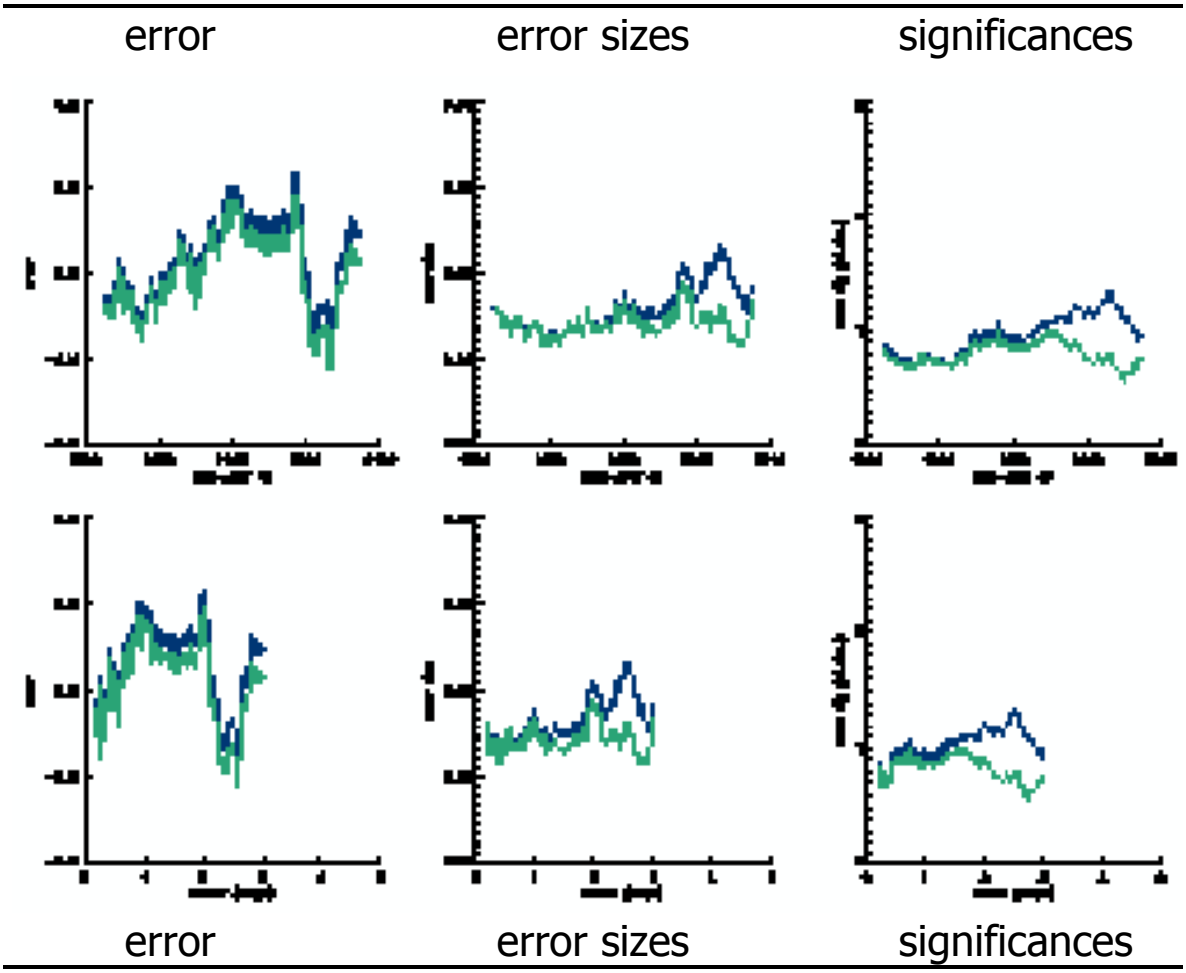
The estimated temperature interannual variance in Ga.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ (y=variance,x=scaler) for **Ga 1+2+3+4** and **Gd 1+2+3+4**
modelled:	variance – pre-industrial
estimated:	$b + a$ * Ga scaler – pre-industrial
error:	each box, and then globally averaged: estimated – modelled (**left**), and its magnitude (**centre**) error magnitude in standard deviations (**right**)
x-axis:	period mid-year (**top**), period mean scaler (°C, **bottom**)

Figure 5.16

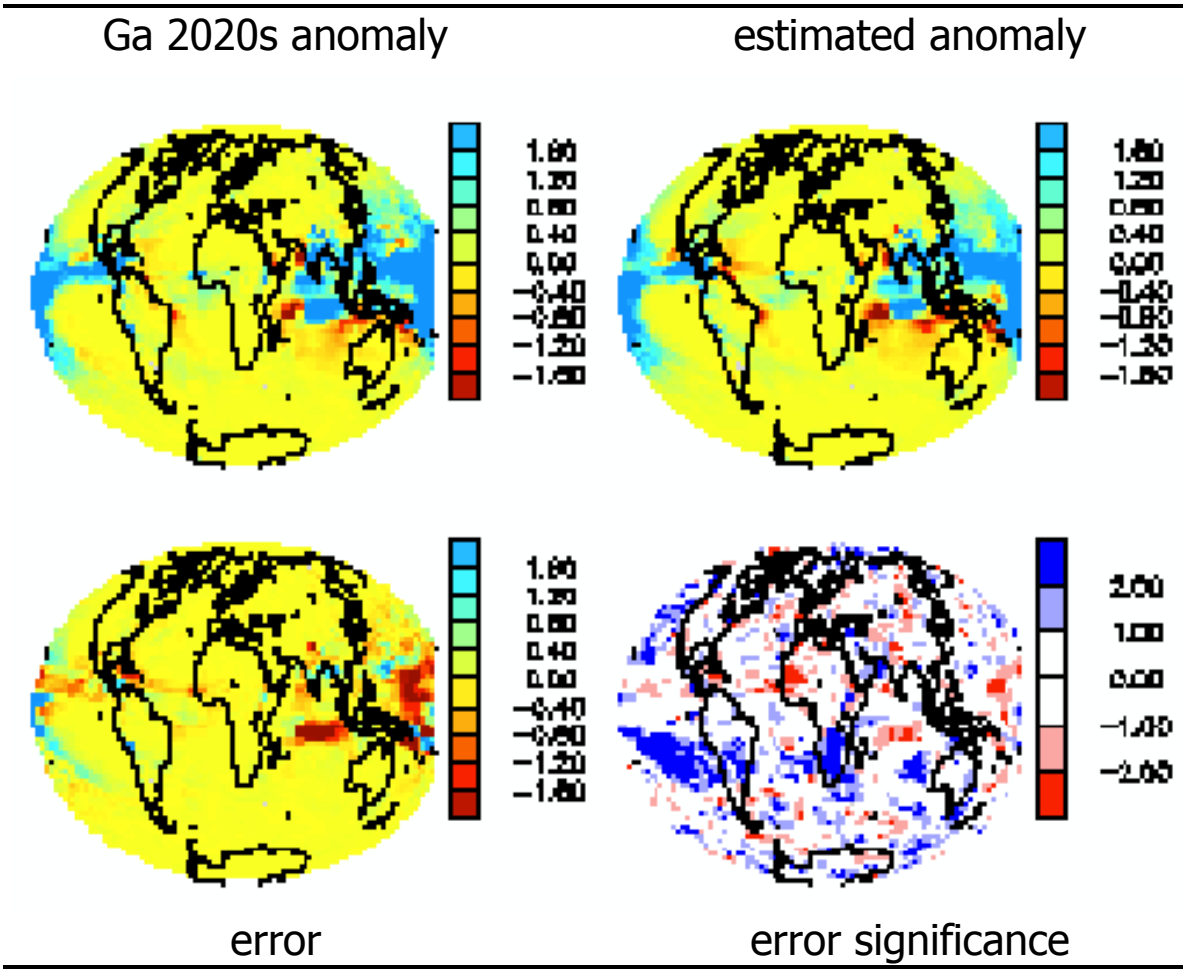
The estimated temperature interannual variance in Gd.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA temperature (°C)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ (y=variance,x=scaler) for Ga 1+2+3+4 and Gd 1+2+3+4
modelled:	variance – pre-industrial
estimated:	$b + a$ * Gd scaler – pre-industrial
error:	each box, and then globally averaged: estimated – modelled (**left**), and its magnitude (**centre**) error magnitude in standard deviations (**right**)
x-axis:	period mid-year (**top**), period mean scaler (°C, **bottom**)

Figure 5.17

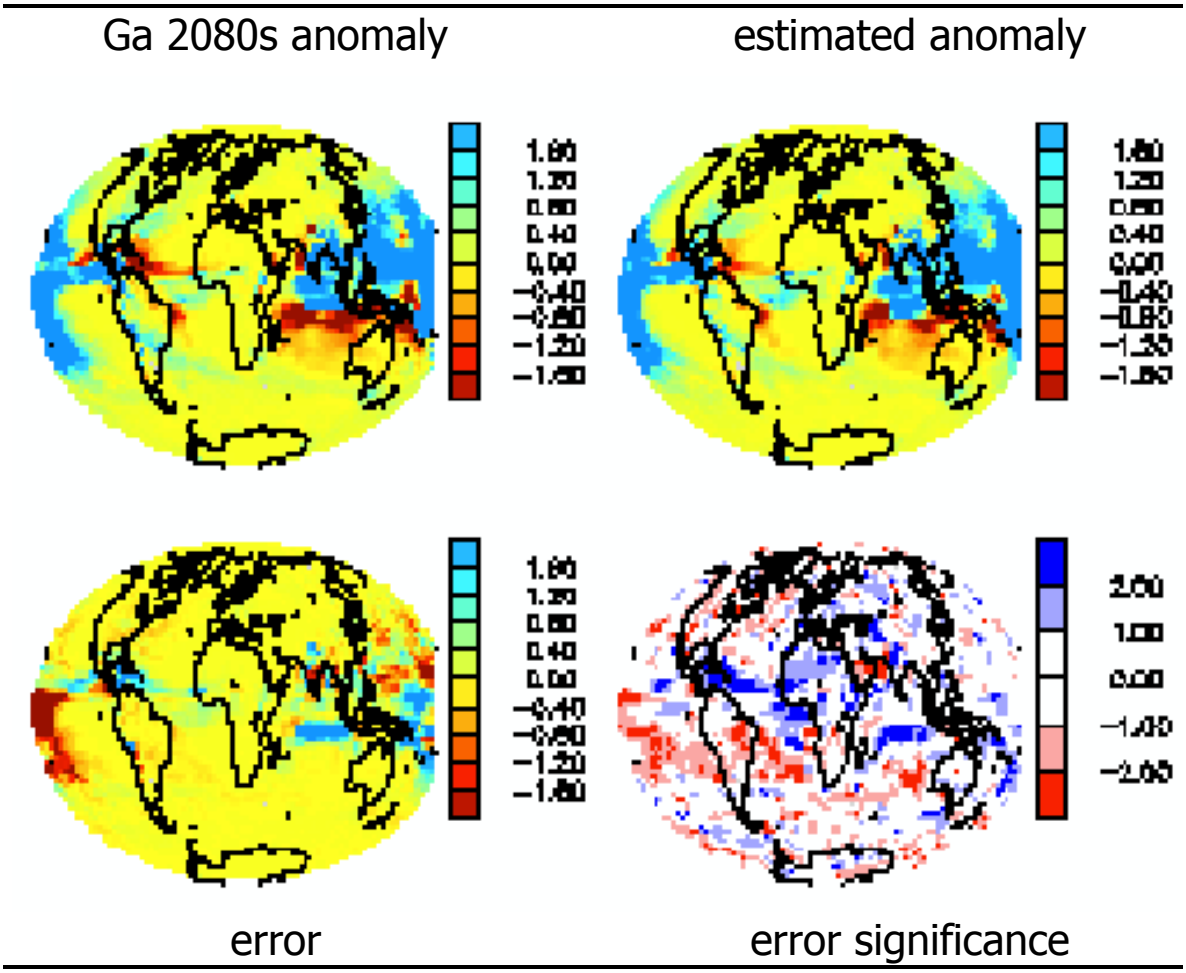
The estimated precipitation interannual variance in Ga in the 2020s.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
modelled:	2020s variance – pre-industrial (**top left**)
estimated:	b + (a * Ga scaler for 2020s) – pre-industrial (**top right**)
error:	estimated – modelled (**bottom left**)
error sig:	error in standard deviations (**bott. right**, Eq. 4.5)
	185 boxes (2.7%) <-2□ ; 605 boxes (8.9%) >2□

Figure 5.18

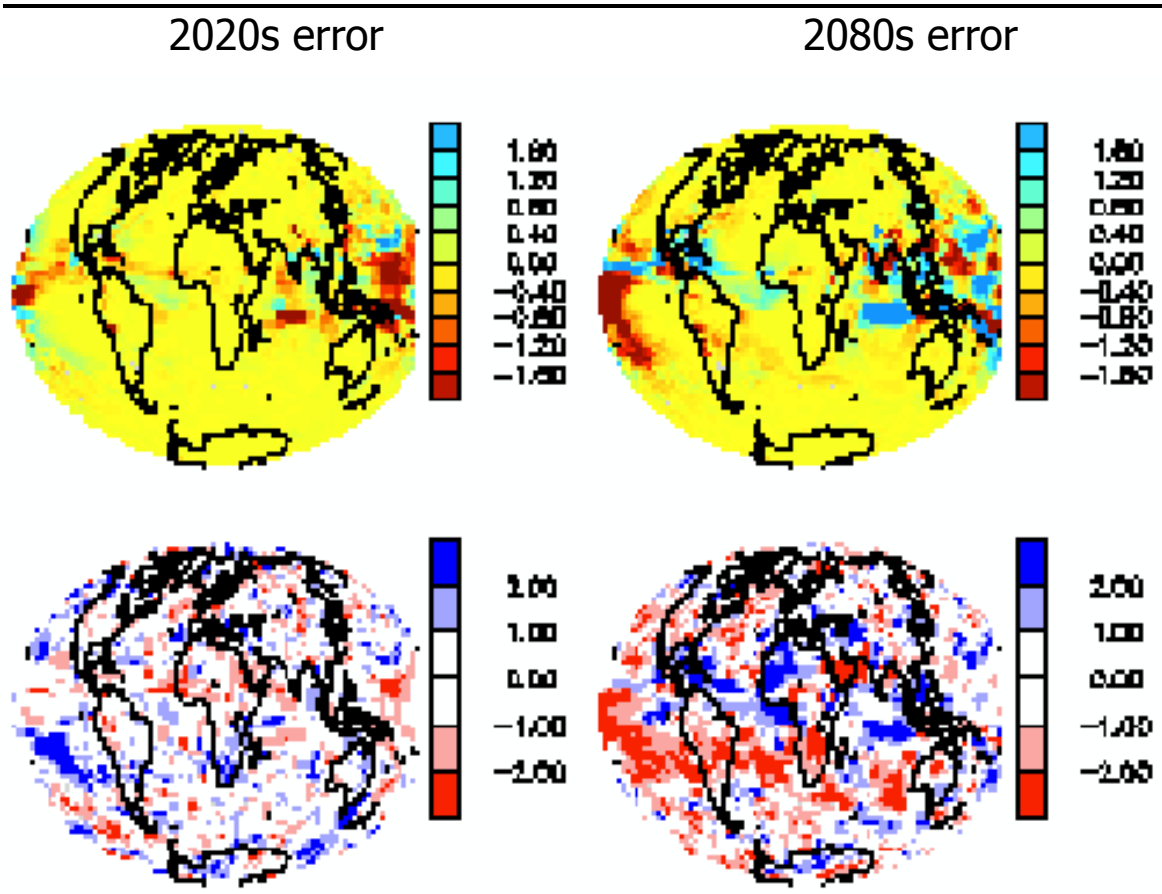
The estimated precipitation interannual variance in Ga in the 2080s.



simulations:	HadCM2 Ga 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler
modelled:	2080s variance – pre-industrial (**top left**)
estimated:	b + (a * Ga scaler for 2080s) – pre-industrial (**top right**)
error:	estimated – modelled (**bottom left**)
error sig:	error in standard deviations (**bott. right**, Eq. 4.5)
	371 boxes (5.4%) <-2□ ; 271 boxes (4.0%) >2□

Figure 5.19

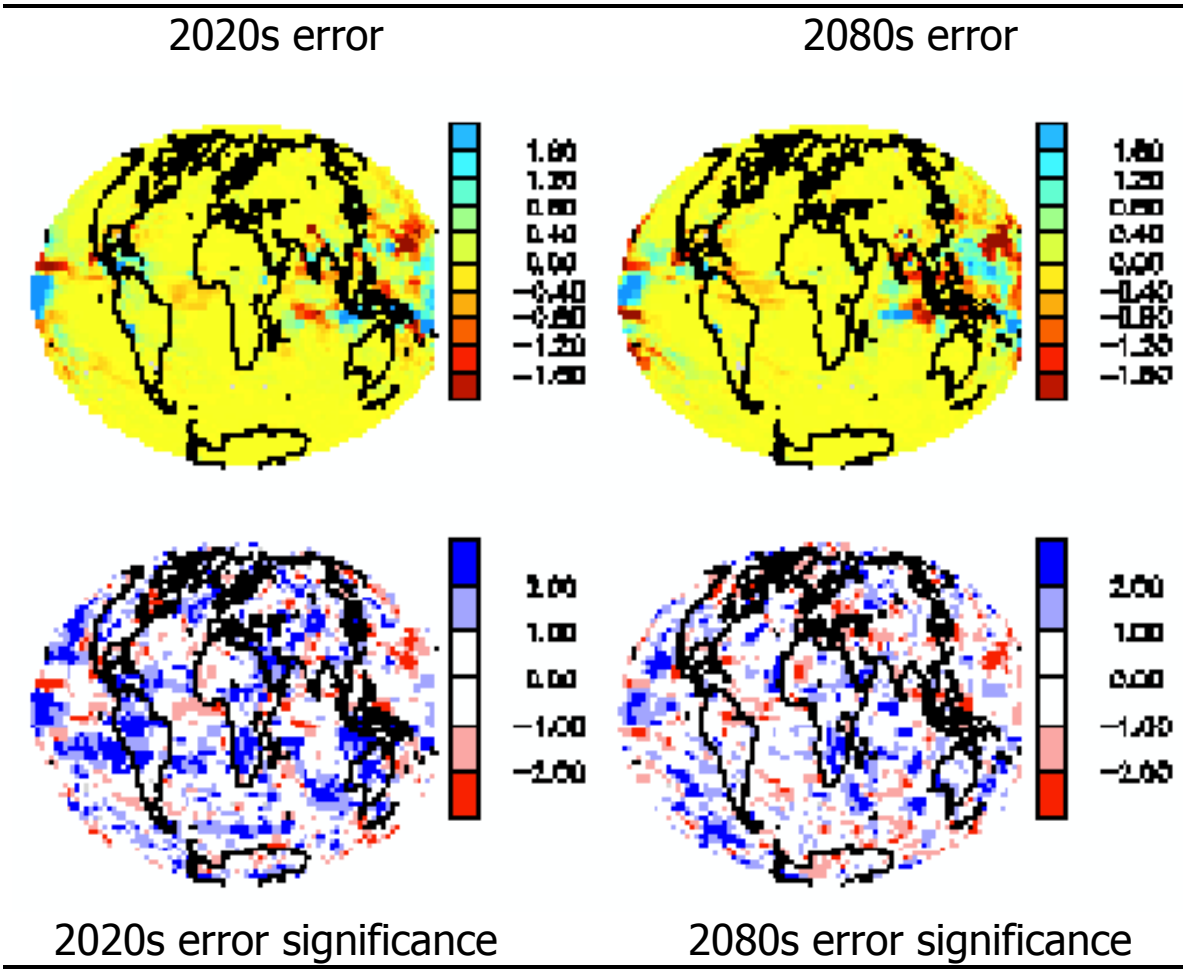
The estimated precipitation interannual variance in Ga by scaling Gd pattern.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler, for Gd
modelled:	2020s [or 2080s] variance – pre-industrial (**top left**)
estimated:	b + (a * Ga scaler for 2020s [or 2080s]) – pre-industrial (**top right**)
error:	estimated – modelled for 2020s (**top left**) and 2080s (**top right**)
error sig:	error in standard deviations (**bottom**, Eq. 4.5)
	2020s (bott. left): 285 boxes (4.2%) <-2□ ; 469 boxes (6.9%) >2□
	2080s (bott. right): 838 boxes (12.3%) <-2□ ; 479 boxes (7.0%) >2□

Figure 5.20

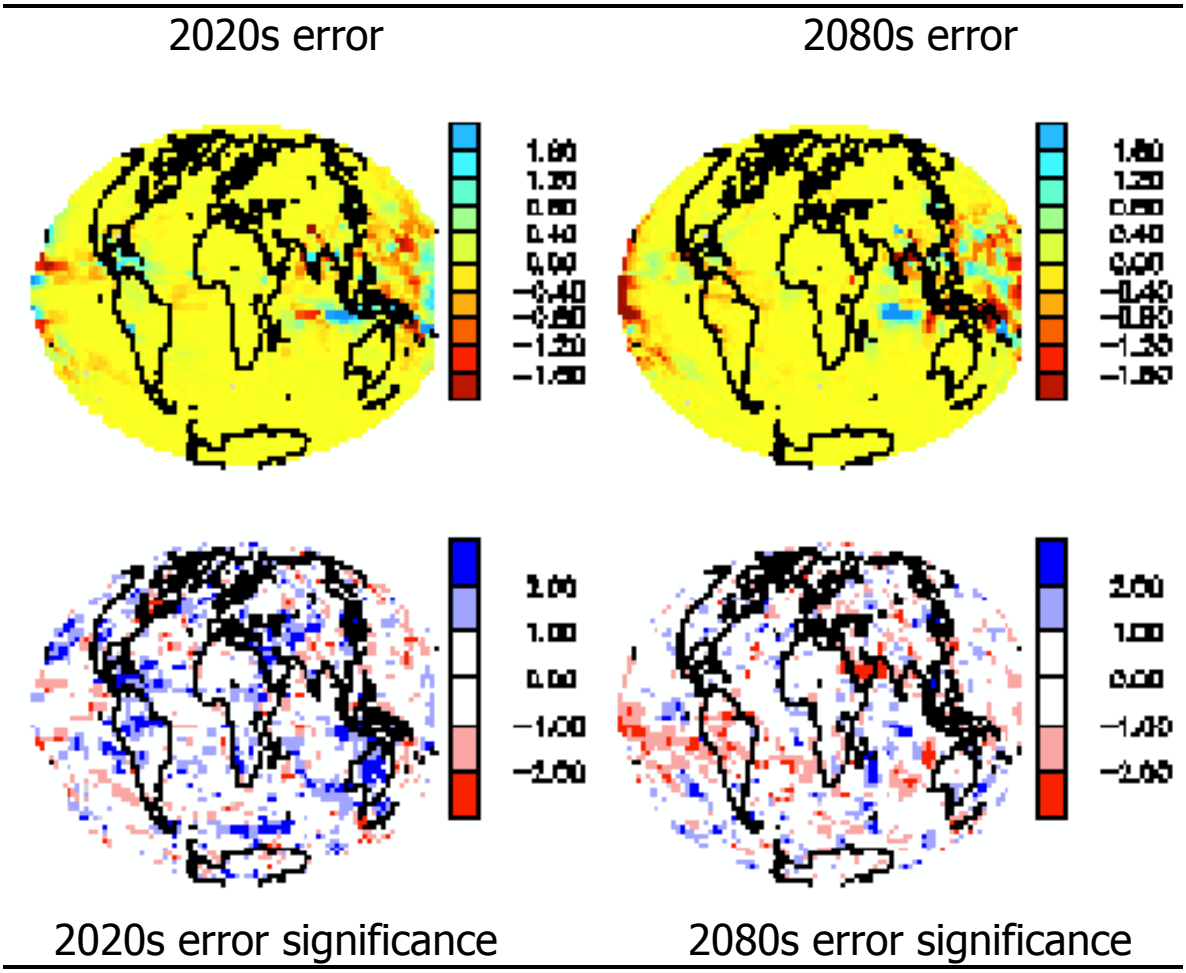
The estimated precipitation interannual variance in Gd by scaling Ga pattern.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler, for Ga
modelled:	2020s [or 2080s] variance – pre-industrial (**top left**)
estimated:	b + (a * Gd scaler for 2020s [or 2080s]) – pre-industrial (**top right**)
error:	estimated – modelled for 2020s (**top left**) and 2080s (**top right**)
error sig:	error in standard deviations (**bottom**, Eq. 4.5)
	2020s (bott. left): 223 boxes (3.3%) <-2□ ; 984 boxes (14.4%) >2□
	2080s (bott. right): 285 boxes (4.2%) <-2□ ; 473 boxes (6.9%) >2□

Figure 5.21

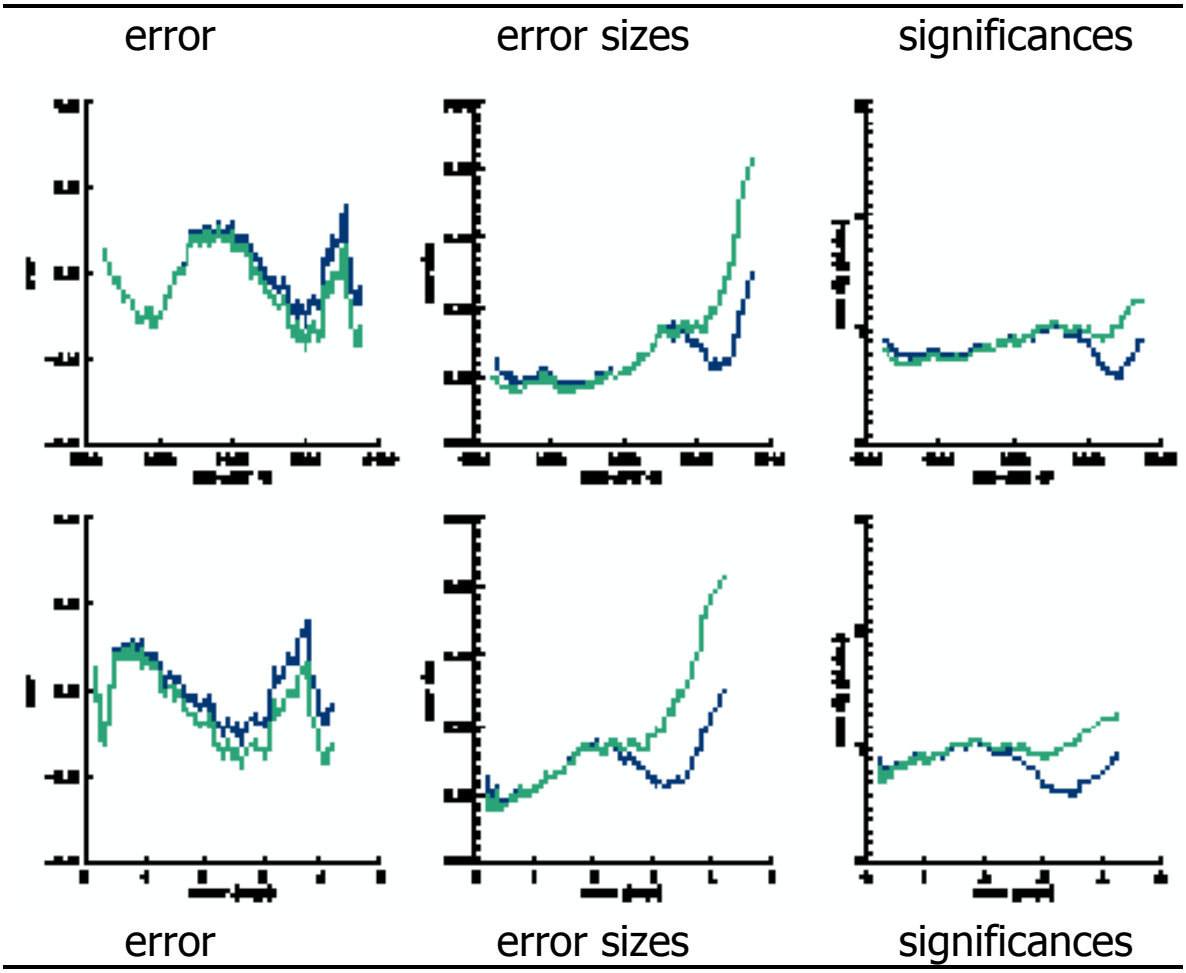
The estimated precipitation interannual variance in Gd by scaling Gd pattern.



simulations:	HadCM2 Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ where y is original variance and x is the scaler, for Gd
modelled:	2020s [or 2080s] variance – pre-industrial (**top left**)
estimated:	b + (a * Gd scaler for 2020s [or 2080s]) – pre-industrial (**top right**)
error:	estimated – modelled for 2020s (**top left**) and 2080s (**top right**)
error sig:	error in standard deviations (**bottom**, Eq. 4.5)
	2020s (bott. left): 111 boxes (1.6%) <-2□ ; 515 boxes (7.6%) >2□
	2080s (bott. right): 192 boxes (2.8%) <-2□ ; 118 boxes (1.7%) >2□

Figure 5.22

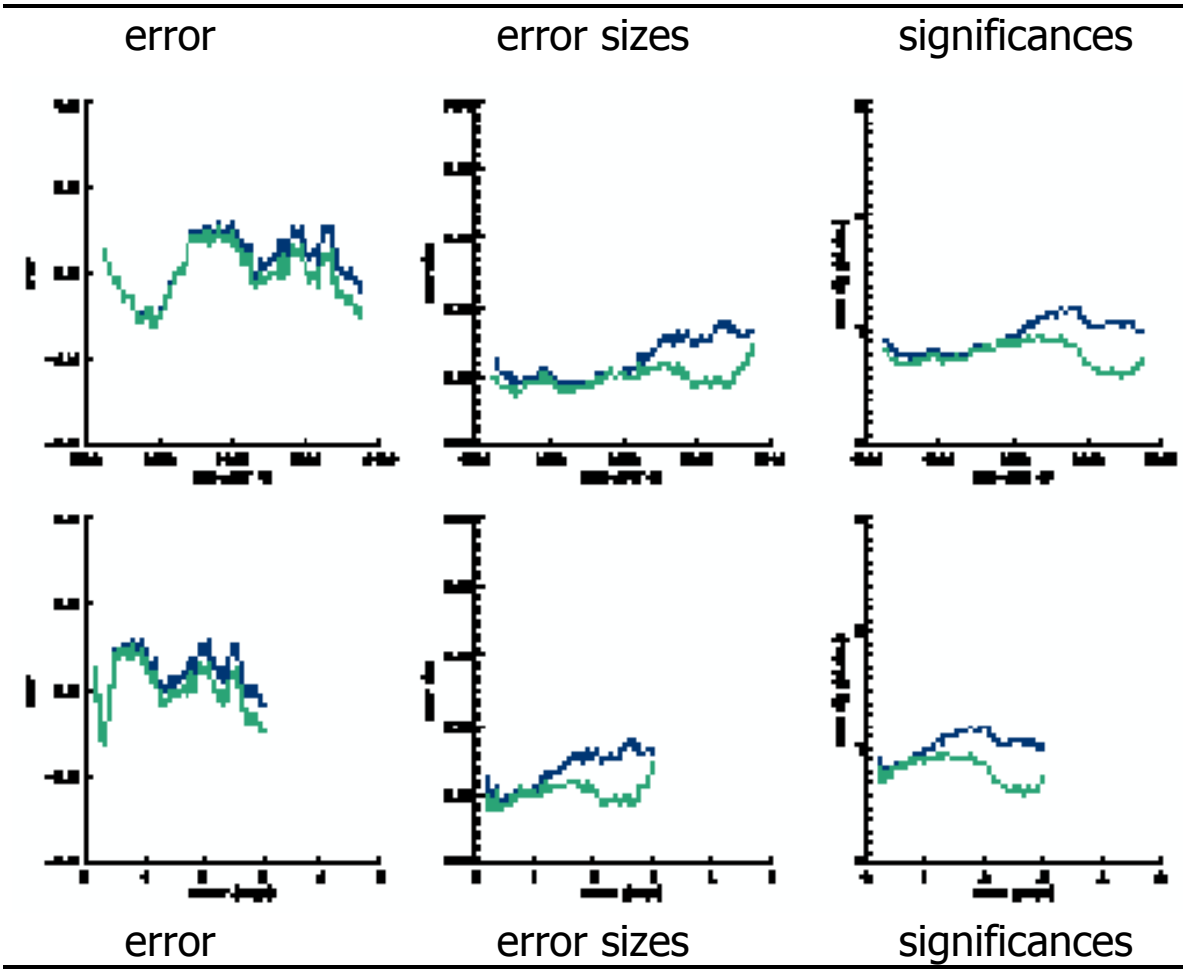
The estimated precipitation interannual variance in Ga.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ (y=variance,x=scaler) for Ga 1+2+3+4 and Gd 1+2+3+4
modelled:	variance – pre-industrial
estimated:	$b + a$ * Ga scaler – pre-industrial
error:	each box, and then globally averaged:
	estimated – modelled (**left**), and its magnitude (**centre**)
	error magnitude in standard deviations (**right**)
x-axis:	period mid-year (**top**), period mean scaler (°C, **bottom**)

Figure 5.23

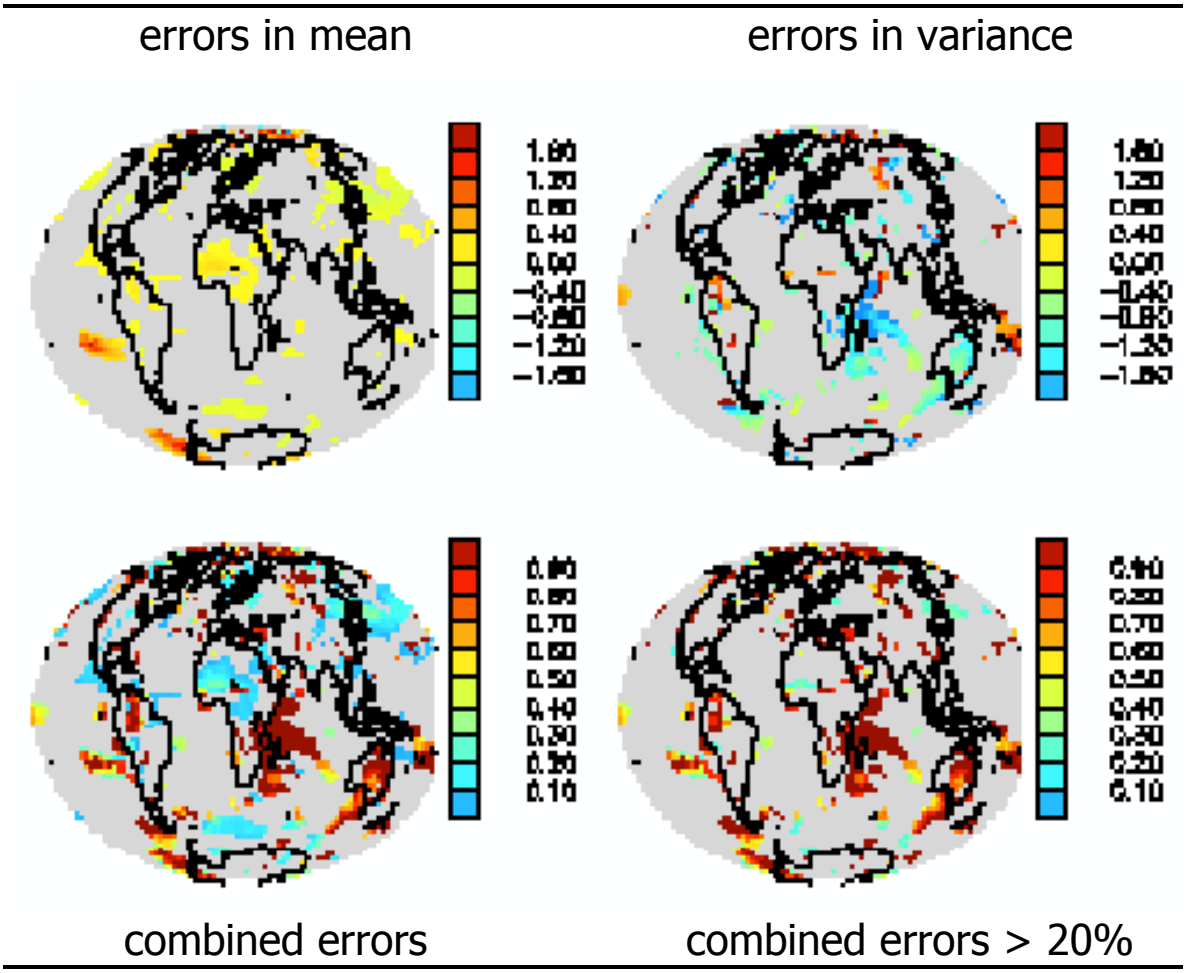
The estimated precipitation interannual variance in Gd.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	interannual variance of grid-box JJA precipitation (mm/day)
original:	overlapping 30y periods, each detrended
variance:	inter-annual variance of 30 values, averaged among 4 simulations
patterns:	$y=ax+b$ (y=variance,x=scaler) for Ga 1+2+3+4 and Gd 1+2+3+4
modelled:	variance – pre-industrial
estimated:	$b + a$ * Gd scaler – pre-industrial
error:	each box, and then globally averaged:
	estimated – modelled (**left**), and its magnitude (**centre**)
	error magnitude in standard deviations (**right**)
x-axis:	period mid-year (**top**), period mean scaler (°C, **bottom**)

Figure 5.24

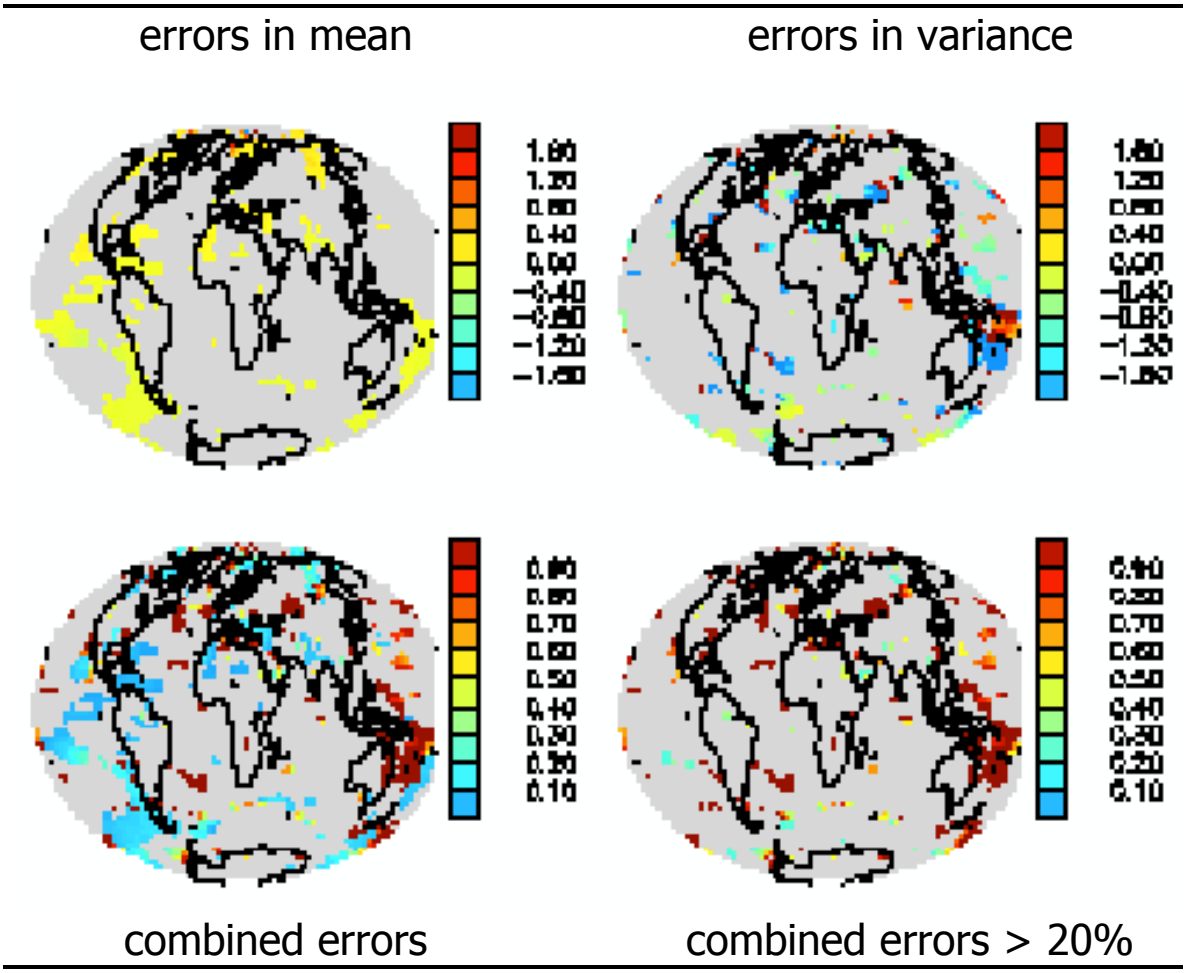
Significant errors in estimating Gd 2020s temperature from the Ga pattern.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	grid-box JJA temperature (°C) mean and interannual variance
errors:	error in estimate of Gd 2020s by scaling Ga response pattern
significance:	error / modelled, masked by statistical significance (**top**)
combined:	maximum absolute value of top left and top right (**bottom left**)
comb > 20%:	bottom left where values are greater than 0.2
	threshold exceeded in 19.9% of boxes

Figure 5.25

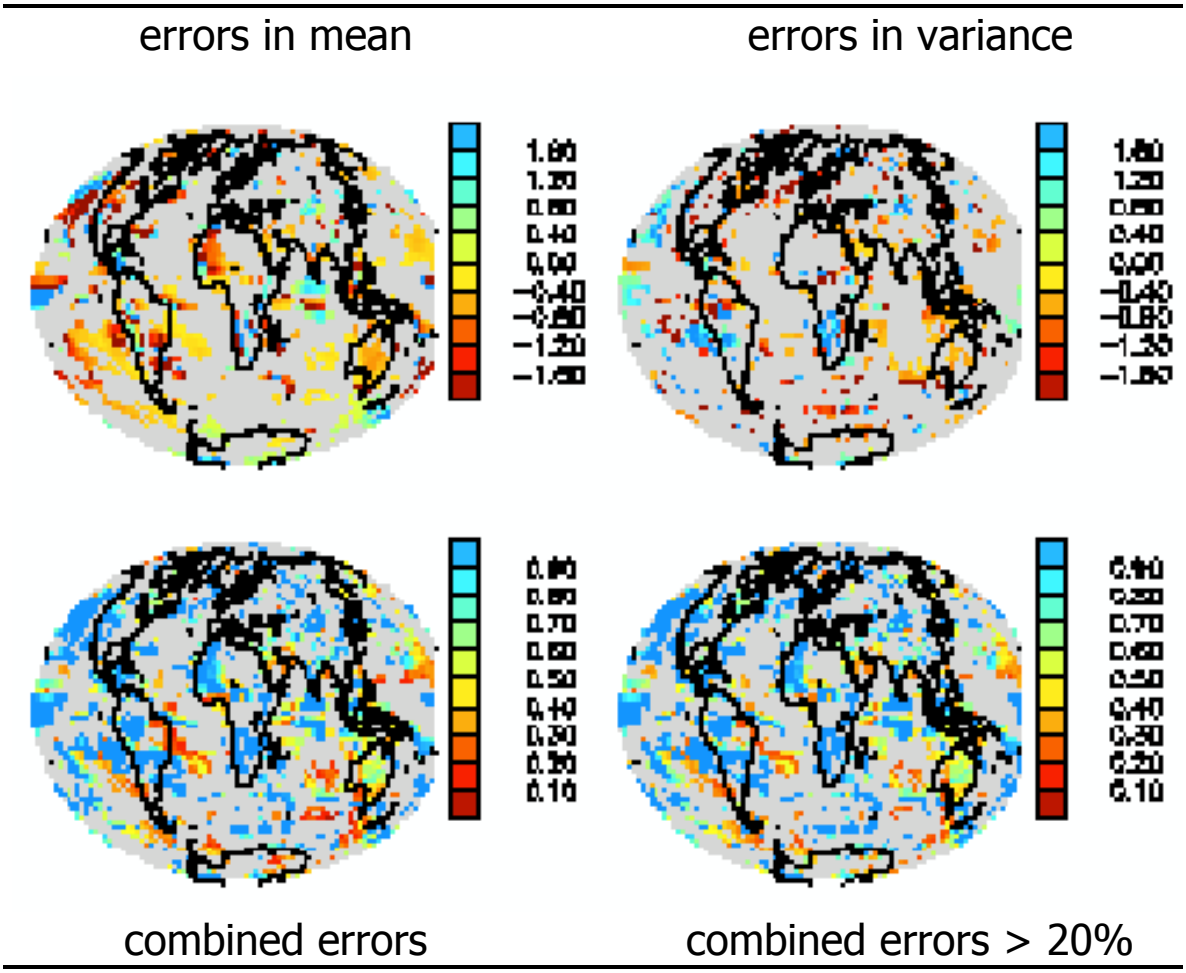
Significant errors in estimating Gd 2080s temperature from the Ga pattern.



simulations: HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic: grid-box JJA temperature (°C) mean and interannual variance
errors: error in estimate of Gd 2080s by scaling Ga response pattern
significance: error / modelled, masked by statistical significance (**top**)
combined: maximum absolute value of top left and top right (**bottom left**)
comb > 20%: bottom left where values are greater than 0.2
threshold exceeded in 12.4% of boxes

Figure 5.26

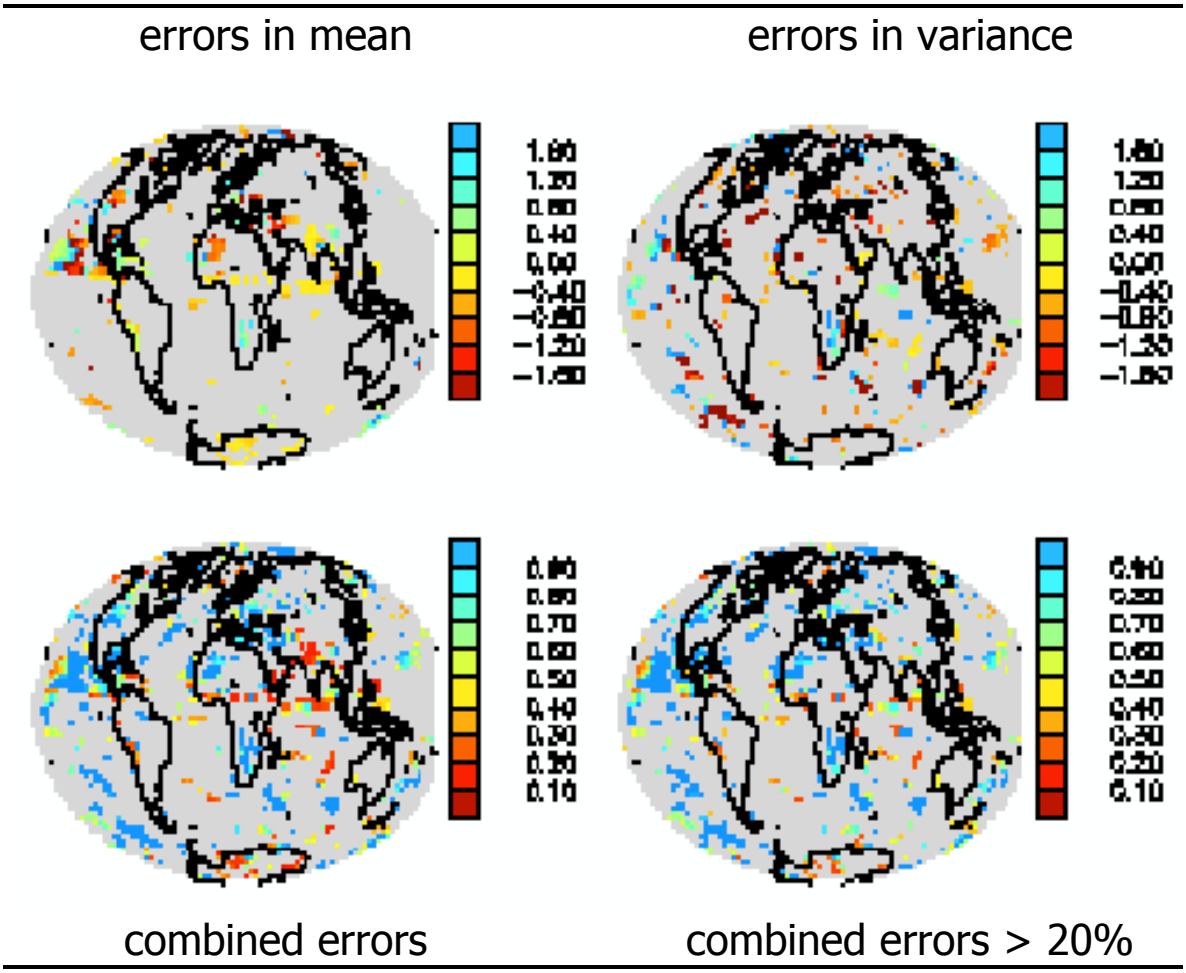
Significant errors in estimating Gd 2020s precipitation from the Ga pattern.



simulations:	HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic:	grid-box JJA precipitation (mm/day) mean and interannual variance
errors:	error in estimate of Gd 2020s by scaling Ga response pattern
significance:	error / modelled, masked by statistical significance (**top**)
combined:	maximum absolute value of top left and top right (**bottom left**)
comb > 20%:	bottom left where values are greater than 0.2
	threshold exceeded in 36.3% of boxes

Figure 5.27

Significant errors in estimating Gd 2080s precipitation from the Ga pattern.



simulations: HadCM2 Ga 1+2+3+4, Gd 1+2+3+4
diagnostic: grid-box JJA precipitation (mm/day) mean and interannual variance
errors: error in estimate of Gd 2080s by scaling Ga response pattern
significance: error / modelled, masked by statistical significance (**top**)
combined: maximum absolute value of top left and top right (**bottom left**)
comb > 20%: bottom left where values are greater than 0.2
threshold exceeded in 18.0% of boxes

6. May we regionalise?

6.1 Introduction

In chapters 4 and 5 we have examined the possibility of scaling spatial patterns of climate change by the global annual temperature, with the aim of enhancing a probabilistic approach to climate scenario construction. Our analysis in those chapters dealt with a single model (HadCM2) at the grid-box scale. Our use of individual grid-boxes did not imply that the model's representation of climatic change accurately represents the real world at the level of the individual grid-box.[208] Indeed, we did not discuss the extent to which a model is consistent with the real world; instead we examined the extent to which the model exhibits a consistent – and scaleable – pattern of change between forcing scenarios at the grid-box scale.

However, doubts over the credibility of GCMs on regional scales are of critical importance for the credibility of climate scenarios. Moreover, the literature contains a wide variety of approaches to the definition of 'regions' for climatic purposes.[209] Therefore the question that we address in this chapter has a wider relevance beyond pattern scaling. We ask whether we may aggregate grid-boxes into regions for the purpose of providing scaled estimates of future regional climate change for climate scenarios. We might put this in abbreviated form: may we regionalise (GCM output)?

To address this question we consider the climate changes found in regions rather than in individual grid-boxes. There is a near-infinite number of possible 'regions' that we might construct on the HadCM2 grid.[210] Yet the climate scenarios that we construct may depend on the principles we employ in constructing regions. There are two key issues involved if we are to select grid-boxes using objective principles:

---

[208] However, at least some climate modellers expect to be able to detect anthropogenic climate change at all spatial scales by the end of the 21st century (*e.g.* Stott and Tett, 1998). Moreover, Osborn *et al.* (1999b) found that HadCM2 accurately represents observed climate at the grid-box scale for at least some grid-boxes. See Mitchell and Hulme (1999) for a review of scientific views on the credibility of GCMs at regional scales.

[209] We documented some of these approaches in section 2.10.

[210] We define a 'region' as an aggregation of one or more grid-boxes.

- the spatial scale: how many grid-boxes should be aggregated together?
- the zoning: which particular grid-boxes should be aggregated together?

In order to investigate these issues we construct a number of sets of regions (section 6.2). In section 6.3 we consider the similarities and differences between the region sets. In the next three sections we examine the effects that our choices of spatial scale and zonation have on our estimates of climate:

- We consider the amount of information that is lost through regionalising, using measures of information loss that are independent of responses to increased radiative forcing (section 6.4).
- We consider the effects of regionalising on signal-to-noise ratios in the context of increased radiative forcing (section 6.5).
- We apply pattern scaling to the different region sets and compare the accuracies achieved (section 6.6).

In the final section (6.7) we draw our conclusions.

6.2 Region set construction

The reason for constructing different region sets is to compare the effects the different principles of aggregation have on our estimates of climate change. Therefore it is important that our region sets are similar, *except* regarding the principles of aggregation they employ. We specify a number of requirements with which each region set must comply. Each region set must be:

- on the HadCM2 model grid;
- comprised of the 2271 land boxes in the model;[211] each box must be included, and no other boxes may be included;
- each grid-box must be allocated to precisely one region.

In the paragraphs below we describe the reasons for constructing each of our seven region sets and the principles by which each was constructed. These principles are summarised in Table 6.1.

1. Land Boxes. This region set provides a reference set against which each of the other region sets may be compared. It treats each of the 2271 land boxes as a region constituted solely by that box.

[211] The total of 2271 excludes the single grid-box at the South Pole.

2. Countries. Many assessments of climate-related impacts are conducted using regions representing national boundaries. In this region set we allocate each grid-box to a region that represents a single country. National boundaries form the sole basis on which a grid-box is allocated to one region rather than another, so the spatial scale varies with the size of a country. Each grid-box is assigned to the country that has the single largest share of the surface area within that grid-box.[212] One constraint that governs how political boundaries are understood is the requirement that any single region must be made up of a contiguous set of grid-boxes.[213] Hence where countries hold sovereignty over areas of land that are distinct from the 'mainland' (such as Alaska or Kaliningrad), the separated territories are treated as regions in their own right, with their own political boundary. In some cases the allocation of a grid-box to one region rather than another depends on which country is judged to hold sovereignty of disputed territories (such as Kashmir or Tibet).

3. Giorgi. The set of regions described in Giorgi and Francisco (2000) was used as the basis for a region set employing the principles followed in the construction of the IPCC regions. The set of regions defined by Giorgi and Francisco (2000) was not precisely the same as the evolving set in the IPCC reports in the 1990s,[214] but those authors consciously followed the lead of the IPCC set in constructing their own set: this was the case not merely in their zoning[215] and scale[216] choices, but

---

[212] Shares of area in a grid-box were judged by visual inspection of a political map overlaid onto the HadCM2 grid. Shares of area do not include ocean areas within a grid-box, but they do include any internal bodies of water. This rule is based on a particular extension to the general rule of national boundaries: internal waters fall within the sovereignty of a nation, but the sovereignty of coastal waters is often not clear.

[213] With reference to the model grid, a region is said to be contiguous when all its constituent grid-boxes adjoin the whole at least one corner. A region may still be considered contiguous if it contains boxes representing the coast and adjacent islands, or an archipelago.

[214] Although the Third Assessment Report of the IPCC is expected to incorporate the set of regions defined by Giorgi and Francisco (2000).

[215] The rectangular regions in Mitchell *et al.* (1990) were selected "to represent a range of climates" (p155). Giorgi and Francisco (2000) selected "regions of simple shape" that "represent different climatic regimes and physiographic settings" (p172).

[216] Mitchell *et al.* (1990) considered "several regions of about 4,000,000 km^2" (p155). In fact, four of their five regions are larger than 4,000,000 km^2, and the average area is larger than 5,000,000 km^2. Giorgi and Francisco (2000) selected a "manageable number" of rectangles of a similar size, each side of which amounted to "several thousand km" (p172).

also in the similarity of some of their regions.[217] We use as our basis the Giorgi and Francisco (2000) set rather than the IPCC set because the former comes much closer to covering the globe than the latter.[218] We are only able to use the Giorgi and Francisco (2000) set as a *basis*, rather than simply reproducing it, because of the omissions in their set. We accounted for Antarctica simply by making it a single region; the surface area it represents lies within the range of the other regions in the set. More notable were the land boxes omitted from the Giorgi and Francisco (2000) set but not explicitly mentioned by them.[219] The requirements that we have laid down for all our region sets are such that the missing boxes must be incorporated. We did this in a way that tampers with the boundaries from Giorgi and Francisco (2000) as little as possible: the missing grid-boxes were all added to existing regions, the regions are still formed by the land boxes within non-overlapping rectangles, and they still represent the areas originally intended.

4. Large Geometric. The Giorgi set described above inherits from Giorgi and Francisco (2000) the use of climatic information to define the regions. Those authors admitted that theirs "was a subjective choice, and different regional configurations could be devised" (p172). In our sixth and seventh sets, described below, we use alternative sources of *climatically*-related information to devise "different regional configurations". In this set and the next we provide the means of examining whether the use of climatically-related information actually adds value to the regional configuration chosen. We achieve this by covering the land grid-boxes outside Antarctica[220] with the same number of regions (21) as were employed in our Giorgi set. However, rather than employing climatic criteria, each

[217] Among the "criteria" for selection followed by Giorgi and Francisco (2000) is that "some of the regions are close" to the IPCC set (p172). (Is this a criterion or a coincidence?) They note Australia (which is identical); SE Asia, Central North America, and the Mediterranean Basin (all of which are similar but not identical); and East Asia (which is not similar at all; the IPCC region corresponds to parts of six of Giorgi and Francisco's regions).

[218] The intention of Giorgi and Francisco (2000, p170) was to cover "all land areas of the globe (with the exception of Antarctica)". At its greatest extent during the 1990s (Kattenberg *et al.*, 1996), the IPCC set only covered a third of the non-Antarctic land grid-boxes in HadCM2.

[219] In their introduction Giorgi and Francisco (2000, p170) claimed to cover "all land areas of the globe (with the exception of Antarctica)", but later they watered down this claim: "we wanted to *approximately* cover all land areas" (p172, our italics). In fact a number of land grid-boxes were omitted, representing SE Alaska, the Caribbean islands, S Central America, New Zealand, N Siberia, and Iceland.

[220] We selected regions for Antarctica that represented surface areas and grid-box totals comparable to the other 21 regions.

region is based on a *rectangle* of grid-boxes[221] selected with as much geometric efficiency as possible:

- to cover a similar number of land grid-boxes in each rectangle;
- to minimise the inclusion of coastal seas within each rectangle;
- to avoid any overlap between rectangles.

Thus the spatial scale of the regions is the same in both this set and the Giorgi set,[222] but the zoning is different.

5. Small Geometric. The countries in the world cover a wide range of sizes, and that range of spatial scales is reflected in the Countries region set. However, given the importance of the choice of spatial scale, it is appropriate to examine particular spatial scales more closely. The Large Geometric set provides an opportunity to examine regions containing approximately 60 to 90 grid-boxes. To complement these larger spatial scales we construct a Small Geometric set using identical zoning principles to those employed for the Large Geometric Set, but constraining the regions outside Antarctica[223] to contain between 20 and 30 grid-boxes, at an average of 25 grid-boxes. It should be noted that since the two Geometric sets and the Countries set have been zoned without any climatically-relevant information, they may be treated as being zoned in the same way. Therefore for zone-independent examinations of spatial scale we have available both a wide range of spatial scales (the Countries set) and more detailed examinations of two particular spatial scales (the Geometric sets).

6. Topographic. Having provided region sets that will permit us to examine the choice of spatial scale, we provide two further region sets that will permit us to examine the value of employing climatically-related information in zoning. Here we zone on the basis of the model topography. Topography is an important climatological influence both in the real world and in GCMs, particularly for precipitation. It therefore seems plausible that we might – in some sense – improve our region set by selecting regions that represent topographic units in the model.

---

[221] The rectangles are made up of both land and ocean grid-boxes from HadCM2, but the region corresponding to a rectangle only includes the land grid-boxes within that rectangle.

[222] That is, the *mean* spatial scale of the regions is the same in both our Giorgi and Large Geometric sets. The *range* and *standard deviation* of sizes is much less within the Large Geometric set than it is in the Giorgi set.

[223] For the Large Geometric set the region sizes for Antarctica represented a compromise between equality in the number of grid-boxes, and in surface area, with the Rest of the World. In the Small Geometric set the numbers of grid-boxes in the Antarctic regions are the same as in the Rest of the World, with no attention paid to surface areas.

In order to make an effective evaluation of the effects of zoning thus we construct the non-Antarctic regions[224] at the same spatial scale as was employed in the Small Geometric set.[225] An example of the information we use is given in Figure 6.1.[226] Each region is zoned to represent a single topographic unit of the appropriate spatial scale. The topographic unit might be a mountain range (*e.g.* Andes), plain, river basin (*e.g.* Amazonia), coastline (*e.g.* Caribbean coast), peninsular (*e.g.* Central America), or archipelago. The regions are contiguous and as compact as possible, although if a marked topographic feature is present its outline is followed despite its irregularities.

Climatic. The zoning of the Giorgi set is loosely climatic, to the extent that the regions appear to be constructed on the basis of areas of coherence under present conditions. We take this approach to its reasonable limits by carrying out a principal component analysis on the HadCM2 control simulation, and identifying regions that exhibit coherent variability in the control. If much of the climate change in response to radiative forcing is manifested through modes of natural variability,[227] it is possible that the optimal method of zoning may be to base the regions on the modes of natural variability in the control. We had to make a number of choices in the principal component analysis:

- **Variable**. We used near-surface (1.5m) temperatures only. This was partly for simplicity, partly because this provided sufficient information to identify appropriately-sized regions, and partly because the Topographic set is more suited to identifying changes in precipitation than in temperatures, which therefore complements this set.
- **Season**. Although in practice information for each of the seasons is likely to be required, it is unlikely that a specific region set would be constructed for each season. Therefore we used annual means.
- **Averaging Period**. We selected the multi-decadal time period as being of primary interest, and smoothed the time series of annual temperature at each grid-box with a 30-year Gaussian filter.

---

[224] In the Topographic set the surface areas of the regions in Antarctica fall within the range of the Rest of the World, with no attention paid to the number of grid-boxes.

[225] In the Small Geometric set the number of grid-boxes in a non-Antarctic region averaged 25, with a range of 20-30. In the Topographic set the average is still 25, but the range is larger (15-35) to allow more flexibility in the attempt to represent topographic units. Some sub-arctic regions are allowed larger numbers of grid-boxes.

[226] The HadCM2 topography was overplotted on the HadCM2 land-sea mask, and major rivers were also overplotted. These plots were used to identify the topographic features.

[227] One advocate of this theory is Tim Palmer. For examples see Palmer (1993) and Corti *et al.* (1999).

- **Detrending**. Since there is a small amount of drift in the control[228] we detrended the smoothed time series: we calculated the best-fit line ($y=ax+b$) by least squares regression and removed the gradient from the time series.

A principal components analysis[229] was carried out on each of eight rectangles of continental size.[230] An example of the information we used is given for South America in Figure 6.2; the information was utilised as follows:

- The first four components were retained. (In South America PC 1 was of little use for region selection because its projection consisted mainly of a land-sea contrast.)
- The contribution of each component was proportional to the proportion of the total variance that it explained.
- Regions were selected on the basis of their coherence in control variability, as judged from the components that explained the largest proportion of the total variance.
- In some places a large area exhibited coherence in one component, but contained a bipole in another component. (The high latitudes of South America form a single area in PC 4, but are split between Patagonia and the Pampas in PCs 2 and 3.) In such cases distinct regions were formed.
- Where there were similar boundaries between centres of variability in two or more components, the precise boundary was primarily determined by the component explaining the largest proportion of variance. (For example, the northern boundary of the Patagonian region was determined primarily by PC 2 rather than PC 3.)

There were no constraints upon the size of a region, except that each had to contain at least half-a-dozen grid-boxes and be smaller than a continent. Neither were there any constraints upon the shape or contiguity of a region, except that coherence was preferred where it was reasonable.

6.3 Comparing region sets

Global plots of the region sets constructed in section 6.2 are displayed in Figures 6.3 and 6.4. Plots for South America to complement those in Figures 6.1 (bottom left) and 6.2 (bottom right) are given in Figure 6.5. Summary statistics for each

---

[228] Johns *et al.* (1997) found a slight drift of +0.02°C per century during the first 400 years of the control simulation.

[229] Carried out on 1400 years using the Karhunen-Loeve expansion on the covariance matrix.

[230] Between them the rectangles covered almost all the land boxes of the HadCM2 grid, together with adjacent ocean grid-boxes; each contained between 400 and 1000 grid-boxes.

region set (including a breakdown of the land area into Antarctica and the Rest of the World) are displayed in Table 6.2, which provides the basis for the inter-comparison of the spatial *scales* of the region sets. The figures provide the basis for the inter-comparison of the *zonal* configurations of the region sets.

6.3a Comparison of groups

There are substantial similarities between the region sets. The spatial scales that we chose are such that five of the sets fall into two distinct groups:

- a large-scale group (Giorgi and Large Geometric),
- a small-scale group (Small Geometric, Topographic, and Climatic).

The large-scale group was formed because we made the Large Geometric set equivalent in scale to the Giorgi set. The Small Geometric and Topographic sets were also deliberately made equivalent in scale, but this was not the case with the Climatic set, which was controlled by the spatial scales within the principal components analysis.

The Countries set is different, covering all spatial scales. It includes a large number of very small regions[231], so it has as many as 130 regions.[232] Nonetheless, the region in the Countries set with the largest surface area (Russia) is comparable in size to the largest in the Giorgi set (North Asia).

However, there are also differences in spatial scale within the two groups. The range and variability of size are much less within the Large Geometric set than in the Giorgi set.[233] Outside Antarctica the Climatic set has slightly larger regions than the rest of the small-scale group, and there is more variability in size among the regions.[234]

6.3b Antarctica

---

[231] There are fifty regions containing fewer than four grid-boxes.

[232] Although 130 is a large number of regions, it is, of course, far less than the number of countries in the world. Thus many countries are unrepresented. This region set is not intended to provide a climate scenario for every country, but to allow us to investigate the effects on climate scenarios of zoning using political boundaries.

[233] For the range, see Table 6.2 ('Rest of World'). Despite the Large Geometric and Giorgi sets having the same number of non-Antarctic regions, the standard deviations of region size (in boxes) for the Large Geometric set is only a fifth (11.70) of that for the Giorgi set (54.13).

[234] The standard deviations of region size (in boxes) for the non-Antarctic regions are as follows: Small Geometric = 2.46, Topographic = 9.03, Climatic =14.00.

There are a number of different ways of treating Antarctica. Antarctica is singled out because of its representation on the HadCM2 grid: although it accounts for less than 9% of the surface area of the land grid-boxes, it accounts for nearly 30% of the grid-boxes (Table 6.2 'Summary').[235] In the Countries and Giorgi sets it is a single region: it is the largest in terms of grid-boxes, but not in terms of surface area. The other sets (Table 6.2 'Antarctica') divide it into between 4 and 25 regions, depending on whether the balance struck with the Rest of the World is in terms of grid-boxes or surface area.

6.3c Large-scale group

A visual inspection of the zonal configuration in the Giorgi set reveals how influenced it has been by climatological and topographic considerations; consider, for example, the division between the Mediterranean basin and Northern Europe, or its identification of the monsoonal areas of India and the Gulf of Guinea (Figure 6.3 bottom left). The zoning of the Large Geometric set pays no regard to such considerations:[236] Northern Africa and Europe are separated, Europe is treated as a whole, and 'East Africa' extends as far north as the Caspian Sea (Figure 6.4 top left). However, despite the different principles employed, in some areas the configuration of land and sea forces the zonal configurations of both sets into similar patterns (*e.g.* Figure 6.5, right).

6.3d Small-scale group

The Small Geometric set (Figure 6.4, top right) pays the same disregard to climatic considerations as the Large Geometric set, although the effects are not as great since the regions are much smaller. In contrast, the Topographic set (Figure 6.4, bottom left), which is equivalent in size to the Small Geometric set, picks out many physiographic features that influence climate processes. If we take South America as an example, we find that unlike the Small Geometric set (Figure 6.5, bottom left), the Topographic set (Figure 6.1, bottom left) identifies the Andes, Brazilian Highlands, Amazon basin, and Pampas. However, the constraints of spatial scale meant that the attempt to identify each region with a single topographic feature was not entirely successful. Examples include the following (see Figure 6.1, bottom left):

---

[235] This is because the HadCM2 grid does not vary the number of longitudinal grid-boxes with changing latitude, so the area represented by a grid-box varies with latitude.

[236] Some of the boundaries look non-geometric on the global plot (Figure 6.4 top left); they are in fact straight lines on the HadCM2 grid, but appear as curves on this equal-area plot. For examples, compare South America on the global (Figure 6.4 top left) and local (Figure 6.5 bottom right) plots.

- islands are treated as part of a mainland region (*e.g.* Caribbean islands);
- some regions are a composite of topographic features (*e.g.* the inclusion of the southern-most Andes in the Patagonian region);
- some regions only represent part of a single topographic feature (*e.g.* the Siberian Plain in Figure 6.4, bottom left).

The Climatic set (Figure 6.4, bottom right) does not identify topographic features nearly as well as the Topographic set. For example, in South America the zonal configuration of the Climatic set (Figure 6.2, bottom right) is more similar to the Small Geometric set (Figure 6.5, bottom left) than it is to the Topographic set (Figure 6.1, bottom left). However, the Climatic set does identify areas of coherent variability in the control surface temperatures (Figure 6.2). The regions are mostly spatially contiguous[237] and compact[238] (Figure 6.4, bottom right).

In summary, varying combinations of region sets are similar in either their spatial scale or else in their principles of zonation, but are different in other respects. A range of spatial scales are included among them, and principles of zonation that incorporate varying amounts of climatically-relevant information. Therefore these region sets provide a suitable basis on which to evaluate the effects of spatial scale and zonation on climate scenario construction.

6.4 Information loss

Our aim in this chapter is to compare the relative worth of different principles by which we might select regions. There are a number of different bases on which we might make such a comparison. In this section we employ a measure of the information that is lost through aggregating grid-boxes together.

6.4a Measure of information loss

Information is lost at the individual grid-box level when we aggregate grid-box values into a regional mean. The 'information' to which we refer is not the response to increased radiative forcing, but the difference between the grid-box value and the global-mean – *i.e.* the spatial pattern. The extent to which information is lost by regionalising provides a measure of the worth of

[237] There are some exceptions, notably in Europe, where the Balkans are separated from the rest of Southern Europe, and where Britain is separated from Northern Europe.

[238] Again, there are some exceptions, notably Japan, central America, and the Namib Desert.

regionalising. Therefore we compare the various region sets using the loss of information incurred in their construction.

Our measure of the loss of *temperature* information is a RMSE,[239] in which the 'error' is the difference between a grid-box anomaly and the regional anomaly to which that grid-box contributes (Equation 6.1.1). This measure is not suitable for precipitation because of precipitation's heteroscedasticity, so our measure of the loss of *precipitation* information is the coefficient of variation among the grid-box values within a region (Equation 6.1.2). Our standard for comparison is the loss of information in the final 30-year period (the 2080s) of the HadCM2 Ga ensemble mean in boreal summer (JJA).

In Figures 6.6 (temperature) and 6.7 (precipitation) we plot the loss of information in each region. We plot the loss of information against the size of the region because of the relationship between the loss of information and the number of grid-boxes in a region. We plot separately the region sets zoned using climatic (top right)[240] and non-climatic (top left)[241] information; thus we may determine whether the differences between the region sets are due merely to different spatial scales, or also to the principles employed in zonation. We also plot the Climatic (Figure 6.6) and Topographic (Figure 6.7) sets separately (bottom left) because it is for these sets that the zonation is most likely – in our judgement – to influence the amount of loss of information. We give a global summary in Figure 6.8.[242]

6.4b Logarithmic relationship with region size

There is a logarithmic relationship between a region's size and the loss of information incurred in its construction (Figure 6.8):

- for a region containing a single grid-box there is no loss of information by regionalising,
- as grid-boxes are added the spread of values within the region rapidly increases;

[239] 'RMSE' denotes root-mean squared error.

[240] The climate-related group is comprised of the Giorgi, Topographic, and Climatic sets.

[241] The non-climatic group is comprised of the Countries and Geometric sets.

[242] In Figure 6.8 we plot the average loss of information (weighted by region size) for each region set against the mean region size. We make the comparison for temperature (top) and precipitation (bottom). Since they are regionalised differently, we make separate comparisons for Antarctica (left) and the Rest of the World (right).

- eventually the region becomes so large that adding extra grid-boxes has little effect on the loss of information from an average grid-box within the region.

Therefore we plot logarithmic best-fit lines in Figures 6.6 and 6.7 and compare them (bottom right).

A visual inspection of Figures 6.6 and 6.7 suggests that the logarithmic best-fit lines capture the general relationship between region size and the loss of information reasonably well, despite the spread of values between similarly-sized regions within a single region set. It is difficult to distinguish the best-fit lines for the group of climate-related sets (brown) from particular members of that group (red, green, bottom right). This is not surprising; since the Climatic and Topographic sets each contribute nearly half the grid points to the climate-related group, the differences among the sets in the climate-related group would have to be substantial to generate different best-fit lines.

6.4c Influence of climatic information

If the zonation employs climate-related information so as – in some way – to identify relatively homogenous regions, we might expect the loss of information in regionalising to be less than for a region of equivalent size that was zoned without using any climatic information. Such a contrast should be identifiable from a visual inspection of the non-climatic (blue) and climate-related (brown) best-fit lines (Figures 6.6 and 6.7). However, although there are differences between these lines it is uncertain whether they are due to sampling or to differences in zoning.

We might also expect to identify a contrast between non-climatic and climate-related sets from their global-means (Figure 6.8). However, in the Antarctic (left) the Climatic and Topographic sets certainly do not improve on the Geometric sets. In the Rest of the World there is no clear improvement for temperature (top right),[243] although there is possibly a slight improvement for precipitation (bottom right).[244]

Therefore we examine the scatter plots more closely (Figures 6.6 and 6.7, top), still searching for a contrast between the climate-related and non-climatic sets. We find

[243] The Topographic and Giorgi sets offer slight improvements on the Small and Large Geometric sets respectively, but there is no clear improvement for the Climatic set, which we might expect *a priori* to offer the greatest improvement.

[244] Each of the climate-related sets (Topographic, Climatic, Giorgi) improves on the non-climatic sets, but not by large amounts.

that in the lowest part of the range of region sizes differences do appear.[245] It is possible that the differences are due to sampling, since relatively few regions are involved, but this seems unlikely because similar differences are present for both temperature and precipitation.

6.4d Quantifying the influence of climatic information

To quantify our comparison we employed permutation testing.[246] To carry out the test we constructed two random samples from the Climatic and Countries sets. Seven regions containing between 8 and 13 grid-boxes (the 'Climatic sub-set') were available from the Climatic set. We selected two equivalent sets of seven from the Countries set (the 'Countries sub-sets'):

- the size of each selected region corresponded to the size of a member of the Climatic sub-set;
- where there was a choice of regions for selection, the selection was made randomly.

For both temperature and precipitation, the loss of information in the Climatic sub-set was half that in the Countries sub-sets.[247] We conducted a permutation test to establish the statistical significance of these differences.[248] The probability that the loss of information at this spatial scale (8-13 grid-boxes) in the Climatic set was less than that lost in the Countries set was very high: 0.984 for temperature, and 0.992 for precipitation.

We repeated this procedure for regions with 20-22 grid-boxes, using a sub-set from the Climatic set and two equivalent sub-sets from the Small Geometric set. However, for this spatial scale the contrast in lost information between the

---

[245] Compare the scatters close to the spatial scale of 10 boxes in the blue and brown plots.

[246] See section 3.5 for our discussion of permutation testing.

[247] The mean temperature RMSE (°C) in the Climatic sub-set was 0.38, and in the two Countries sub-sets it was 0.58 and 0.71. The mean precipitation coefficient of variation in the Climatic sub-set was 0.58, and in the two Countries sub-sets it was 1.02 and 1.28.

[248] We tested the null hypothesis that equal information is lost in the Climatic and Countries sets at the spatial scale of 8-13 grid-boxes. Two random samples were available from the Climatic (n=7) and Countries (n=14) sets. We carried out a pooled permutation test on sums to establish the likelihood of the Climatic sample being a random permutation from the pooled sample. We calculated 100,000 permutations to establish the reference distribution.

Climatic and Small Geometric sub-sets was not nearly so great.[249] The probabilities that less information was lost in the Climatic set were only 0.422 for temperature and 0.067 for precipitation.

6.4e Implications of the quantified comparison

It appears that there is a contrast between spatial scales of approximately 10 and approximately 20 (or greater) grid-boxes. At the smallest scales represented here (8-13 grid-boxes) the climate-related group loses less information than the non-climatic group, but at larger spatial scales there is no identifiable difference between the groups. This contrast explains why the climate-related best-fit lines (brown, red, green) have steeper gradients than the non-climatic best-fit lines (blue).

It is conceivable that the differences between the non-climatic and climate-related groups at the smaller scale might be explained by the Climatic regions concerned all belonging to a particular part of the world, but in fact this is not the case.[250] It appears that the contrast between the smaller and larger scales is genuine, and that the reduced loss of information (relative to the non-climatic group) at the smaller scale is due to the principles of zonation employed. We conclude that:

- Climatic information only 'improves' the zonation of regions when the spatial scale is small.
- Contiguous areas in HadCM2 with homogenous natural variability are on a relatively small spatial scale of less than 15 grid-boxes.

6.5 Signal-to-noise ratios

In section 6.4 we examined the loss of information arising from the aggregation of grid-boxes, regardless of any climate change. However, since our investigation is partly motivated by the possibility of aggregating grid-boxes for pattern scaling purposes, it is appropriate to employ a measure related to climate change to compare the relative worth of different methods of region selection.

---

[249] The mean temperature RMSE (°C) in the Climatic sub-set was 0.57, and in the two Small Geometric sub-sets it was 0.41 and 0.67. The mean precipitation coefficient of variation in the Climatic sub-set was 0.86, and in the two Small Geometric sub-sets it was 0.45 and 0.67.

[250] The seven-member Climatic sample is made up of regions representing: Japan, the Cherskiy Range, and Rub al Khali (all in Asia); the Namib Desert (Africa), the Mexican Plateau (North America), Patagonia (South America), and Northern Queen Maud (Antarctica). Not only do these regions represent a wide range of continents; they also represent a wide range of latitudes and topographic features.

6.5a Globally-averaged ratios

The measure we choose is the signal-to-noise ratio described in the latter part of section 4.4 (Equations 4.3), which we re-express in Equations 6.2.[251] We calculate our signal-to-noise ratios for the 2080s in JJA from the HadCM2 Ga ensemble mean and 1400-year control. We calculate for both temperature and precipitation, and we calculate for each of the regions in each of the region sets. We plot the signal-to-noise results by calculating the mean ratio (weighted by size) and mean region size for each region set (Figure 6.9).

There is not the simple logarithmic relationship between signal-to-noise ratio and region size (Figure 6.9)[252] that there is between lost information and region size (Figure 6.8). The relationship between signal-to-noise ratio appears to be approximately linear, and the *zonation* appears to be as important to the signal-to-noise ratio as the spatial *scale*. For example, in two of the four plots in Figure 6.9 (top right, bottom left) the Small Geometric set gives signal-to-noise ratios that are at least the equal of the Large Geometric set.

6.5b Influence of climatic information

We compare the signal-to-noise ratios in the region sets zoned with and without climate-related information (Figure 6.9). There is no difference between the temperature signal-to-noise ratios in the Antarctic (top left), and in the Rest of the World the climate-related sets have smaller ratios than those zoned without climatic information (top right). For precipitation it is possible that the climate-related information does improve the signal-to-noise ratios: in the Antarctic the Topographic set gives better signal-to-noise ratios than the Geometric sets (bottom left), and in the Rest of the World this is the case for both the Topographic and Giorgi sets.

However, there is another plausible explanation for the different precipitation signal-to-noise ratios. Although particular sets have equal mean region sizes, these sets may have different distributions of region sizes, which may affect the mean

---

[251] It is worth noting that the signal-to-noise ratio is different from the measure of information loss described in the previous section. The 'information' referred to there concerned the distinctiveness of the spatial pattern. Here the 'signal' is the response to increased radiative forcing. In the former case the 'information' became less as the spatial scale increased; in the latter case the signal-to-noise ratio may increase as the spatial scale increases.

[252] Except, perhaps, for Antarctic temperatures (Figure 6.9, top left).

signal-to-noise ratio. Upon investigation we found that the distribution of region sizes in the Giorgi set had a much larger variance than the Large Geometric set,[253] and that the signal-to-noise ratios of the five largest regions in the Giorgi set[254] contributed 52.8% of the weighted mean of the set, despite only contributing 44.8% of the grid boxes. On the basis of this evidence it seems plausible that the larger mean signal-to-noise ratio in the Giorgi set is due to the different region sizes in the Giorgi set, rather than to the use of climate-related information.

The obstacle to using a permutation test[255] in these circumstances is posed by the very different region sizes within the Giorgi and Large Geometric sets, which makes it very difficult to draw appropriate equivalent samples from those sets. We attempted to overcome this obstacle through removing the influence of size (x) on a region's signal-to-noise ratio (y), by calculating a best-fit line ($y=ax+b$) for the pooled sets. We removed this best-fit line from the regional signal-to-noise ratios, randomly subdivided the two sets into two halves, and conducted a permutation test on each half.[256] We found no evidence that the Giorgi set has larger signal-to-noise ratios than the Large Geometric set, once the influence of varying region sizes had been removed.

When we compare the distributions of region sizes in the Topographic and Small Geometric sets we find that although they have the same mean, the Topographic set has a much higher variance, and is positively skewed. However, there is a considerable correspondence in region sizes between these two sets. So we randomly selected samples of regions with equivalent sizes in the 20-22 box and 23-26 box ranges from the Topographic and Small Geometric sets. The permutation tests suggested that in the 20-22 box range the Topographic set

---

[253] Although the Large Geometric and Giorgi sets have identical mean region sizes, their standard deviations of region size are 11.7 boxes and 54.1 boxes respectively.

[254] An example of the contrast in variances is that the Large Geometric and Giorgi sets have one (121 boxes) and five (93, 95, 114, 154, 275 boxes) regions larger than 90 boxes respectively.

[255] A permutation test could be used to determine whether one sample (the Giorgi set) might have been a random selection from the pooled sample (the Giorgi and Large Geometric sets).

[256] We tested the null hypothesis that the Giorgi and Large Geometric sets have equivalent signal-to-noise ratios, once the influence of region size had been removed by detrending the ratios with respect to region size. There was sufficient data for us to divide the samples (n=21) into two and conduct two tests (n=10, n=11). We carried out two pooled permutation tests on sums to establish the likelihood of the Giorgi sample being a random permutation from the pooled sample. We calculated 100,000 permutations to establish the reference distribution. The p-values in the two tests were 0.5683 and 0.5501.

definitely gives larger signal-to-noise ratios than the Small Geometric set, but that in the 23-26 box range the difference is insignificant.[257] The inconclusive results prompt us to repeat the detrending of the signal-to-noise ratios with respect to size for the Topographic and Small Geometric sets. Having conducted six permutation tests,[258] we concluded that there was no evidence that the Topographic set has larger signal-to-noise ratios than the Small Geometric set, once the influence of varying region sizes had been removed.

6.6 Pattern scaling

Thus far we have examined the loss of information and changes in signal-to-noise ratio that follow from regionalising. We have found that although the choice of spatial scale is very influential, zonation does not appear to be influential.

In this section we examine the accuracy of pattern scaling for different spatial scales and zonations. We apply the pattern scaling technique evaluated in chapter 4 to some of the region sets described earlier in this chapter. Since we have found the spatial scale to be particularly important, we select the Land Boxes, Small Geometric, and Large Geometric sets. In order to compare zoning methods we also include the Climatic and Topographic sets for temperature and precipitation respectively. The standard for comparison is the estimate made for each region set of the mean JJA temperature and precipitation anomalies in the 2080s under the Gd scenario. A worked example is given for Small Geometric temperatures in Figure 6.10.

---

[257] We tested the null hypothesis that the Topographic and Small Geometric sets have equivalent signal-to-noise ratios. We carried out tests for region sizes of 20-22 boxes (n=9), and 23-26 boxes (n=11). We carried out two pooled permutation tests on sums to establish the likelihood of the Topographic sample being a random permutation from the pooled sample. We calculated 100,000 permutations to establish the reference distribution. The p-values in the two tests were 0.9899 and 0.6745 respectively.

[258] We tested the null hypothesis that the Topographic and Small Geometric sets have equivalent signal-to-noise ratios, once the influence of region size had been removed by detrending the ratios with respect to region size. There were sufficient data to sub-divide the samples and conduct six tests (n=10 for each). We carried out six pooled permutation tests on sums to establish the likelihood of the Topographic sample being a random permutation from the pooled sample. We calculated 100,000 permutations to establish the reference distribution. The p-values in the six tests were 0.2472, 0.4365, 0.4516, 0.6994, 0.7977, 0.8352.

We obtain a response pattern from the Ga ensemble for each region set by calculating the mean response for each region from its constituent grid-boxes.[259] We estimate the mean JJA temperature and precipitation anomalies in the 2080s under the Gd scenario for each region set by multiplying the response pattern by the scaler in the 2080s under the Gd scenario (*e.g.* Figure 6.10, top right). We express the error in the estimate (bottom left) relative to the modelled values (top left), and calculate its significance (bottom right).[260]

We use the error significance to compare the accuracy of the estimates. We plot this statistic for the four selected region sets for temperature (Figure 6.11)[261] and precipitation (Figure 6.12). These temperature and precipitation plots are summarised in Figures 6.13 and 6.14 respectively, where each land grid-box (excluding Antarctica)[262] is classed by the magnitude of its error significance.

In all cases the proportion of grid-boxes where the error represents at least two standard deviations is relatively small, approximately 10-20%. These grid-boxes are found in similar areas in different region sets.[263] An important measure is the proportion of grid-boxes where there is no evidence that the error is significant, which we define as the class where the magnitude of the error is less than one standard deviation. The spatial scale clearly influences this measure. For both temperature and precipitation an increase in spatial scale from Land Boxes to Small Geometric to Large Geometric decreases the proportion of grid-boxes where the error magnitude is less than a standard deviation: for temperature the decrease from Land Boxes to Large Geometric is approximately 10%, for precipitation it is 25%. At first glance this might be thought a little surprising, because the error might be expected to decrease as the spatial scale increases. In fact, the error does indeed decrease as the spatial scale increases, but the internal variability decreases even further.[264]

---

[259] The response for each grid-box under the Ga scenario was calculated in section 4.5 and plotted in Figure 4.10. It is expressed in °C or mm/day per °C of change in the scaler.

[260] The calculation of the significance follows Equations 4.5; we estimated σ_M from the control.

[261] In Figure 6.11, the top right plot is a repeat of Figure 6.10 (bottom right).

[262] Antarctica is excluded to ease the comparison between the Small Geometric and climatically-zoned (Climatic, Topographic) sets.

[263] For example, most of the grid-boxes where the error in the temperature estimate represents at least two standard deviations are in West Africa, the Middle East, India, eastern Russia, and western Canada (Figure 6.11).

[264] The mean error among the 2271 land grid-boxes is 0.0047 °C for Land Boxes, and 0.0043 °C for Large Geometric. The standard deviation by which the error is divided to obtain the significance of

When we consider the influence of zonation in a similar way to the influence of spatial scale, we find no evidence of any improvement in the accuracy of the pattern scaling. The proportion of grid-boxes in the climatically-zoned sets where the error magnitude is less than a standard deviation is no greater than in the Small Geometric set: for precipitation the Topographic set has a similar proportion (Figure 6.14), and for temperature the Climatic set actually has a smaller proportion (Figure 6.13).

6.7 Conclusions

In this chapter we have addressed the possibility of aggregating grid-boxes into regions for the purposes of pattern scaling. In particular, we have examined the influence of the choices we make of spatial scale and zonation. We constructed seven region sets on the HadCM2 grid using different spatial scales and methods of zonation (section 6.2). In reviewing the sets in section 6.3 we identified a group[265] in which only the spatial scale varies, and two groups[266] in which the spatial scale is similar but the zoning varies.

When we compared the loss of information[267] in different region sets (section 6.4) we found a logarithmic relationship between the loss of information and the spatial scale. The method of zonation had a comparatively minor influence, although there is evidence that zoning climatically gives a small reduction in the loss of information at smaller spatial scales of less than 15 grid-boxes. In section 6.5 we found an approximately linear relationship between the spatial scale and the signal-to-noise ratio that represents the response to increased radiative forcing. In this case the variations in signal-to-noise ratios between sets with equivalent spatial scales suggest that the signal-to-noise ratio is nearly as dependent on the zonation as on the spatial scale. However, there is little or no evidence that using specifically *climatic* information in the zonation leads to an increase in the signal-to-noise ratio.

---

the error (see Equations 4.5) is 0.1644 °C for Land Boxes, and 0.1046 °C for Large Geometric. Therefore although the error decreases with increasing spatial scale, it increases relative to internal variability.

[265] The group is formed by the Land Boxes, Countries, Small and Large Geometric sets.

[266] One group is formed by the Small Geometric, Topographic, and Climatic sets; the other group is formed by the Large Geometric and Giorgi sets.

[267] *i.e.* the loss of the spatial pattern

Against this background we carried out pattern scaling for some of the region sets (section 6.6). We found that although the estimation errors decreased as the spatial scale increased, the background of internal variability decreased even further. Therefore the proportion of grid-boxes with significant errors increased slightly with increasing spatial scale. We did not find any evidence that the method of zonation improves the accuracy.

The first major conclusion is that there is little evidence that we can improve climate scenarios by defining regions using climatically-relevant information. The region set employed by the IPCC is made up of rectangles that loosely represent a range of different climates. However, our conclusion suggests that the use of climatic information does not improve the results; similar results would be obtained if similarly-sized regions were zoned using geometric principles. If our conclusion is correct then the only conceivable way by which climatically-informed zoning might increase signal-to-noise ratios and the accuracy of pattern scaling is to construct the regions on the basis of the response pattern itself. There is, of course, a strong element of circularity in using the HadCM2 response pattern to define regions, employing them in pattern scaling, and then pronouncing them better at approximating the response to forcing than any other region set!

The second major conclusion is that there is no optimal spatial scale for pattern scaling, in the sense that any change in spatial scale involves a trade-off. Increasing the spatial scale decreases the information available from the spatial pattern, but also increases the signal-to-noise ratio that represents the response to increased radiative forcing. Increasing the spatial scale decreases the magnitudes of the errors introduced by pattern scaling, but the errors become more statistically significant. However, this conclusion ought to be qualified by the greater agreement between models and with observations that we might expect to find at larger spatial scales.

The answer to the question that we posed as the title to this chapter ('May we regionalise?') is 'yes'. Aggregating grid-boxes into regions improves the signal-to-noise ratio and decreases the errors introduced by pattern scaling. However, the individual grid-box remains a suitable subject for pattern scaling from one forcing scenario to another; naturally the signal-to-noise ratios are smaller and the errors from pattern scaling are larger, but a large amount of useful information is still retained. Whether pattern scaling for the individual grid-box is a worthwhile activity ultimately depends on the view one takes of the intrinsic value of model information at the grid-box level. From the limited perspective of the pattern

scaling technique the grid-box and region may both be legitimately used in climate scenarios.

Equations 6.1

1. The information lost (expressed as a RMSE in °C) by a region (r, with n_g grid-boxes) through regionalising the grid-box temperatures (°C, t_g) into a regional mean (t_r).

$$RMSE = \sqrt{\frac{1}{n_g}\sum_{g=1}^{n_g}(t_g - t_r)^2}$$

2. The information lost (expressed as the coefficient of variation) by a region (r, with n_g grid-boxes) through regionalising the grid-box precipitation (mm/day, p_g) into a regional mean (p_r).

$$p_r = \frac{1}{n_g}\sum_{g=1}^{n_g} p_g$$

$$c.v. = \frac{1}{p_r}\sqrt{\left(\frac{1}{n_g}\sum_{g=1}^{n_g} p_g^2\right) - p_r^2}$$

Equations 6.2

1. The signal (s_r) in °C in a region (r, with n_g grid-boxes) for the Ga ensemble mean in a 30-year period (j). It was constructed by anomalising the mean grid-box temperature in that period (°C, t_g) against the mean from the first 240 years of the control (°C, t_b). An equivalent precipitation signal is constructed similarly.

$$s_{rj} = \frac{1}{n_g} \sum_{g=1}^{n_g} \left(t_g - t_b \right)$$

2. The noise (σ_r) in °C in the same region, using a set of periods[268] ($c=1 \ldots n_c$) in the 1400-year control. The period-mean for the region (t_r) is the mean of the region's constituent grid-boxes, and the anomalies are calculated with reference to the first 240 years of the control (b). The noise is the estimate of the population standard deviation from the sample of t_r (size n_c):

$$a_{rc} = t_{rc} - t_{rb}$$

$$\sigma_r = \sqrt{\frac{n_c}{n_c - 1} \left(\frac{1}{n_c} \sum_{c=1}^{n_c} a_{rc}^2 - \left(\frac{1}{n_c} \sum_{c=1}^{n_c} a_{rc} \right)^2 \right)}$$

$$\frac{s_r \sqrt{4}}{\sigma_r}$$

3. The signal-to-noise ratio for a region (r). The multiplication by two is to allow for the signal being constructed from a four-member ensemble.

[268] We detrend the 1400-year control by calculating the best-fit line ($y=ax+b$) through least squares regression, and removing b. We use the mean from the first 240 years of the control as the reference (b) against which to anomalise. The remaining 1160 years of the control are divided into 40-year sequential segments, and the periods ($c=1 \ldots n_c$) are the first 30 years of each segment.

Table 6.1

Criteria upon which regions were selected for the seven region sets.

The zoning constraints employed when constructing the various region sets were as follows. The columns refer to constraints derived from published criteria (left), model characteristics, geometric criteria, and contiguity of shape (right). A blank entry in a column indicates that in that respect no constraint was set.

zoning constraints	published Globe	model Globe	geometric Globe	contiguity Globe
Land Boxes				
Countries	yes			yes
Giorgi	yes		yes	yes
Large Geometric			yes	yes
Small Geometric			yes	yes
Topographic		elevations		yes
Climatic		climate		

The spatial scale constraints were as follows, given separately for Antarctica (Ant) and the Rest of the World (RoW). In the right-hand column we state whether equality between region sizes in Antarctica and the Rest of the World was achieved in terms of the number of boxes in a region, or in the spatial area it represents. In the other three columns we note the constraints that we followed concerning the total number of regions in the set, the average number of boxes in a region, and the minimum and maximum number of boxes permitted in a region. A blank entry in a column indicates that in that respect no constraint was set.

size	total regions		mean boxes		min/max box		RoW-Ant eq	
constraints	RoW	Ant	RoW	Ant	RoW	Ant	boxes	area
Land Boxes	1631	640	1	1			yes	
Countries		1		640				
Giorgi		1		640				yes
Large Geometric	21		78					yes
Small Geometric	66	25	25	26	20,30	20,30	yes	
Topographic	66		25		15,35			yes
Climatic								

Table 6.2

Summary characteristics for the seven region sets.

The full set of land grid-boxes (Global Land) is subdivided between Antarctica and the Rest of the World. 'Area' (in millions of km^2) refers to model surface area.

Summary	boxes total	boxes %	area m km2	area %
Global Land	2271	100.00	147.69	100.00
Antarctica	640	28.18	13.20	8.94
Rest of World	1631	71.82	134.49	91.06

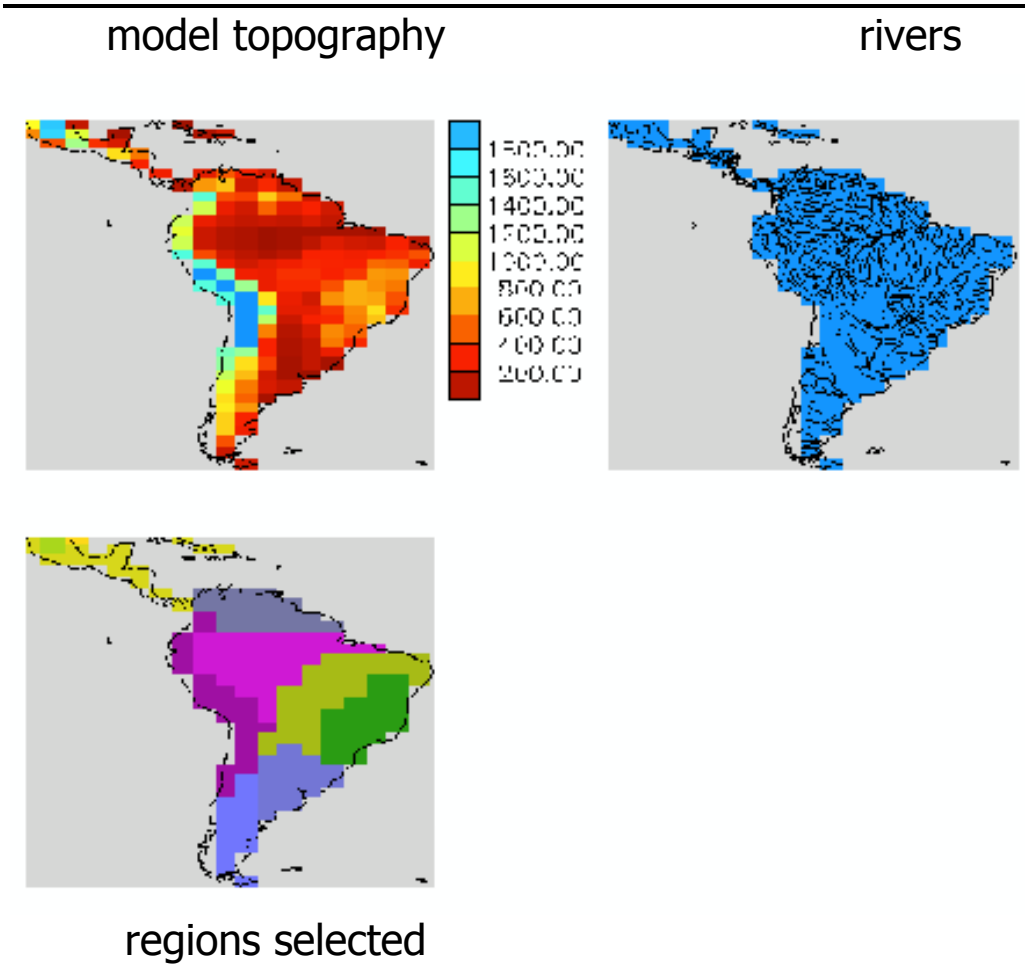
Global Land	regions	mean boxes	min. boxes	max. boxes	mean m km2	min. m km2	max. m km2
Land Boxes	2271	1	1	1	0.07	0.01	0.12
Countries	130	17	1	640	1.14	0.06	16.66
Giorgi	22	103	24	640	6.71	2.58	14.65
Large Geometric	25	91	62	177	5.91	2.77	9.43
Small Geometric	91	25	20	30	1.62	0.15	3.29
Topographic	72	32	15	156	2.05	0.76	4.04
Climatic	63	36	8	184	2.34	0.40	6.05

Antarctica	regions	mean boxes	min. boxes	max. boxes	mean m km2	min. m km2	max. m km2
Land Boxes	640	1	1	1	0.02	0.01	0.03
Countries	1	640	640	640	13.20	13.20	13.20
Giorgi	1	640	640	640	13.20	13.20	13.20
Large Geometric	4	160	124	177	3.30	2.77	3.96
Small Geometric	25	26	20	30	0.53	0.15	1.05
Topographic	6	107	35	156	2.20	1.23	3.10
Climatic	9	71	11	184	1.47	0.40	3.36

Rest of World	regions	mean boxes	min. boxes	max. boxes	mean m km2	min. m km2	max. m km2
Land Boxes	1631	1	1	1	0.08	0.03	0.12
Countries	129	13	1	305	1.04	0.06	16.66
Giorgi	21	78	24	275	6.40	2.58	14.65
Large Geometric	21	78	62	121	6.40	3.63	9.43
Small Geometric	66	25	20	30	2.04	0.64	3.29
Topographic	66	25	15	72	2.04	0.76	4.04
Climatic	54	30	8	63	2.49	0.71	6.05

Figure 6.1

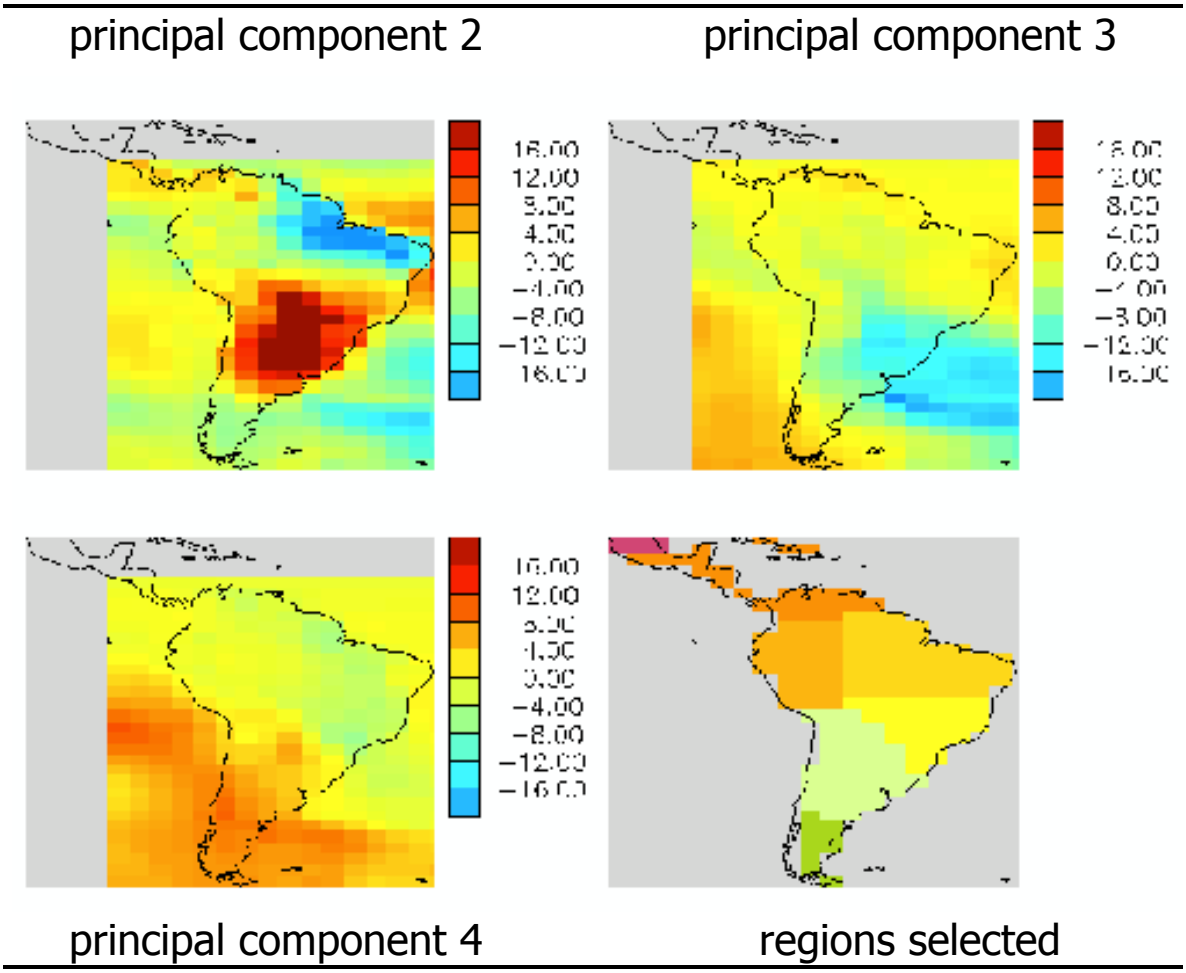
The selection of regions for the Topographic set for South America.



model grid:	HadCM2
region set:	Topographic
topography:	metres above sea level for each HadCM2 grid-box (**top left**)
rivers:	major rivers (**top right**)
regions:	regions selected (**bottom left**)

Figure 6.2

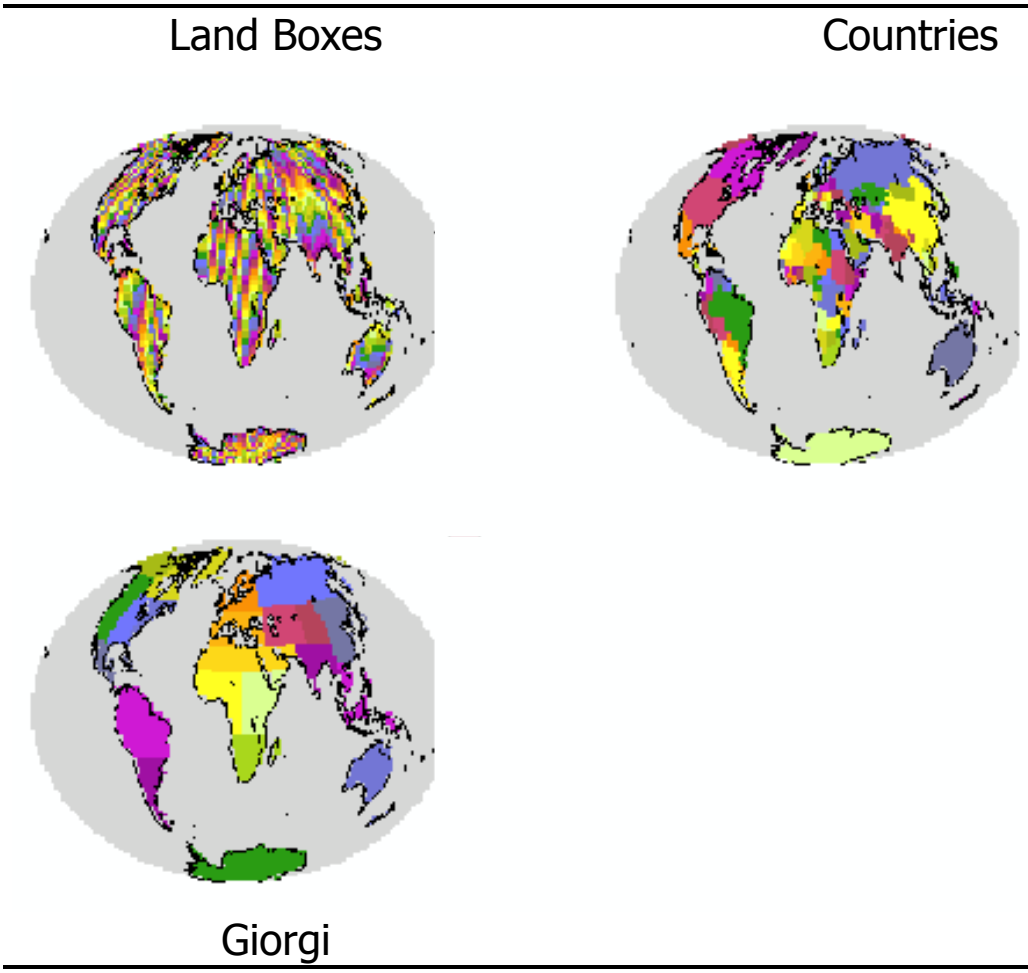
The selection of regions for the Climatic set for South America.



model grid:	HadCM2
region set:	Climatic
data:	annual mean temperatures (°C) from 1400y control
processing:	smoothed with 30y Gaussian filter , and detrended
PC analysis:	PC analysis for the rectangle of grid-boxes displayed
projections:	of the first 4 PCs, we plot numbers 2-4 (**top**, **bottom left**)
	(PC 1 is a land-sea contrast and played no role in selection)
var. expl.:	percentages of the variance explained:
	PC 1: 48.7%, PC 2: 9.9%, PC 3: 7.7%, PC 4: 5.7%
regions:	regions selected (**bottom right**)

Figure 6.3

The regions selected for the Land Boxes, Countries, and Giorgi sets.

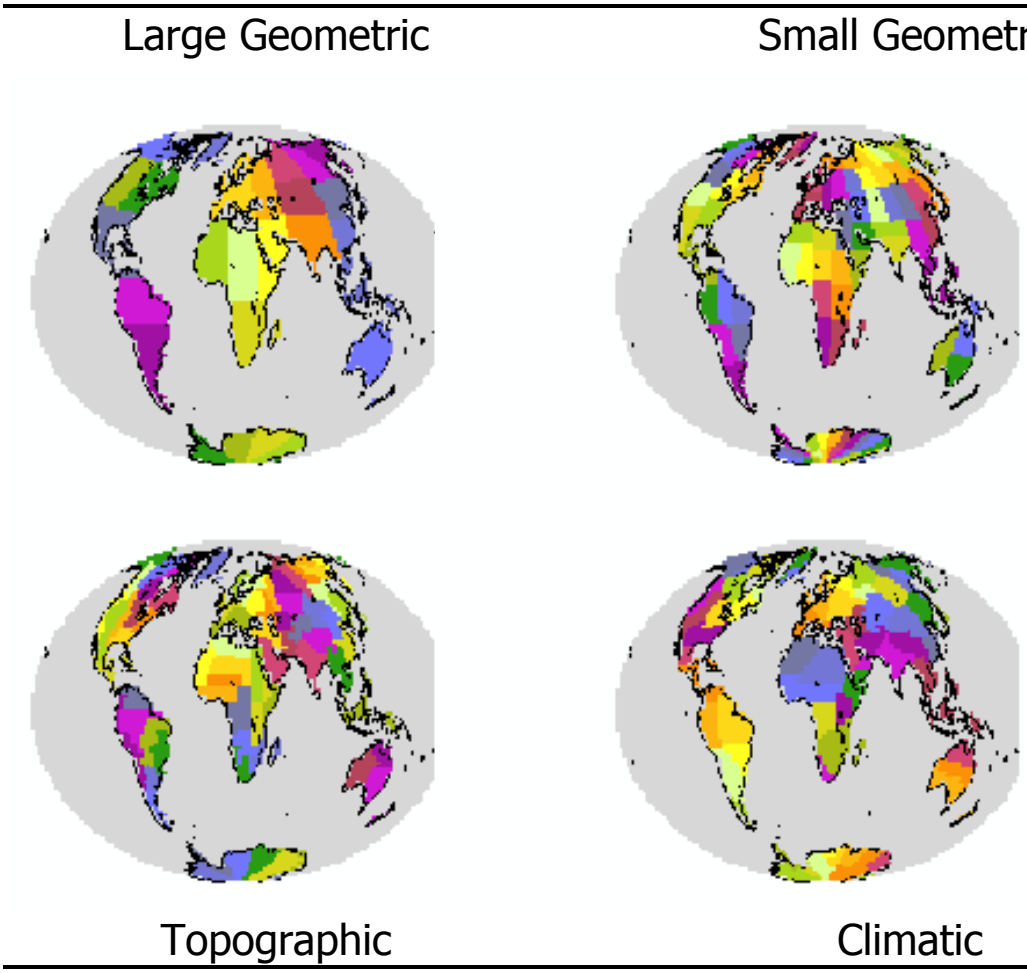


model grid: HadCM2

region sets: Land Boxes (**top left**), Countries (**top right**), Giorgi (**bottom left**)

Figure 6.4

The regions selected for the Geometric, Topographic, and Climatic sets.

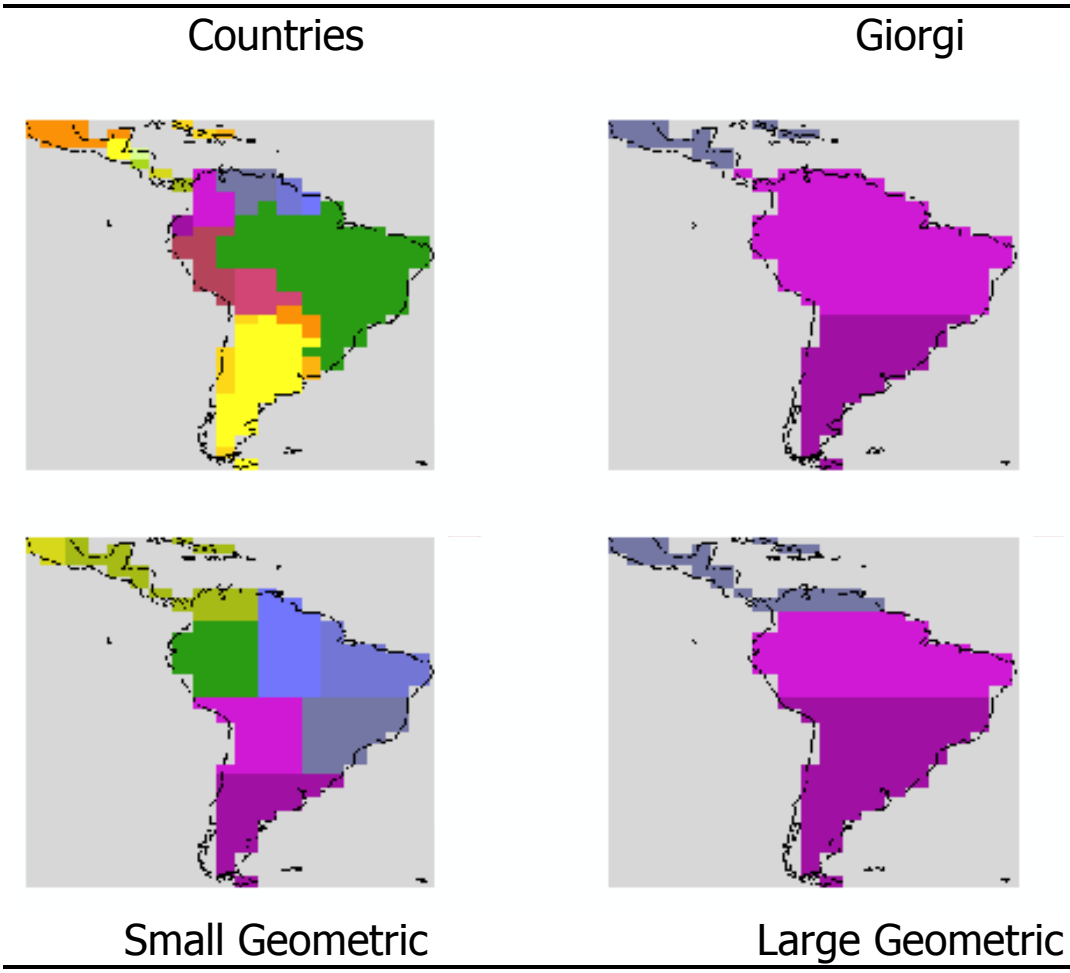


model grid: HadCM2

region sets: Large Geometric (**top left**), Small Geometric (**top right**) Topographic (**bottom left**), and Climatic (**bottom right**)

Figure 6.5

The regions selected for the Countries, Giorgi, and Geometric sets.

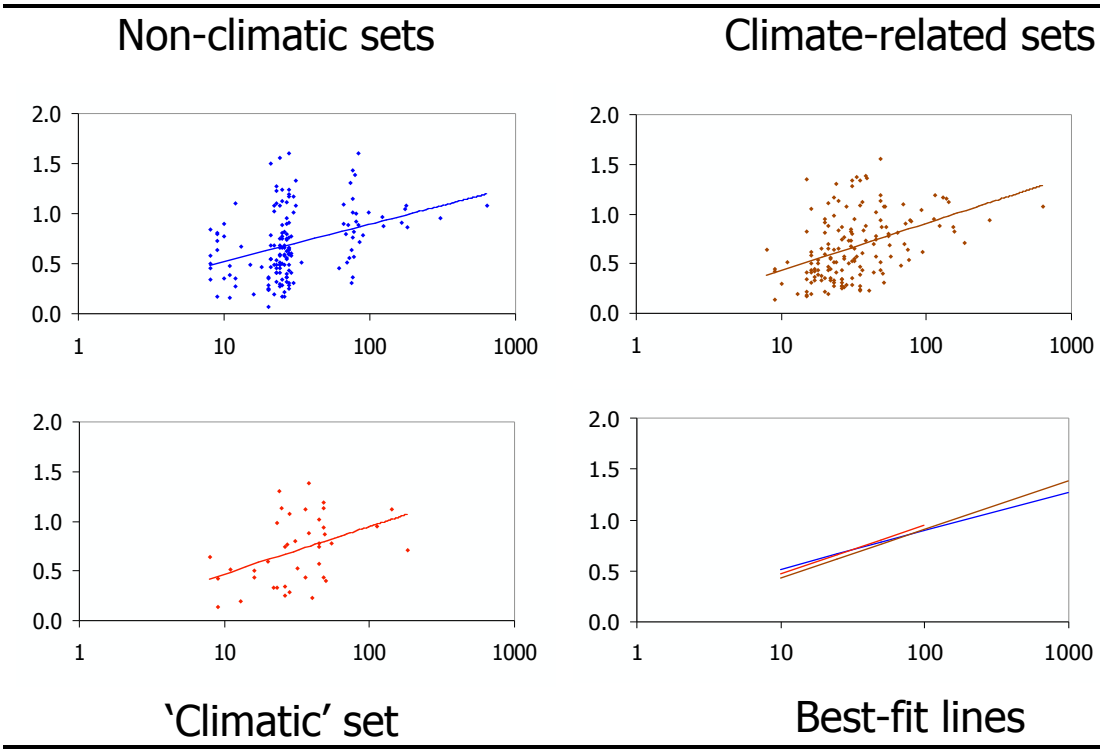


model grid: HadCM2

region sets: Countries (**top left**), Giorgi (**top right**),
Small Geometric (**bottom left**), Large Geometric (**bottom right**)

Figure 6.6

The temperature information lost through regionalising.



model grid: HadCM2

region sets: Countries, Small Geometric, Large Geometric (**top left**)
(restricted to regions with at least 8 grid-boxes)
Giorgi, Climatic, Topographic (**top right**)
Climatic (**bottom left**)

variable: Ga ensemble JJA temperature (°C) mean for 2070-2099

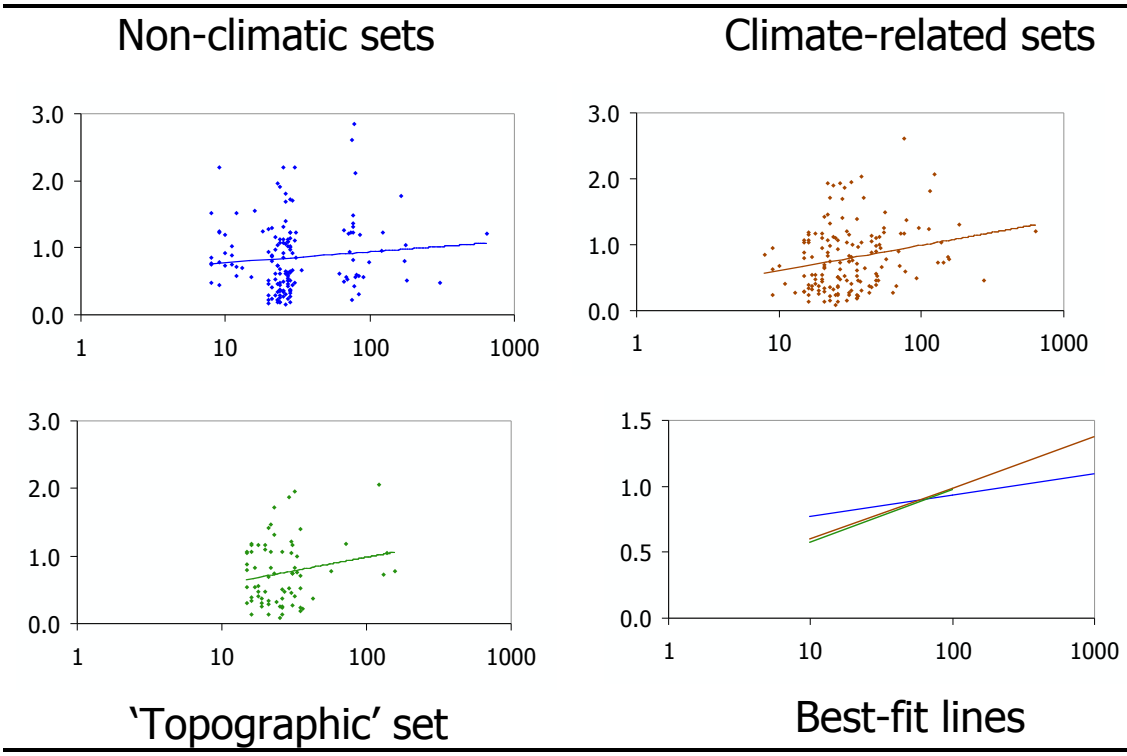
y-axis: RMSE for set of grid-boxes in a region (Equation 6.1.1)
'error' is regional value minus grid-box value

x-axis: number of grid-boxes in the region

best-fit: logarithmic best-fit lines overplotted for each diagram and compared (**bottom right**)

Figure 6.7

The precipitation information lost through regionalising.



model grid: HadCM2

region sets: Countries, Small Geometric, Large Geometric (**top left**)
(restricted to regions with at least 8 grid-boxes)
Giorgi, Climatic, Topographic (**top right**)
Topographic (**bottom left**)

variable: Ga ensemble JJA precipitation (mm/day) mean for 2070-2099

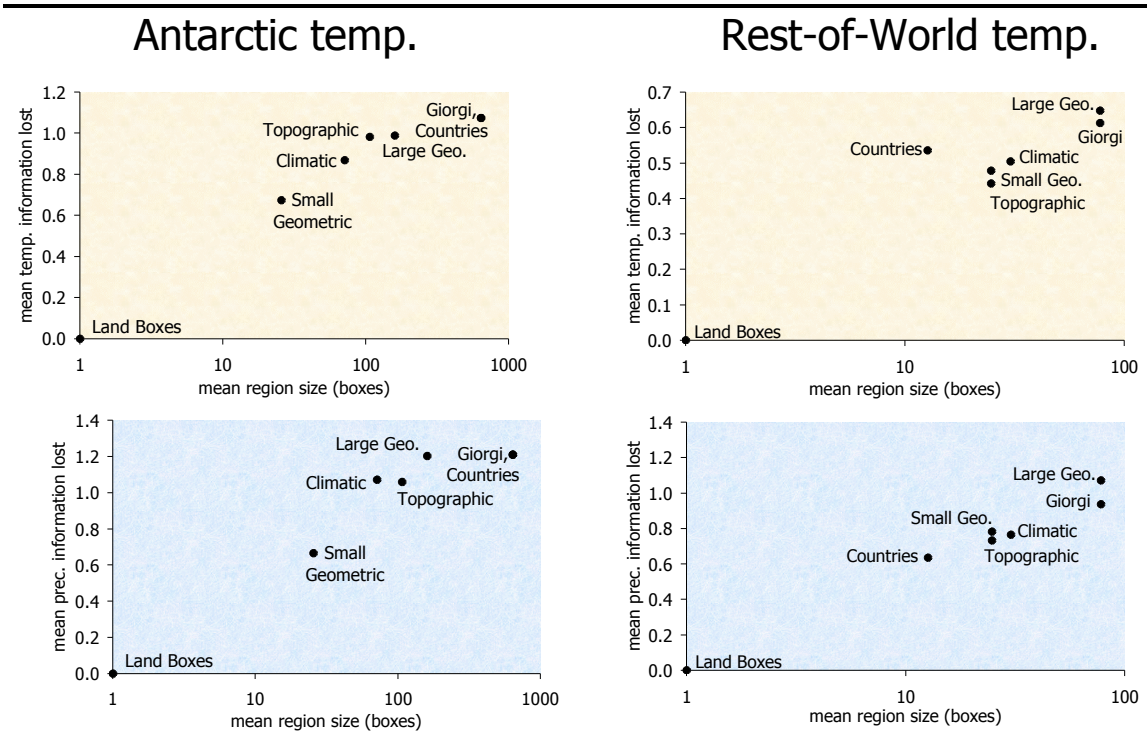
y-axis: coefficient of variation for set of grid-boxes in a region

x-axis: number of grid-boxes in the region (Equation 6.1.2)

best-fit: logarithmic best-fit lines overplotted for each diagram and compared (**bottom right**)

Figure 6.8

The temperature and precipitation information lost through regionalising.



model grid: HadCM2

region sets: Countries, Small Geometric, Large Geometric, Giorgi, Climatic, Topographic, Land Boxes
divided into Antarctica (**left**) and Rest of World (**right**)

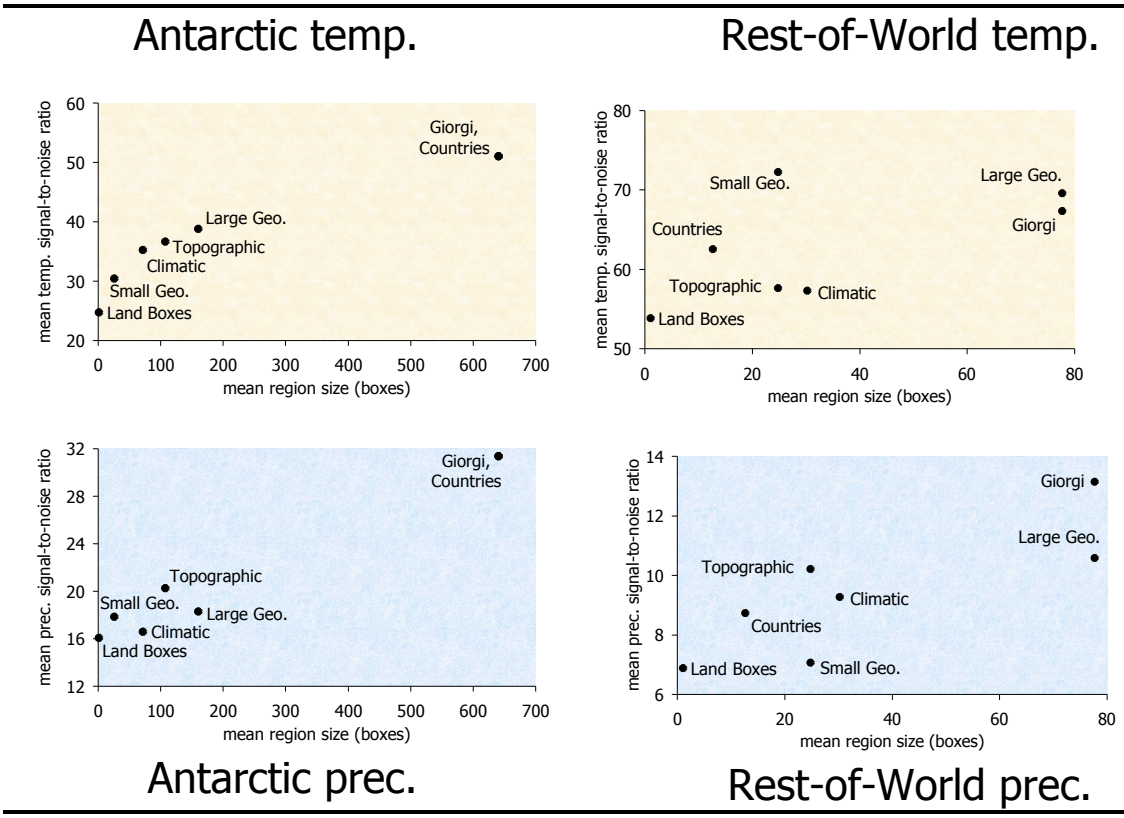
variables: JJA regional temp. information lost (°C, as Figure 6.6) (**top**)
JJA regional prec. information lost (as Figure 6.7) (**bottom**)

y-axis: weighted mean for each set of information lost
(weighted by the number of grid-boxes in each region)

x-axis: average number of grid-boxes per region in each set

Figure 6.9

Temperature and precipitation signal-to-noise ratios.

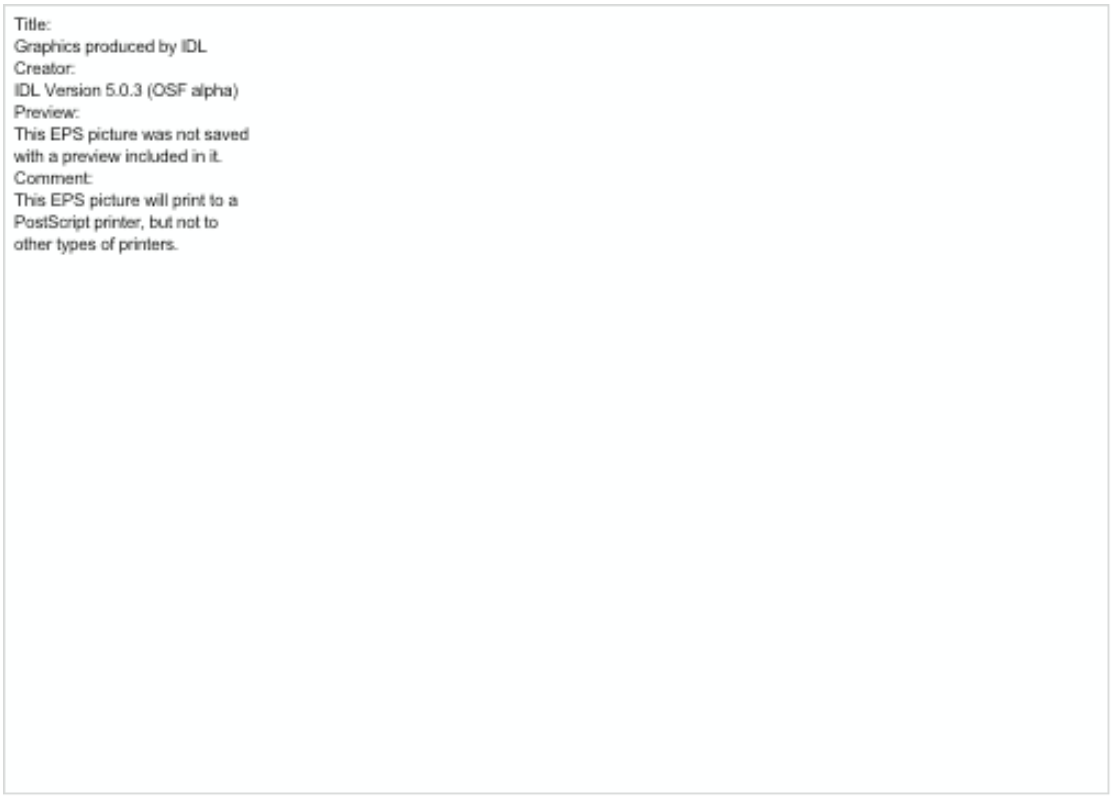


model grid:	HadCM2
region sets:	Countries, Small Geometric, Large Geometric, Giorgi, Climatic, Topographic, Land Boxes divided into Antarctica (**left**) and Rest of World (**right**)
variables:	JJA temperature (**top**) and precip (**bottom**) signal-to-noise ratios
signal:	Ga ensemble mean anomaly (*v* control 1-240) in 2080s (2070-99)
noise:	population standard deviation of 30 periods (10y gaps) in detrended 1400y control anomaly means (*v* control 1-240)
ratio:	absolute value of signal, multiplied by two (due to ensemble), and divided by noise
y-axis:	weighted mean for each set of signal-to-noise ratios (weighted by the number of grid-boxes in each region)
x-axis:	average number of grid-boxes per region in each set

Figure 6.10

The estimate of Gd temperatures in the 2080s for the Small Geometric set.

modelled in Gd estimated with Ga pattern



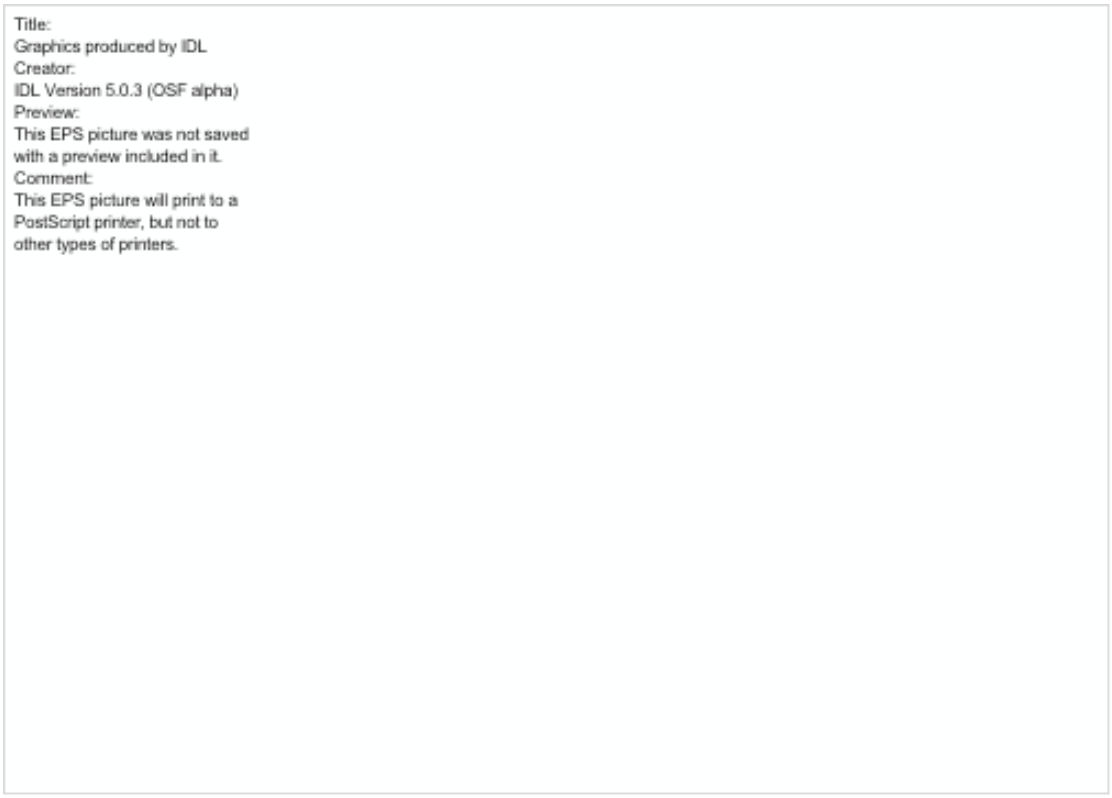
error significance

model grid:	HadCM2
region sets:	Small Geometric
modelled:	Gd 2080s JJA temperature anomalies (°C, **top left**)
Ga pattern:	for each region, the mean of the temperature responses under Ga from its constituent grid-boxes (section 4.5, Figure 4.10)
estimate:	Ga pattern * scaler in Gd 2080s (°C, **top right**)
error:	estimate – modelled (°C, **bottom left**)
error sig:	error / pop. standard deviation (□, Equations 4.5)
classed sig:	error sig. classed with bounds: -2, -1, 1, 2 (□, **bottom right**)

Figure 6.11

Accuracy of temperature estimates made using pattern scaling.

Land Boxes Small Geometric



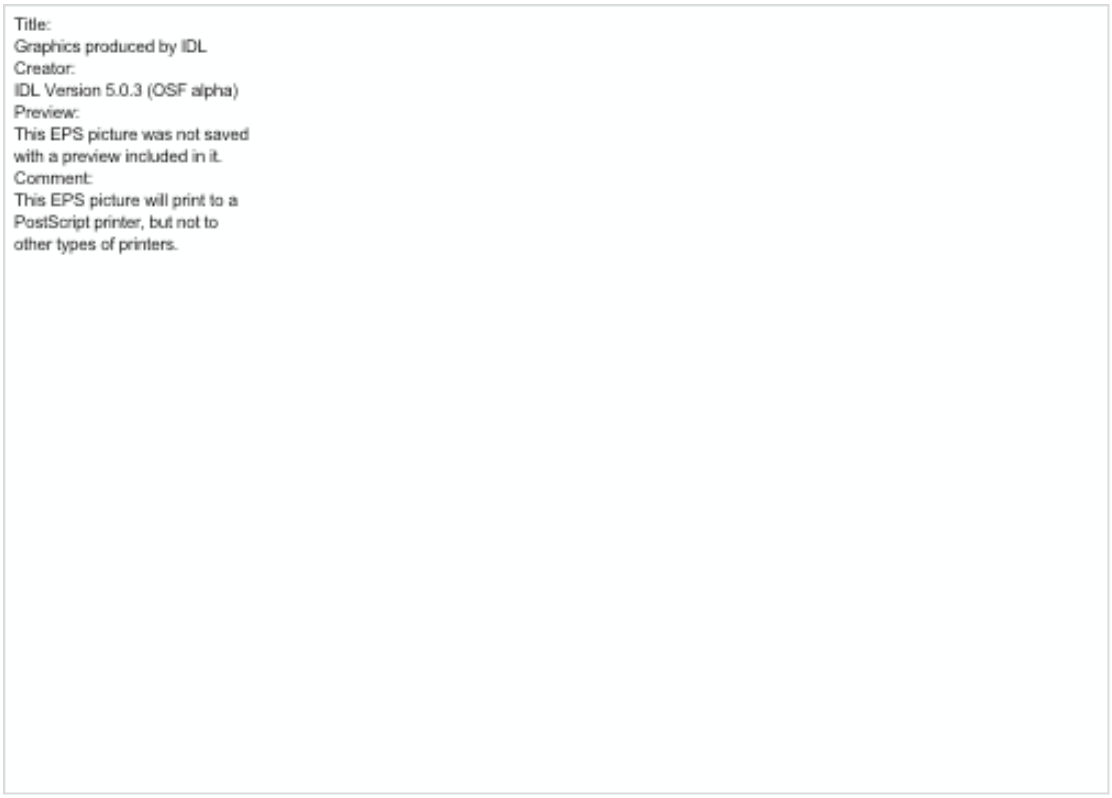
Large Geometric Climatic

model grid:	HadCM2
region sets:	Land Boxes, Small Geometric, Large Geometric, Climatic
modelled:	Gd 2080s JJA temperature anomalies (°C)
Ga pattern:	for each region, the mean of the temperature responses under Ga from its constituent grid-boxes (section 4.5, Figure 4.10)
estimate:	Ga pattern * scaler in Gd 2080s (°C)
error:	estimate – modelled (°C)
error sig:	error / pop. standard deviation (□, Equations 4.5)
classed sig:	error sig. classed with bounds: -2, -1, 1, 2 (□, **plotted**)

Figure 6.12

Accuracy of precipitation estimates made using pattern scaling.

Land Boxes | Small Geometric

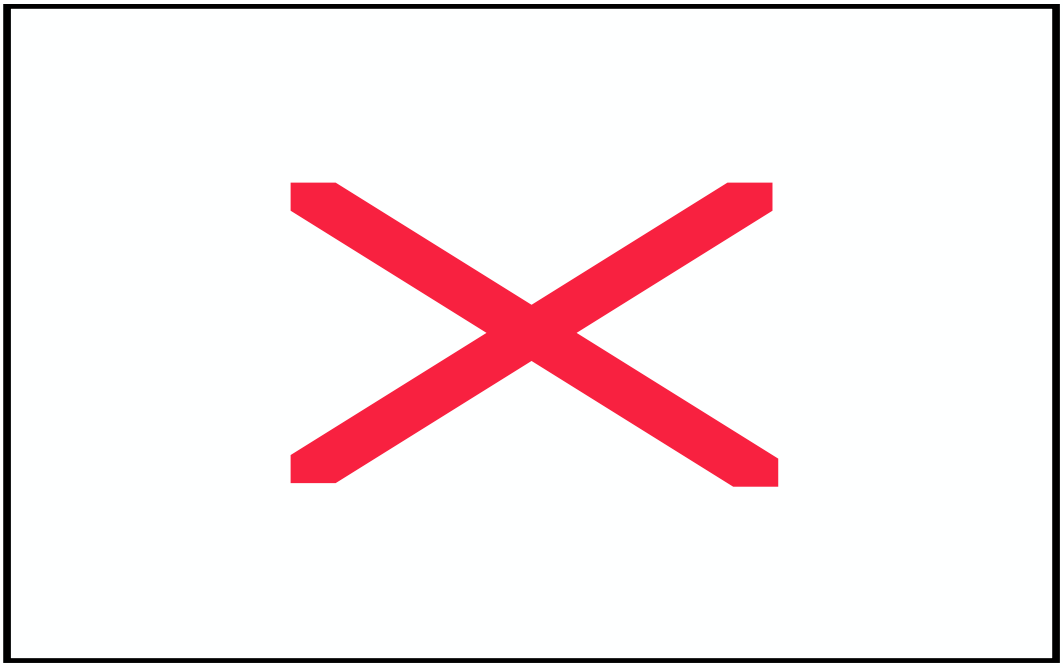


Large Geometric | Topographic

model grid:	HadCM2
modelled:	Gd 2080s JJA precipitation anomalies (mm/day)
Ga pattern:	for each region, the mean of the temperature responses under Ga from its constituent grid-boxes (section 4.5, Figure 4.10)
estimate:	Ga pattern * scaler in Gd 2080s (mm/day)
error:	estimate – modelled (mm/day)
error sig:	error / pop. standard deviation (□, Equations 4.5)
classed sig:	error sig. classed with bounds: -2, -1, 1, 2 (□, **plotted**)

Figure 6.13

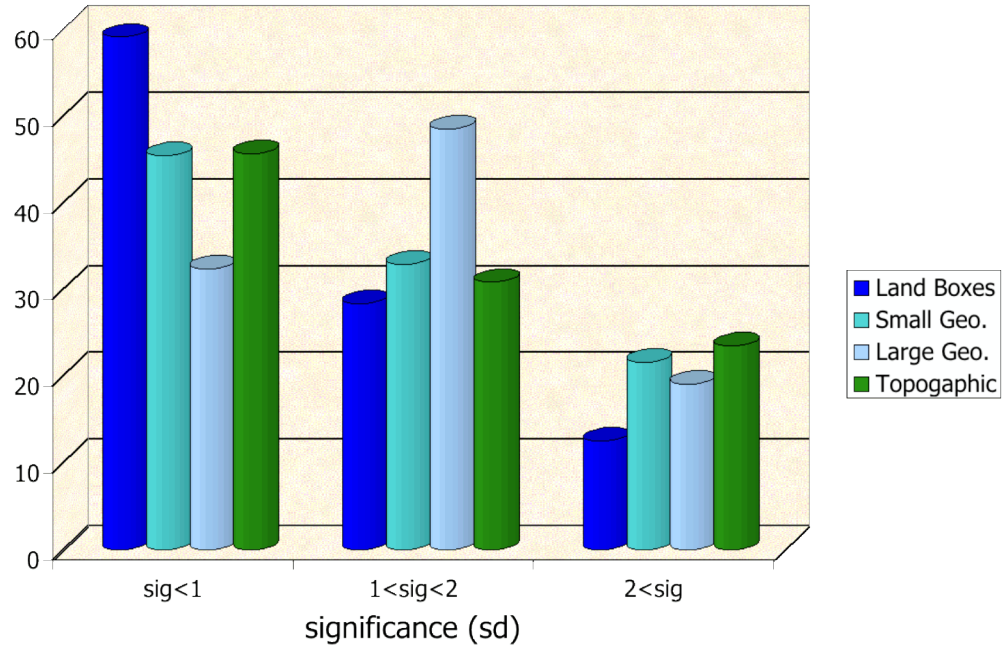
Accuracy of temperature estimates made using pattern scaling.



model grid:	HadCM2
region sets:	Land Boxes, Small Geometric, Large Geometric, Climatic
modelled:	Gd 2080s JJA temperature anomalies (°C)
Ga pattern:	for each region, the mean of the temperature responses under Ga from its constituent grid-boxes (section 4.5, Figure 4.10)
estimate:	Ga pattern * scaler in Gd 2080s (°C)
error:	estimate – modelled (°C)
error sig:	error / pop. standard deviation (□, Equations 4.5)
classed sig:	magnitudes of error sig. classed with bounds: 1, 2 (□)
percentages:	non-Antarctic grid-boxes in each class (%, **plotted**)

Figure 6.14

Accuracy of precipitation estimates made using pattern scaling.



model grid:	HadCM2
region sets:	Land Boxes, Small Geometric, Large Geometric, Topographic
modelled:	Gd 2080s JJA precipitation anomalies (mm/day)
Ga pattern:	for each region, the mean of the temperature responses under Ga from its constituent grid-boxes (section 4.5, Figure 4.10)
estimate:	Ga pattern * scaler in Gd 2080s (mm/day)
error:	estimate – modelled (mm/day)
error sig:	error / pop. standard deviation (, Equations 4.5)
classed sig:	magnitudes of error significance grouped with bounds: 1, 2 ( )
percentages:	non-Antarctic grid-boxes in each class (%, **plotted**)

7. May we scale different models?

7.1 Introduction

Thus far we have examined pattern scaling using a single model – HadCM2. However, if we are to conclude that the technique is generally applicable, we must demonstrate that it is applicable to a wider range of GCMs. Moreover, a full assessment of future climate change requires the use of multiple GCMs.[269] In this chapter we apply the pattern scaling technique developed in chapter 4 to five more models, and consider results not just for individual grid-boxes, but also for two of the region sets developed in chapter 6.

We begin by carefully selecting the data we will use for our inter-model comparison, so as to make the comparison as fair as possible (section 7.2). We establish in section 7.3 that the differences between the data used here and the data used in previous chapters do not invalidate any comparison with previous chapters. In section 7.4 we construct response patterns both for a sample of GCMs, and for an ensemble of simulations from HadCM2. This combination allows us to evaluate any inter-model differences in the light of the internal variability of one of the models. We also calculate response patterns particularly for the 'Giorgi' region set from chapter 6, so as to enable a comparison to be made with the regional assessments of the IPCC (section 7.5). In section 7.6 we examine the accuracy with which the response patterns are able to represent climate anomalies in the 2080s in the simulations from which they were drawn. Again, the availability of the HadCM2 ensemble enables us to place the errors from pattern scaling into a context of internal variability. Once again we apply these methods particularly to the 'Giorgi' region set (section 7.7). Finally we draw our conclusions (section 7.8).

7.2 Making a fair comparison

The IPCC established the Data Distribution Centre (DDC) to facilitate the use of GCM output in climate impact analyses. Thus recent results from a number of coupled GCMs have become available for comparison. If we are to carry out a fair comparison of the *spatial patterns* of climate change exhibited by the models, we

[269] We discussed the need for multiple GCMs in section 2.11.

need the model data on a common grid.[270] The necessary work to achieve this has already been carried out by the DDC, which has transformed much of the model output from its native grid to a common grid by a Gaussian interpolation: the grid measures 2.5° latitude by 3.75° longitude. Therefore we use this regridded model output, together with the HadCM2 and HadCM3 model output already on that grid, as the raw material that we analyse in this chapter. In this section we further narrow down the raw material that will enable us to make a fair comparison of the models.

7.2a Nine models

To make a fair comparison of the models, we must employ simulations with a common magnitude and pattern of radiative forcing. We rule out the use of simulations forced with a mixture of greenhouse gases and sulphates because of the variety of sulphate forcings employed by the different modelling centres. At present there are no regridded control simulations available to us, so the use of control simulations is also ruled out. However, there are nine models for which there are simulations on the common grid forced solely with greenhouse gases under the 'GG' scenario (Figure 7.1).[271]

Under the GG scenario the radiative forcing in the model is that observed until (about) 1990, and thereafter the concentration of greenhouse gases is increased at the rate of about 1% per annum. The radiative forcing in this set of simulations is not only reasonably homogenous, it is also reasonably strong, and thus provides a good opportunity to identify any divergence of behaviour between different models. Therefore we use the GG simulations for our model inter-comparison.

Inter-model differences in architecture may be assessed not just in terms of the model structure, but also by measures of the model climate. In Figure 7.2 we include both kinds of measure for the nine models with GG simulations on the common grid: we plot the horizontal resolution of the model atmosphere on the axes, and we plot the model's equilibrium climate sensitivity as the circle

---

[270] It is common practice to regrid model data to facilitate inter-comparison (*e.g.* Lambert and Boer, 2000). It is particularly important here because of our aggregation of grid-boxes into region sets.

[271] We recognise that this greenhouse-gas-only scenario neglects some of the radiative forcings that have contributed to changes in the 20th century. However, our purpose here is merely to consider the applicability of pattern scaling to more than one model, rather than to develop definitive climate scenarios for the 21st century. Moreover, in the latest emissions scenarios (IPCC, 2000) the greenhouse gases dominate all other radiative forcings in the 21st century.

diameter. Although it may not be necessary to compare all nine models, the models we select should span the range displayed here.

Our pattern scaling method employs anomalies. In previous chapters we have anomalised relative to the first 240 years of the long HadCM2 control simulation, but our choices here are restricted by the limited data available for the different models. If we anomalised relative to a control simulation, we would be restricted to two models. If we anomalised relative to a 30-year period from a 19th century 'pre-industrial' climate we would be restricted to four of the nine models (Figure 7.1). Therefore we anomalise relative to a recent 30-year period (1960-1989) that is included in all nine models (Figure 7.1).[272]

7.2b Six models

Our pattern scaling method employs a response pattern calculated from a time series of overlapping 30-year means. In order to provide a sufficiently strong signal for the response pattern this time series must be reasonably lengthy, and span a reasonably wide range of values of the scaler (global annual temperature). In our judgement, the 110-year period from 1990-2099 satisfies these requirements; extending the period further back into the 20th century will not greatly extend the range of scaler values, and shortening it will make the response pattern more dependent upon inter-decadal variability.

In three of the nine models the GG simulation does not extend as far as 2099 (Figure 7.1), so the time series requirements must be balanced against the need to span the range displayed in Figure 7.2. However, if we do insist on the time series extending from 1990 to 2099, the three models 'lost' (ECHam3, GFDL-R15, and NCAR-DOE) do not prevent us from almost spanning the ranges of horizontal resolution and climate sensitivity.[273] Therefore we restrict our inter-model comparison to the six models (CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, and HadCM2) where the GG simulation extends to 2099. There is a single GG simulation available for each of the first five models, but for HadCM2 there is an

[272] We examine the effects on the response patterns of using this recent base period from which to calculate anomalies in section 7.3.

[273] The range of climate sensitivities for all nine models is 2.5-4.6 °C, and the number of grid-boxes at the surface ranges from 1920 to 8192. Without the three models 'lost' the sensitivity range is 2.5-4.3 °C, and the grid-box range is 2048-8192.

ensemble of four simulations.[274] Therefore we will be able to compare both inter- and intra-model variations.

It is beyond the scope of this thesis to enter into a detailed examination of the six models.[275] Here we merely summarise their features in Tables 7.1 (spatial specification) and 7.2 (dynamical specification). Table 7.2 also gives references for the models and their GG simulations. This summary is sufficient to indicate the variety of implementations among the six selected models.

7.3 Periods for pattern construction

We have already identified the data that we will use to construct response patterns, namely the 110-year period 1990-2099, anomalised against the period 1960-1989. This ('chapter 7') data is different to previous chapters (particularly chapter 4), where we have used the 240-year period 1860-2099, anomalised against the first 240 years of the control simulation. Therefore, if the results from this chapter are to be directly comparable with the preceding chapters, we must establish what effect the different data have upon the response patterns.

We begin by replotting the 'chapter 4' response patterns (from section 4.5) for the HadCM2 Ga ensemble: the temperature and precipitation patterns from Figure 4.10 are replotted in the top left of Figures 7.3 and 7.4 respectively. We then recalculate the response patterns using the 'chapter 7' data:[276] the temperature and precipitation patterns are in the top right of Figures 7.3 and 7.4 respectively. The differences between the 'chapter 7' and 'chapter 4' patterns are plotted in the bottom left. To provide an indication of the practical importance of these differences we replot from Figure 4.11 the difference between the temperature response patterns for the 2020s and 2080s (Figure 7.3, bottom right), and add its precipitation equivalent (Figure 7.4, bottom right).

---

[274] In previous chapters this ensemble has been referred to as Ga 1,2,3,4.

[275] Detailed comparisons of coupled models are the subject of the Coupled Model Inter-Comparison Project (CMIP). This project is described briefly by Meehl *et al.* (1997) and key tables are provided by Phillips (1998). Lambert and Boer (2000) have published a report.

[276] We calculate grid-box JJA anomalies for the period 1990-2099 against the base period mean (1960-1989), and average them into 30-year period means. Then we calculate the best-fit line for each grid-box ($y=ax$) between the regional anomaly (y) and global annual temperature (x) by least squares regression.

The differences between the 'chapter 7' and 'chapter 4' patterns are small, and difficult to identify from a visual inspection of the patterns (Figures 7.3 and 7.4, top), although they are apparent when plotted (bottom left). These differences are actually *smaller* than the differences between instantaneous patterns from the 2020s and 2080s (bottom right).[277] This indicates that any differences in the anomalies estimated using the 'chapter 7' pattern rather than the 'chapter 4' pattern are likely to be smaller than those arising from the slight non-linearities of the response to increased radiative forcing in the 21st century. Therefore we are justified in using the 'chapter 7' data, and in drawing comparisons with previous chapters.

7.4 Patterns for different models

7.4a Two samples

To make a formal comparison of model response patterns we construct two samples of response patterns. For each individual in a sample we calculate response patterns for temperature and precipitation using the method of section 7.3. The 'HadCM2 sample' is drawn from the individual members of the HadCM2 Ga ensemble, which gives us four possible climatic outcomes from a single representation of the climate system (*i.e.* the HadCM2 model).[278] The 'DDC sample' is drawn from the GG simulations from five models and the Ga 1 pattern from HadCM2, which gives us six different representations of the climate system.[279] The DDC sample gives equal weight to each model, which would not be the case if we included within it all four HadCM2 members. We include the first member of the HadCM2 ensemble rather than the ensemble mean because using the ensemble mean would confuse the contrast between the HadCM2 and DDC samples.

---

[277] The average magnitude among the 6816 grid-boxes in the difference plots in Figure 7.3 is 0.059 °C (bottom left) and 0.100 °C (bottom right). In Figure 7.4 the average magnitude is 0.040 mm/day (bottom left) and 0.067 mm/day (bottom right).

[278] In this discussion we could have referred to the 'HadCM2 sample' as the 'HadCM2 ensemble'. We chose the former term in order to be consistent with the 'DDC sample' with which we compare the 'HadCM2 sample', and in order to avoid ambiguity, since the preceding chapters have considered *two* ensembles from HadCM2.

[279] See Tables 7.1 and 7.2 and Figure 7.2 for examples of differences between models in their representations of the climate system.

In the analysis below we summarise the properties of the HadCM2 and DDC samples using different statistics, because of the different natures of the two samples. We assume that the HadCM2 sample of patterns forms a random sample from the population of all possible response patterns from this model, and that the population has a Gaussian distribution. Therefore we may estimate the population parameters of the distribution from the HadCM2 sample. However, we cannot assume that the DDC sample is a random sample from the population of all plausible models,[280] nor that the population of all plausible models forms a Gaussian distribution. Therefore we treat the DDC sample differently from the HadCM2 sample. We neither draw inferences about the population of all plausible models, nor calculate the sample's mean and standard deviation. The sample median and range are more appropriate measures of the central tendency and spread.

We present the temperature patterns (odd-numbered figures) and precipitation patterns (even-numbered figures) based on individual grid-boxes[281] from the single simulations available for CCSR-NIES, CGCM1, CSIRO2, ECHam4 (Figures 7.5 and 7.6), and HadCM3 (Figures 7.7 and 7.8, bottom right). These five patterns, together with the HadCM2 Ga 1 pattern,[282] make up the DDC sample. We present the HadCM2 sample in Figures 7.7 and 7.8. Rather than plotting the individual patterns from the HadCM2 sample (which are barely distinguishable from each other), we plot our estimate of the population mean (bottom left), together with the lower (top left) and upper (top right) bounds of the 95% confidence interval for our estimate of the population mean.

We present summary statistics for the patterns from the DDC sample in Figures 7.9 – 7.10 (for individual grid-boxes) and 7.11 – 7.12 (for the Countries set). We plot the sample median (top left) and range (bottom left). We also plot the difference between the limits of the HadCM2 95% confidence interval (bottom right) described above and plotted in Figures 7.7 and 7.8. Finally we compare the DDC range and the HadCM2 95% confidence interval (top right).

7.4b Comparison of models

[280] We discussed the reasons for regarding the DDC sample as a non-random sample in section 2.11.

[281] We calculated similar patterns and sample statistics for the Countries set (not shown).

[282] The HadCM2 Ga 1 pattern is not shown, but the HadCM2 Ga ensemble mean pattern (which is similar) is shown in Figures 7.7 and 7.8 (bottom left).

The models agree on the large-scale features of the response to radiative forcing. In their inter-model comparison, Kattenberg *et al.* (1996) considered a sample of sixteen models (their Table 6.3) including three of the DDC sample.[283] Their conclusions regarding temperature (section 6.2.2.2) are similar to those that may be identified from a visual inspection of the DDC sample (Figures 7.5, 7.7):[284]

- "a greater warming over land than over the sea" (p301);
- "a minimum warming or even regions of cooling in the high latitude southern ocean" (p302);
- "a region of minimum warming in the northern North Atlantic" (p304).

The conclusions of Kattenberg *et al.* (1996) regarding precipitation (6.2.2.3) may also be identified from the DDC sample (Figures 7.6, 7.8):

- "an increase in global-mean precipitation" (p307);[285]
- "increases in high latitudes in winter" (p307);[286]
- "more rainfall over India and/or south-east Asia" (p309);[287]
- "changes in the dry subtropics are small" (p309).

When the same authors turned to regional scales they concluded from the inter-model range of responses and the size of the model biases (with respect to observations) that "confidence in the regional scenarios simulated by AOGCMs remains low" (p339). A visual inspection of the patterns from the DDC sample (Figures 7.5-7.8) seems to confirm that the inter-model range of responses at regional scales remains large.

7.4c Contribution of internal variability

Differences between the models have been established by Kattenberg *et al.* (1996) and demonstrated in our DDC sample. However, we cannot infer merely from differences between models that the models disagree, because internal variability has not been taken into account. Our HadCM2 sample enables us to place the inter-model differences in the context of the differences that might be expected

[283] The three models are CGCM1, CSIRO2, and HadCM2.

[284] Although the conclusions of Kattenberg *et al.* (1996) are drawn with respect to annual means, they also apply to the JJA patterns presented for the DDC sample.

[285] The JJA global-mean precipitation changes (mm/day) per degree of annual global warming are as follows: 0.01 (CCSR-NIES), 0.03 (CGCM1), 0.06 (CCSR-NIES), 0.02 (ECHam4), 0.03 (HadCM3), 0.04 (HadCM2 Ga 1). These numbers are the global-means of Figures 7.6 and 7.8.

[286] Since our plots are for JJA this only applies to the high latitudes of the southern hemisphere.

[287] This conclusion particularly concerns the monsoon in JJA (*e.g.* their Figure 6.11).

merely from internal variability. A visual inspection of Figures 7.7 and 7.8 (top) provides an indication at the level of the individual grid-box of the range of responses arising from internal variability.[288]

We quantify the comparison between inter-model differences and internal variability by plotting the ratio between the DDC range and the HadCM2 95% confidence interval. In some grid-boxes (12.5%) the DDC range of temperature responses is actually smaller than the HadCM2 confidence interval (Figure 7.9, top right), but in most (54.4%) the DDC range is at least twice the HadCM2 interval. However, this result adds an important caveat to the conclusions drawn by Kattenberg *et al.* (1996). It suggests that for as much as half the globe, the inter-model differences might be at least halved if internal variability were considered. The result for precipitation (Figure 7.10, top right) is even more striking. The DDC range of precipitation responses is smaller than the HadCM2 confidence interval for almost half the grid-boxes (46.4%), and in three-quarters (77.2%) the DDC range is less than twice the HadCM2 interval.

We cannot prove from these results that the inter-model differences may be explained by internal variability, even for the proportion of the grid-boxes where the DDC range is less than the HadCM2 confidence interval, because we only have a measure of internal variability from a single model – HadCM2. To give such a proof we would need a measure of internal variability from each of the models. However, the comparison of the internal variability of HadCM2 with the inter-model differences does indicate that a large proportion of the 'disagreements' between models on regional scales may not, in fact, be 'disagreements' at all.[289]

7.4d Dependence on spatial scale

[288] It should be noted that Figures 7.7 and 7.8 (top) are not response *patterns* as such, because taken *as a whole* they are not plausible representations of HadCM2's response to forcing. They are representations of the 95% confidence interval for *individual* grid-boxes, independently of any other grid-boxes, and should only be considered at regional scales.

[289] This conclusion differs, at least in emphasis, from that of Giorgi and Francisco (2000): "the uncertainty in internal model variability for temperature is much lower than the uncertainty in inter-model variability" (p179). Their results, partly drawn from Kittel *et al.* (1998), arise from a method that differs from our own in two key respects. Firstly, they used a larger spatial scale than ourselves (*cf.* the next section). Secondly, their comparison was based on the regional climate sensitivity, whereas we used a 'purer' measure of the spatial pattern by calculating the regional climate response accompanying each °C change in the scaler.

There is considerable interest in repeating this analysis for the Countries set, not merely to express the relationship between inter-model differences and internal variability for political entities, but also because of the increase in spatial scale. If we increase the spatial scale we may expect both the DDC range and the HadCM2 95% confidence interval to decrease, but not necessarily by equal amounts.

Since the Countries set only includes land areas, we first repeat the previous analysis for land grid-boxes only.[290] Under this restriction the ratio of DDC range to HadCM2 interval decreases for temperature and increases for precipitation.[291] We repeat the entire sequence of operations for the Countries set, from anomalising, through pattern construction, to ratio calculation, and plot the results in Figures 7.11 and 7.12. The result is an increase (relative to the land grid-boxes) in the ratio between DDC range and HadCM2 interval for both temperature and precipitation.[292] In other words, any decrease in the DDC range brought about by the increased spatial scale is more than matched by a corresponding decrease in the HadCM2 interval. This result presents the possibility that inter-model agreement may not always be greater as the spatial scale increases, provided one expresses inter-model agreement in terms of the internal variability of the models.

7.4e Inferential statistics

Although we are restricted in the inferential statistics we may calculate by the non-random nature of the DDC sample, we can draw inferences in relation to the HadCM2 sample. We have already defined the 95% confidence interval for the HadCM2 population mean. By comparing the individuals within the DDC sample with the HadCM2 confidence interval we may determine whether the grid-box response from another model may plausibly be drawn from the HadCM2 population. We present the results in Figures 7.13 and 7.14 for four models. A

---

[290] That is, the land boxes on the common grid. The contributory boxes from an individual model grid to a particular 'land' box on the common grid may not be (entirely) land boxes.

[291] When we consider only land grid-boxes rather than all grid-boxes, the proportion of temperature (precipitation) grid-boxes where the DDC range is less than the HadCM2 95% confidence interval increases from 12.5% to 21.7% (decreases from 46.4% to 34.4%). The proportion where the DDC range is more than twice the HadCM2 95% confidence interval decreases from 54.4% to 37.4% (increases from 22.8% to 38.3%).

[292] When we consider the Countries set rather than land grid-boxes, the proportion of temperature (precipitation) grid-boxes where the DDC range is less than the HadCM2 95% confidence interval decreases from 21.7% to 2.2% (decreases from 34.4% to 10.3%). The proportion where the DDC range is more than twice the HadCM2 95% confidence interval increases from 37.4% to 49.2% (increases from 38.3% to 46.7%).

visual inspection shows that in each case the pattern from the DDC model lies outside the 95% confidence interval from HadCM2. This is true for temperature (Figure 7.13) for the vast majority of grid-boxes, less overwhelmingly so for precipitation (Figure 7.14).

7.5 IPCC regional responses

The model inter-comparisons in the literature on regional climate change mostly use the IPCC region set,[293] and it is mostly on the basis of these inter-comparisons that models are generally believed to disagree on regional scales. We have already developed a region set (the 'Giorgi' set) similar to the IPCC set in section 6.2. By calculating our normalised response patterns for the Giorgi set, and by comparing the differences between the models with the differences arising from internal variability, we have an opportunity to extend the literature on model inter-comparisons. Thus we build upon the comparison of single simulations from nine models (Kittel *et al.*, 1998) and the comparison of ensembles from a single model (Giorgi and Francisco, 2000).

We calculate response patterns for the mean changes in JJA temperature and precipitation in the 2080s for each member of the DDC and HadCM2 samples, using the Giorgi region set. We plot the normalised temperature (Figure 7.15) and precipitation (Figure 7.16) responses for each region (Table 7.3) in a manner that mimics the presentations of Kittel *et al.* (1998) and Giorgi and Francisco (2000). The scatter of circles represents the responses among the members of the DDC sample; the filled circle and bar represent our estimate of the HadCM2 population mean and the 95% confidence interval for our estimate.

The Second Assessment Report (SAR) of the IPCC emphasised the differences between the models:

> "Scenarios produced by these transient experiments varied widely among models and from region to region, both for temperature and precipitation … For most regions, the inter-model range of simulated temperature increase was rather pronounced, about 3–5°C" (Kattenberg *et al.*, 1996, p337)

Further details were given in Kittel *et al.* (1998), who were more explicit about inter-model differences in precipitation. They added that in most regions "there was little agreement among models in response sign or magnitude" (p10).

Our results *appear* to be somewhat different:

[293] We introduced the IPCC region set in section 2.10.

- The typical inter-model range in regional temperature anomalies for a doubling of CO_2 is 3-5°C. The mean inter-model range of normalised responses from our DDC sample is merely 0.57°C.
- The typical inter-model range in regional precipitation anomalies for a doubling of CO_2 is 30%. The mean inter-model range of normalised responses from our DDC sample is merely 10.2%.

The results appear to be different because of the differences in method:

- The SAR employed nine models, rather than six, which would make the SAR ranges slightly larger than ours.
- The SAR considered seven regions, rather than twenty-two, which would make the SAR ranges slightly smaller than ours.
- The SAR calculated anomalies at CO_2 doubling, whereas we calculated anomalies per degree of global warming, which eliminates most of the inter-model differences arising from different model sensitivities. This would tend to make our ranges smaller than the SAR ranges.
- The SAR calculated anomalies at the time of CO_2 doubling, when the global-mean temperature change averaged 2.1°C among the nine transient experiments. Our anomalies are expressed per °C of global-mean warming.

The SAR did not explicitly consider the contribution of internal variability to the inter-model ranges, although Kittel *et al.* (1998) super-imposed an observed measure of decadal variability. For some regions internal variability might be sufficient to explain the inter-model differences. A notable example is Australian precipitation, for which the 95% confidence interval for the HadCM2 mean straddles all six model values. The average HadCM2 interval among the 22 regions amounts to 0.26°C and 6.4%, approximately half the inter-model ranges given above.

We draw two conclusions from these comparisons.

i. A substantial proportion of the differences between single simulations from different models appears to arise from internal variability.

ii. If we treat regional manifestations of climate change *separately* from the global-mean response to climate change, then the differences between models are greatly reduced. This is of considerable importance in the context of pattern scaling, where we *assume* that we may treat regional manifestations of climate change separately from the global-mean response to climate change.

It appears that the inter-model ranges that are often quoted in support of the argument that models remain in disagreement on regional scales, are in fact much smaller than is often thought. A large proportion of the 'disagreements' may be attributed to different climate sensitivities within the models, or simply to internal variability.

7.6 Estimation by scaling patterns

The objective of this chapter is to establish whether or not it is reasonable to apply pattern scaling to a wider range of GCMs than merely HadCM2. In this section we attempt to reproduce the climate anomalies for the late 21st century simulated by each model. We estimate by scaling each pattern by the scaler from the same model. We estimate for temperature (odd-numbered figures) and precipitation (even-numbered figures).

7.6a Estimating for ECHam4

We illustrate our method in Figures 7.17 and 7.18 using one of the members of the DDC sample: ECHam4. We scale the ECHam4 temperature and precipitation response patterns calculated in section 7.3 (and plotted in Figures 7.5 and 7.6) and by the global-mean temperature in the ECHam4 GG simulation in the 2080s. We plot the product (top right), together with the modelled value (top left) and the estimation error (bottom left). We cannot calculate the *statistical* significance of the error as we did in chapter 4 because we only have a single simulation for each model. However, we can consider the *practical* significance of the errors. As we noted in chapter 4, practical significance is heavily dependent on the context, but the proportion of the modelled value made up by the error does provide one indication of the practical significance of the error. Therefore we also plot the fraction of the modelled anomaly that the error represents (bottom right).

A visual inspection of Figures 7.17 and 7.18 (top) shows a strong degree of correspondence between the modelled and estimated anomalies. However, errors can be identified (bottom left), notably in temperatures over the Southern Ocean and in tropical precipitation. As a fraction of the anomalies being estimated (bottom right), the precipitation errors are larger than the temperature errors. However, outside the Southern Ocean the temperature errors do not exceed a fifth

of the value being estimated, and many of the largest errors in estimating precipitation are small in relation to the anomaly being estimated.[294]

We repeat these analyses for the Countries set, rather than individual grid-boxes, and we plot the results in Figures 7.19 and 7.20. The increase in spatial scale may be expected to reduce levels of internal variability, and thus to reduce any contribution from internal variability towards the error obtained from pattern scaling. This does appear to have occurred, particularly for precipitation,[295] but it is difficult to identify differences from visual inspection. Later we give a more quantitative evaluation of the influence of internal variability on the error obtained from pattern scaling, but first we broaden the scope to the rest of the DDC sample.

7.6b Estimating for the DDC sample

We repeat the estimation method applied above to ECHam4 to four other models from the DDC sample: CCSR-NIES, CGCM1, CSIRO2, and HadCM3. We do not plot the fractional errors because they are very similar to those obtained for ECHam4, indicating that the practical significance of the errors – particularly for temperature – is quite small. Instead we examine the possibility that the models may exhibit a common non-linearity in their responses to forcing, which – if it occurs – would limit the application of pattern scaling. We plot the estimation errors for the four models in Figures 7.21 and 7.22.[296]

By visually inspecting Figures 7.21 and 7.22 we may identify any regions where the errors in the various models have the same sign. In fact there are not many regions where this is the case: the most obvious examples are the temperature estimates being too cool in Australia and around Lake Chad, and the precipitation estimates being too moist in West Africa. The fact that in these regions the errors from pattern scaling all have the same sign suggests that the errors may not arise merely from internal variability, but from a non-linearity in the response to forcing in those regions. The fact that the West African estimates are both too cool and too

[294] In Figure 7.18, among the 194 grid-boxes (2.8% of the total) with the largest errors (defined as those with a magnitude greater than 0.4 mm/day), in 115 of the grid-boxes (59.3%) this error represents less than two-fifths of the value being estimated.

[295] Compare Figures 7.18 and 7.20, particularly South America and Australia.

[296] Note the different scale to the plots of the ECHam4 errors in Figures 7.17 – 7.20, where the scale was chosen to optimise the comparison with the modelled and estimated values. Here the scale was chosen to optimise the comparison between model errors.

moist is particularly suggestive. However, a non-linear response to forcing in one model may not be present in another, so the fact that there are errors of both signs in most regions implies neither that the response is linear, nor that the errors may be completely explained by internal variability.

7.6c Incorporating internal variability

In order to directly address the contribution of internal variability to the errors we turn to the HadCM2 sample. By repeating the estimation procedure for each of the four individuals we obtain a sample of four errors (for each grid-box) that differ from each other solely through internal variability.

From the sample of four errors we calculate the 95% confidence interval into which we expect the HadCM2 mean error to fall. We plot the limits of the 95% interval in Figures 7.23 and 7.24 (bottom). It is worth noting that this interval fails to span zero in only a few regions (notably West Africa again, and Central America), indicating that in most regions the HadCM2 pattern scaling error is not statistically significant. For comparison we plot the minimum and maximum errors found at each grid-box among the six individuals in the DDC sample (Figures 7.23 and 7.24, top). A visual inspection shows that the DDC and HadCM2 samples have similar ranges of error both over the globe as a whole, and in terms of the spatial distribution of error. In both samples the temperature errors are greatest over land, and the precipitation errors are greatest in the tropics.

We make a quantitative comparison in Figures 7.25 and 7.26, where we express the DDC range (top left) and the HadCM2 95% confidence interval (top right) as a ratio (bottom left). For temperature errors, the DDC range is less than the HadCM2 interval in half the grid-boxes (51.0%), and for precipitation the proportion is two-thirds (67.3%). We emphasise again that this does not *prove* that the errors from applying pattern scaling to DDC models in those grid-boxes can be explained by internal variability, because the measure of internal variability is only taken from a single model. However, it does *suggest* that many of the errors may be explained by internal variability, and that even where the DDC range exceeds the HadCM2 interval, a proportion of the errors may be due to internal variability.

Finally we compare our estimate of statistical significance with our estimate of practical significance.[297] Our measure of statistical significance is the ratio of DDC range to HadCM2 interval (Figures 7.25 and 7.26, bottom left), and we make the threshold for significance a ratio of 1.0. Our measure of practical significance is the median fractional error among the DDC sample, and the threshold for significance is 20% of the modelled value. The proportion of grid-boxes where the error is both statistically and practically significant under these definitions is 2.8% (temperature) or 23.7% (precipitation). If we raise the threshold for practical significance to 40%, the proportions drop to 0.0% (temperature) or 8.8% (precipitation). Therefore we suggest that for most grid-boxes, the errors introduced by applying pattern scaling to a GCM are not simultaneously statistically and practically significant.

7.7 IPCC estimation by scaling patterns

We extend our analysis by applying pattern scaling to each region in the 'Giorgi' set. The response patterns are those described in section 7.5 and displayed in Figures 7.15 and 7.16. We scale these patterns by the global annual temperature in the 2080s from the same simulations, and thus obtain estimates of temperature (Figure 7.27) and precipitation (Figure 7.28) in the 2080s, which we plot as plus signs. The anomalies simulated by the models are plotted as squares. As in Figures 7.15 and 7.16, we do not plot the HadCM2 sample members individually; instead we plot the sample mean and the 95% confidence interval (as a bar) for the population mean.

Our purpose in presenting Figures 7.27 and 7.28 is not to emphasise the range of modelled values among the DDC sample,[298] but to point out the close correspondence between the modelled and estimated values. In almost all cases the error is less than 0.5°C or 0.5 mm/day. We identify the errors more clearly in the accompanying Figures 7.29 and 7.30. In almost all regions there are both positive and negative errors among the DDC models, and in almost all cases the HadCM2 95% confidence interval straddles zero. Where the HadCM2 interval

[297] As above, we emphasise that practical significance is context-dependent, and the measure that we present here is but one possible measure of practical significance.

[298] It should be noted that the range of modelled values in each region is not merely due to the combination of internal variability and genuine model disagreements (as in Figures 7.15 and 7.16). Here the climate sensitivity is also a contributory factor, as may be seen from the scaler by which the patterns are multiplied to obtain estimates. For CGCM1 the scaler is 5.6°C, whereas for ECHam4 it is 3.4°C.

straddles zero the errors introduced by applying pattern scaling to HadCM2 are not statistically significantly. Once again, this measure of internal variability only directly applies to HadCM2, but it does provide an indication of the levels of internal variability that may be present among the DDC members. If other models have the same level of internal variability as HadCM2, then almost all the errors may be explained by internal variability.

One of the principal exceptions to these results is Central America (CAM). The range of errors from applying pattern scaling to the DDC sample in Central America is much greater than the 95% confidence interval from HadCM2 (Figures 7.29 and 7.30). These errors may prove to be statistically significant, but the relatively narrow range of modelled temperatures (Figure 7.27) suggests that the temperature errors may not be practically significant. On the other hand, the precipitation errors add 50% to the range of modelled values. West Africa presents us with a different example. Here the primary problem is – as noted in section 7.6 – the consistently cold (Figure 7.29) and moist (Figure 7.30) errors among the models, suggesting a consistently non-linear response among the models.

7.8 Conclusions

In this chapter we have examined the technique of pattern scaling for a sample of six models, and using a sample of four simulations for one of those models. The detailed choices were intended to provide us with a fair comparison of the models (section 7.2). We established that the differences between the patterns constructed here and those constructed in previous chapters would not divorce this chapter from the rest of the thesis (section 7.3).

We extracted JJA temperature and precipitation response patterns from each of the models and examined them (section 7.4). The response patterns were in broad agreement with IPCC assessments of model responses to radiative forcing, but the additional information about internal variability suggested that a large proportion of the inter-model differences might be attributable to internal variability, and might not be 'disagreements' at all. Another interesting result was found when the spatial scale was increased: any decrease in the inter-model differences was more than matched by corresponding decreases in the internal variability of HadCM2. This suggests that although the inter-model differences may become smaller as the spatial scale increases, they may also become more statistically significant.

We also constructed response patterns for regions approximating to the IPCC set (section 7.5). We presented results in a similar format to that employed in the IPCC assessments. By calculating normalised patterns we removed the contribution to inter-model differences from their different climate sensitivities. By including a HadCM2 ensemble we also took into account the contribution to inter-model differences from internal variability. As a result, the inter-model 'disagreements' appear to be much smaller than those discussed in the IPCC Second Assessment Report.

In section 7.6 we applied the response patterns to the task of estimating the modelled climate anomalies in the 2080s. We found that the errors from pattern scaling were generally much smaller than the anomalies being estimated. Although practical significance is heavily dependent on the context, we considered one measure of practical significance which indicated that the practical significance of the errors introduced by scaling may be fairly low. We also found that there were few regions where the errors from pattern scaling were all of the same sign, suggesting that it is unlikely that the models exhibit non-linearities in the same ways. We evaluated the contribution of internal variability to the errors, and found that in many regions the range of errors from internal variability is comparable to the inter-model range. We concluded that the errors are likely to be both statistically and practically significant in only a small proportion of the grid-boxes. Finally (section 7.7) we repeated the estimation procedure for the IPCC region set, and demonstrated the validity of the points made above for these particular regions.

The general conclusion we might draw from these analyses is that scaling appears to 'work' for a variety of models, not just HadCM2. It ought to be noted that we have only considered two variables and a single season. Given the different behaviour displayed by temperature and precipitation, the technique ought only to be applied to other variables with considerable caution. However, Mitchell *et al.* (1999) obtained accurate *annual* estimates by pattern scaling, so we would not expect very different results for seasons other than JJA.

It should also be noted that we have only scaled a pattern from a simulation to estimate for a period at the end of *the same* simulation. It is very likely that if the same patterns were used to estimate anomalies under a weaker forcing scenario, larger errors would result. We cannot evaluate this for the various models because of the lack of suitable simulations. However, when we examined this question for HadCM2 (in chapter 4), we found that the errors so introduced did not increase so

much that the technique became invalid. The broadly linear responses to forcing evident from the small size of the errors from estimating for GG, and the similarity between the errors found for HadCM2 and the other models, both suggest that the conclusions drawn in chapter 4 with respect to HadCM2 may apply to the other models also.

The answer to the question that we posed in the title of this chapter, namely 'May we scale other models (than HadCM2])?', is 'yes'. Within the limits of the material we have examined in this chapter, there is no evidence that the development and application of response patterns for various DDC models introduces any more errors than are introduced when using HadCM2. The practical importance of this conclusion may be noted in the context of the considerable uncertainties in regional climate assessments that stem from differences between models. Our analysis suggests that we may apply pattern scaling to a variety of models and obtain estimates of change from each model. It is a reasonable to suggest that we might treat such estimates as equally plausible, and present all of the results in an attempt to represent the uncertainty arising from differences between models.

Another possibility that we do not explore is to scale a response pattern from one model in an attempt to estimate changes in another model. This latter approach could only conceivably be useful if we were attempting to estimate changes that would be simulated by a model with defined parameters that had never been constructed. The reason for attempting such a thing would stem from the recognition that there is an infinite variety of possible models that would provide plausible interpretations of the climate system; we might attempt to represent the uncertainty arising from the infinite variety of plausible models by estimating the changes simulated by those plausible models.

Table 7.1

Spatial specifications of six models with GG scenarios extending to 2099.

atmosphere

model	res	X-long	Y-lat	Z-vert	total
CCSR-NIES	T21	64	32	20	40,960
CGCM1	T32	96	48	10	46,080
CSIRO2	R21	64	56	9	32,256
ECHam4	T42	128	64	19	155,648
HadCM2	T42	96	73	19	133,152
HadCM3	T42	96	73	19	133,152

ocean

model	res	X-long	Y-lat	Z-vert	total
CCSR-NIES	2.8	128	64	17	64,100
CGCM1	1.8	192	96	29	256,600
CSIRO2	5.6*3.3	64	56	21	39,400
ECHam4	2.8	128	64	11	43,300
HadCM2	2.5*3.8	96	73	20	73,200
HadCM3	1.3	288	145	20	439,200

coupled

model	total	ratio	sensitivity
CCSR-NIES	105060	0.64	3.5
CGCM1	302680	0.18	3.5
CSIRO2	71656	0.82	4.3
ECHam4	198948	3.59	2.6
HadCM2	206352	1.82	2.5
HadCM3	572352	0.30	3.3

After Lambert and Boer (2000) Table 1.

Some entries have been modified, HadCM3 has been added to the models, and equilibrium climate sensitivity has been added to the variables.

Ocean resolution given in degrees between grid points.

Coupled ratio represents atmospheric points / oceanic points.

Climate sensitivity (°C) added from Kattenberg *et al.* (1996) Table 6.3.

Table 7.2

Dynamical specifications of six models with GG scenarios extending to 2099.

model	model reference	GG reference
CCSR-NIES	Emori *et al.* (1999)	Emori *et al.* (1999)
CGCM1	Flato *et al.* (2000)	Boer *et al.* (2000)
CSIRO2	Gordon and O'Farrell (1997)	Gordon and O'Farrell (1997)
ECHam4	Roeckner *et al.* (1996)	Roeckner *et al.* (1999)
HadCM2	Johns *et al.* (1997)	Mitchell *et al.* (1999)
HadCM3	Gordon *et al.* (2000)	Mitchell *et al.* (1998)

model	ocean dynamics	ocean boundary	ocean vertical	ice
CCSR-NIES	primitive	rigid lid	height	basic
CGCM1	primitive	rigid lid	height	basic
CSIRO2	primitive	rigid lid	height	rheology
ECHam4	primitive	free surface	density	rheology
HadCM2	primitive	rigid lid	height	drift
HadCM3	primitive	rigid lid	height	drift

model	fluxes adjustments	land surface	land stomata	land routing	initial
CCSR-NIES	heat, water	bucket		ocean	equilib
CGCM1	heat, water	bucket+	without	ocean	equilib
CSIRO2	heat, water, mom	canopy	included	ocean	equilib
ECHam4	heat, water	canopy	without	ocean	equilib
HadCM2	heat, water	canopy	included	ocean	partial
HadCM3	none	canopy	included	ocean	obs

The lower two tables were constructed following Kattenberg *et al.* (1996) Table 5.1, and Lambert and Boer (2000) Table 2.

ocean dynamics: primitive or quasi-geostrophic equations

ocean boundary: a rigid lid or a free surface

ocean vertical co-ordinate: height or density/isopycnal

ice: thermodynamic (basic), with dynamic rheology or ocean-forced-drift

flux adjustments: heat, water, and momentum

land surface: simple bucket, modified bucket (+), or canopy model

land stomata: stomatal resistance included or not included

land routing: run-off routed to ocean or not routed

initialisation: equilibrated ocean and atmosphere, partially equilibrated based on the upper ocean, or observations

Table 7.3

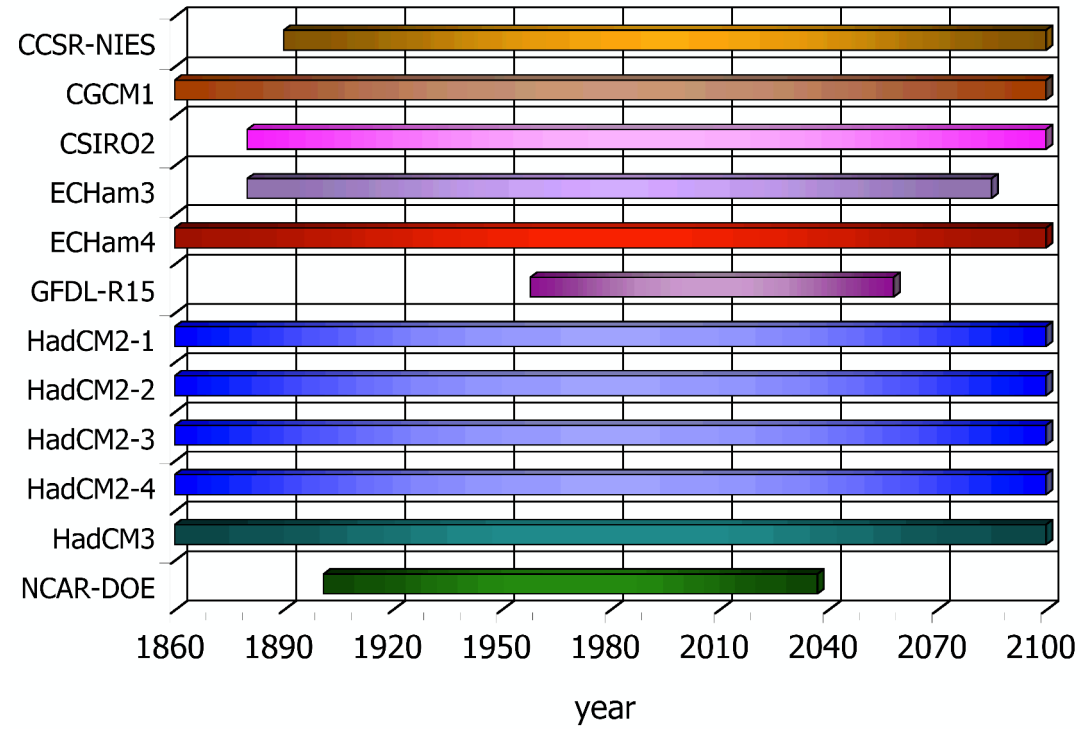
The regions in the Giorgi set.

As constructed in section 6.2 (3) and displayed in Figure 6.3 (bottom left).
The labelling is similar to Giorgi and Francisco (2000, see their Table 2), but the regions vary slightly from those they defined (see section 6.2 (3) for details). We also include the number of grid-boxes and the approximate surface area of the grid boxes (km^2).

label	region	boxes	area
ALA	Alaska	77	3,782,000
AMZ	Amazon Basin	114	13,020,000
ANT	Antarctica	640	13,195,000
AUS	Australia	75	7,761,000
CAM	Central America	24	2,594,000
CAS	Central Asia	76	6,604,000
CNA	Central North America	43	3,819,000
EAF	Eastern Africa	78	8,928,000
EAS	East Asia	93	8,323,000
ENA	Eastern North America	31	2,580,000
GRL	Greenland	154	6,177,000
NAS	North Asia	275	14,647,000
NEU	Northern Europe	67	4,072,000
SAF	Southern Africa	55	5,940,000
SAH	Sahara	95	9,986,000
SAS	South Asia	48	5,081,000
SEA	South East Asia	35	4,004,000
SEU	Southern Europe	51	4,455,000
SSA	Southern South America	52	5,010,000
TIB	Tibet	48	4,172,000
WAF	Western Africa	69	7,863,000
WNA	Western North America	71	5,680,000

Figure 7.1

The greenhouse-gas-only simulations from the regridded DDC models.



model grids: CCSR-NIES, CGCM1, CSIRO2, ECHam3, ECHam4, GFDL-R15, HadCM2, HadCM3, NCAR-DOE

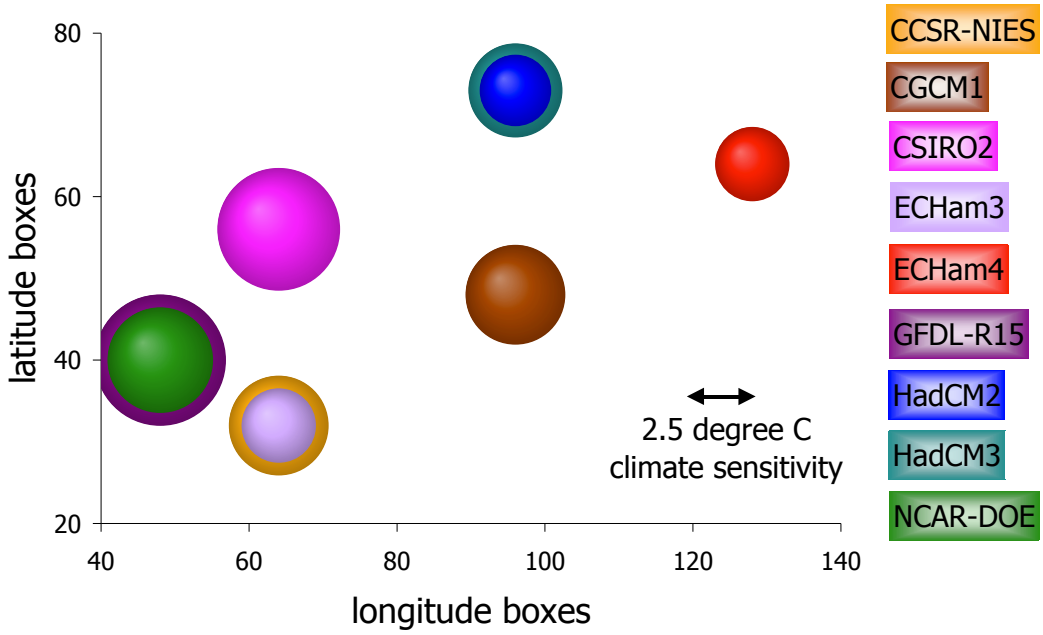
simulations: 'GG' simulations:

→1990: with observed radiative forcing

1990→: increasing radiative forcing of approx. 1%/annum

Figure 7.2

The horizontal resolution of the atmosphere in the regridded DDC models.



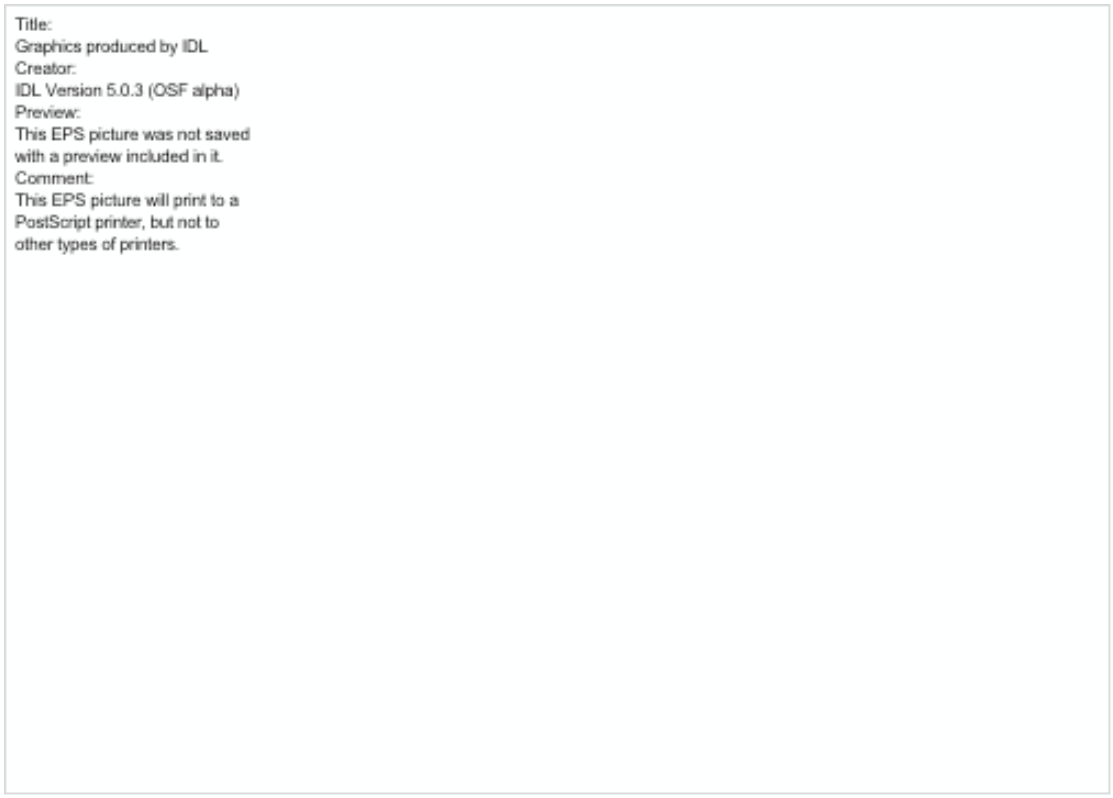
model grids: CCSR-NIES, CGCM1, CSIRO2, ECHam3, ECHam4, GFDL-R15, HadCM2, HadCM3, NCAR-DOE

sensitivity: equilibrium climate sensitivity (**diameter**, °C), sourced from: Kattenberg *et al.* (1996) Table 6.3 (pp 298-299)
DDC website: http://ipcc-ddc.cru.uea.ac.uk

dimensions: number of longitude (**x-axis**) and latitude (**y-axis**) grid-boxes

Figure 7.3

Temperature response patterns using sub-sets of the Ga ensemble.



model: HadCM2

simulations: Ga 1,2,3,4

temporal: JJA, 30-year means

patterns: *a* from least-squares regression ($y=ax$) of grid-box temperature (°C) and the scaler (global annual temperature, °C) for:

1860-2099, anomalised against 240y control (**top left**)

1990-2099, anomalised against 1960-1989 (**top right**)

grid-box temperature (°C) divided by the scaler (°C) for:

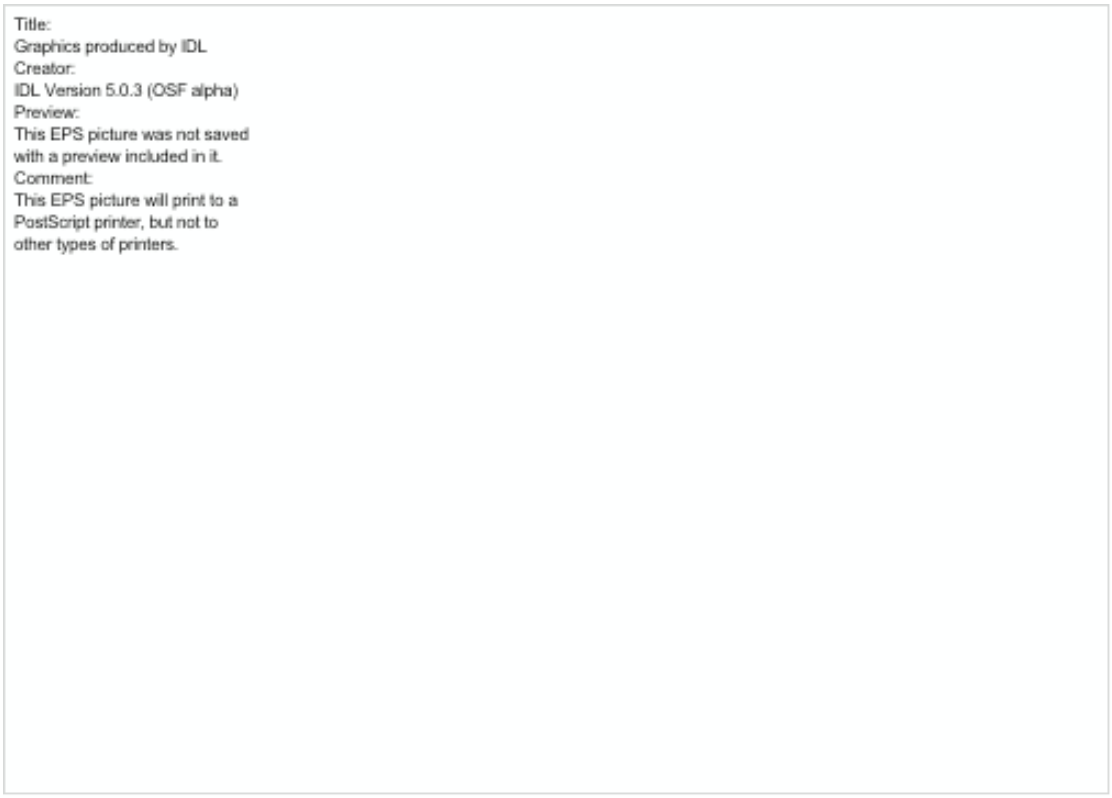
2010-2039 ('2020s'), and 2070-2099 ('2080s')

differences: top right – top left (**bottom left**)

2080s – 2020s (**bottom right**)

Figure 7.4

Precipitation response patterns using sub-sets of the Ga ensemble.

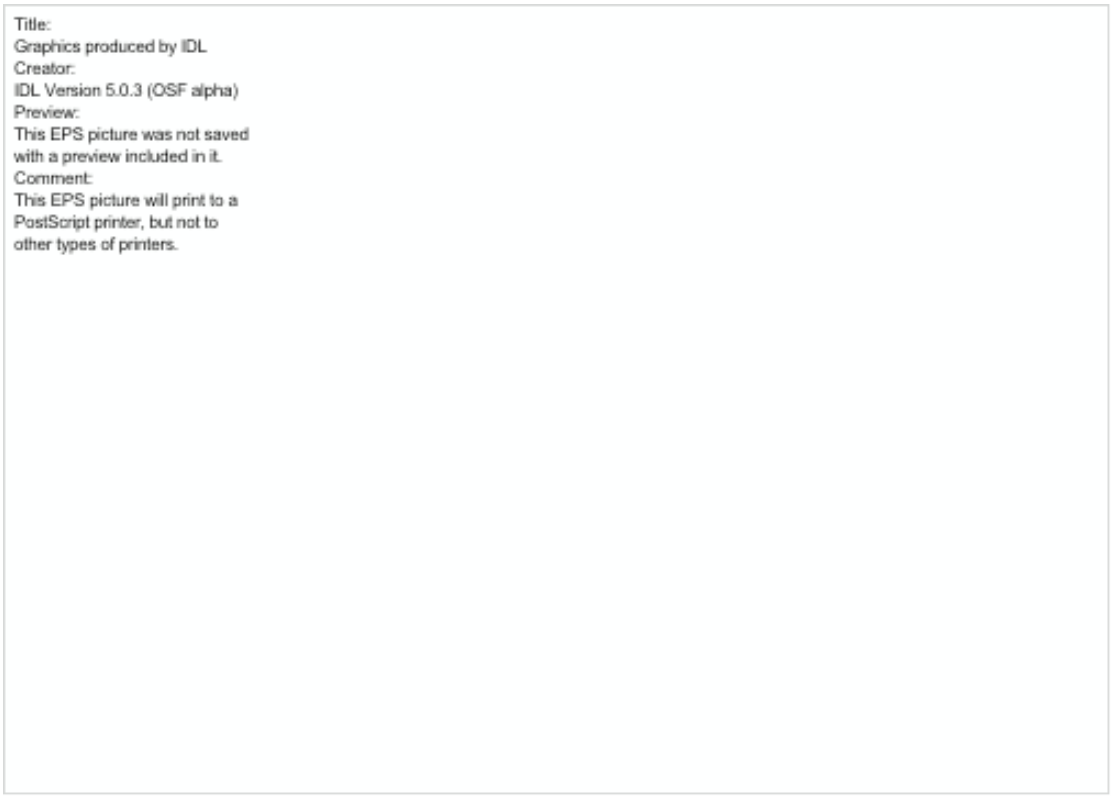
‘chapter 4’ pattern	‘chapter 7’ pattern
	
‘chapter 7’ – ‘chapter 4’	‘2080s’ – ‘2020s’

model:	HadCM2
simulations:	Ga 1,2,3,4
temporal:	JJA, 30-year means
patterns:	*a* from least-squares regression ($y=ax$) of grid-box precipitation (mm/day) and the scaler (global annual temperature, °C) for:
	1860-2099, anomalised against 240y control (**top left**)
	1990-2099, anomalised against 1960-1989 (**top right**)
	grid-box precipitation (mm/day) divided by the scaler (°C) for:
	2010-2039 (‘2020s’), and 2070-2099 (‘2080s’)
differences:	top right – top left (**bottom left**)
	2080s – 2020s (**bottom right**)

Figure 7.5

Temperature response patterns from four DDC models.

CCSR-NIES CGCM1



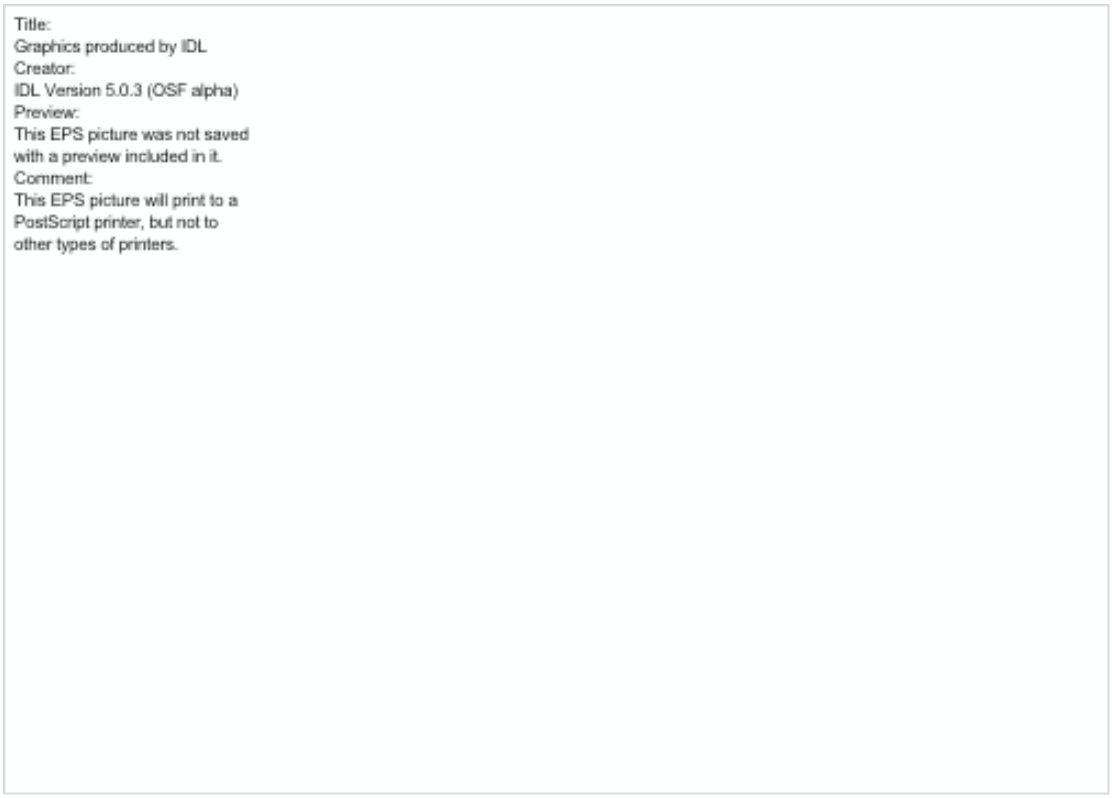
CSIRO2 ECHam4

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4
simulations:	GG on a common grid
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
patterns:	a from least-squares regression ($y=ax$) of grid-box temperature (y) and the scaler (global annual temperature, x)

Figure 7.6

Precipitation response patterns from four DDC models.

CCSR-NIES CGCM1

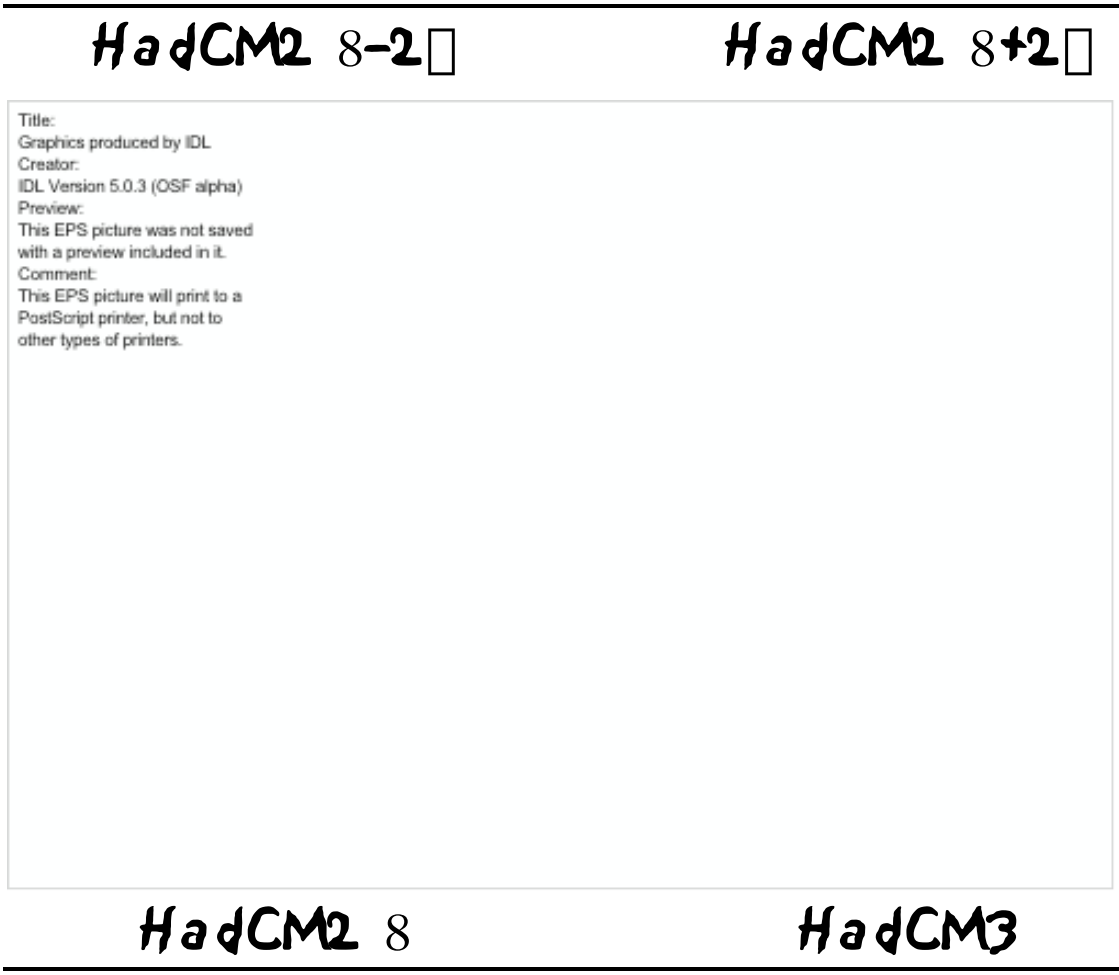


CSIRO2 ECHam4

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4
simulations:	GG on a common grid
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
patterns:	a from least-squares regression ($y=ax$) of grid-box precipitation (y) and the scaler (global annual temperature, x)

Figure 7.7

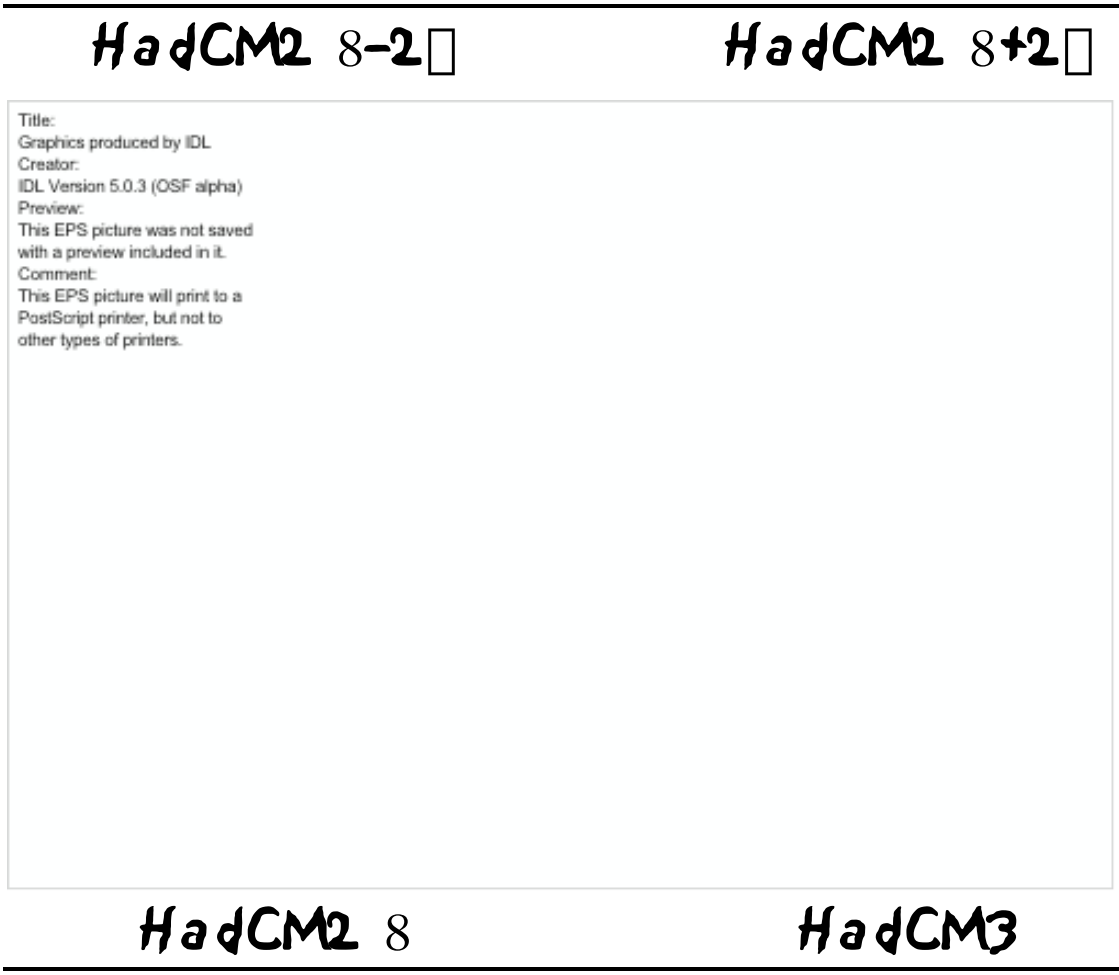
Temperature response patterns from two Hadley Centre models.



models:	HadCM2, HadCM3
simulations:	GG (HadCM3), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
patterns:	a from least-squares regression ($y=ax$) of grid-box temperature (y) and the scaler (global annual temperature, x)
means:	GG (**bottom right**), Ga 1,2,3,4 patterns (8, **bottom left**)
sd:	estimate of pop. standard deviation from Ga 1,2,3,4 patterns (□)
95% limits:	8 – 2□ (**top left**), 8 + 2□ (**top right**),

Figure 7.8

Precipitation response patterns from two Hadley Centre models.

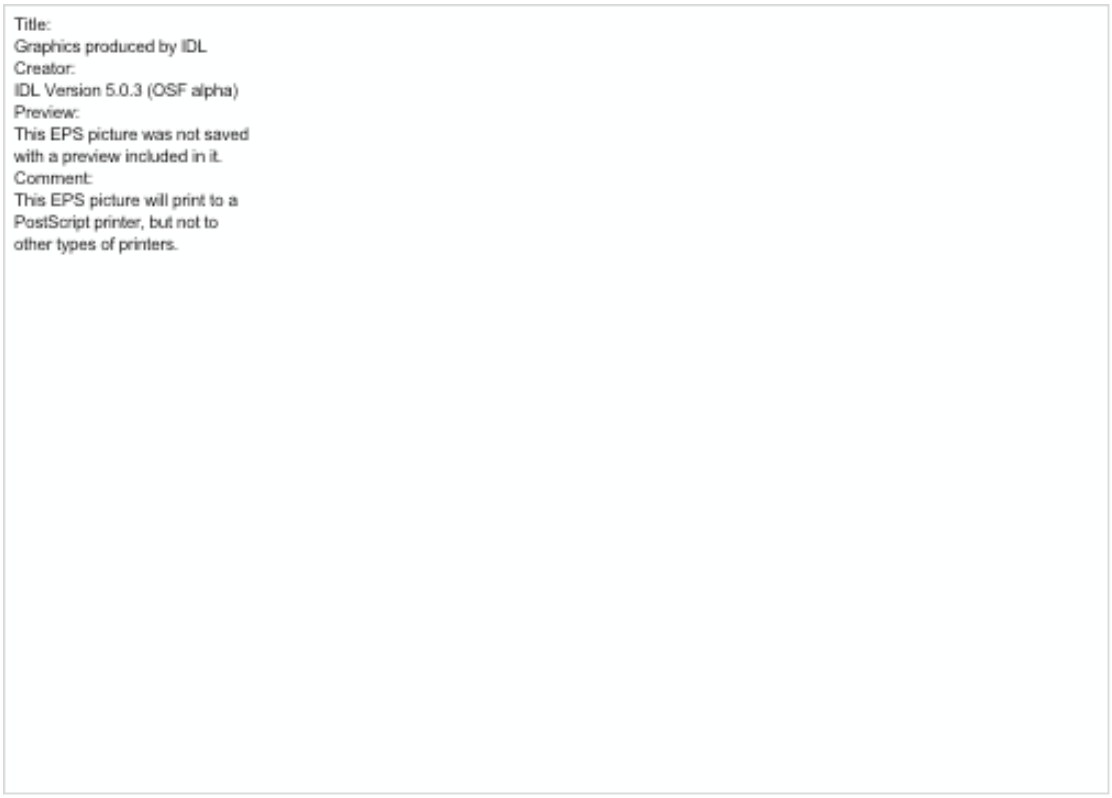


models:	HadCM2, HadCM3
simulations:	GG (HadCM3), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
patterns:	a from least-squares regression ($y=ax$) of grid-box precipitation (y) and the scaler (global annual temperature, x)
means:	GG (**bottom right**), Ga 1,2,3,4 patterns (8, **bottom left**)
sd:	estimate of pop. standard deviation from Ga 1,2,3,4 patterns (□)
95% limits:	8 – 2□ (**top left**), 8 + 2□ (**top right**),

Figure 7.9

Inter- and intra-model variations in temperature response patterns.

DDC median bottom left / bottom right



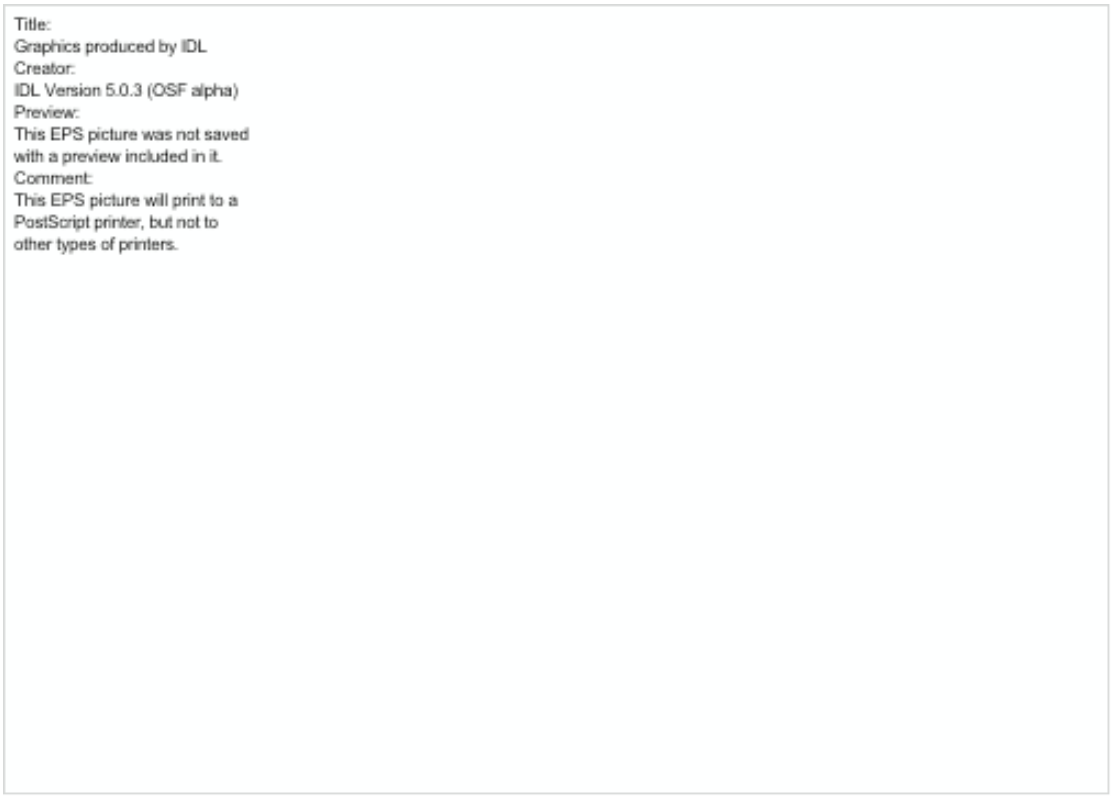
DDC range HadCM2 95% interval

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	grid-box values
patterns:	as calculated in Figures 7.5 and 7.7
DDC:	from the five GG patterns and HadCM2 Ga 1 pattern: median (**top left**), and range (**bottom left**)
95% interval:	from HadCM2 Ga 1,2,3,4 (**bottom right**, derived from Figure 7.7)
inter:intra	ratio: DDC range / HadCM2 95% interval (**top right**)

Figure 7.10

Inter- and intra-model variations in precipitation response patterns.

DDC median bottom left / bottom right



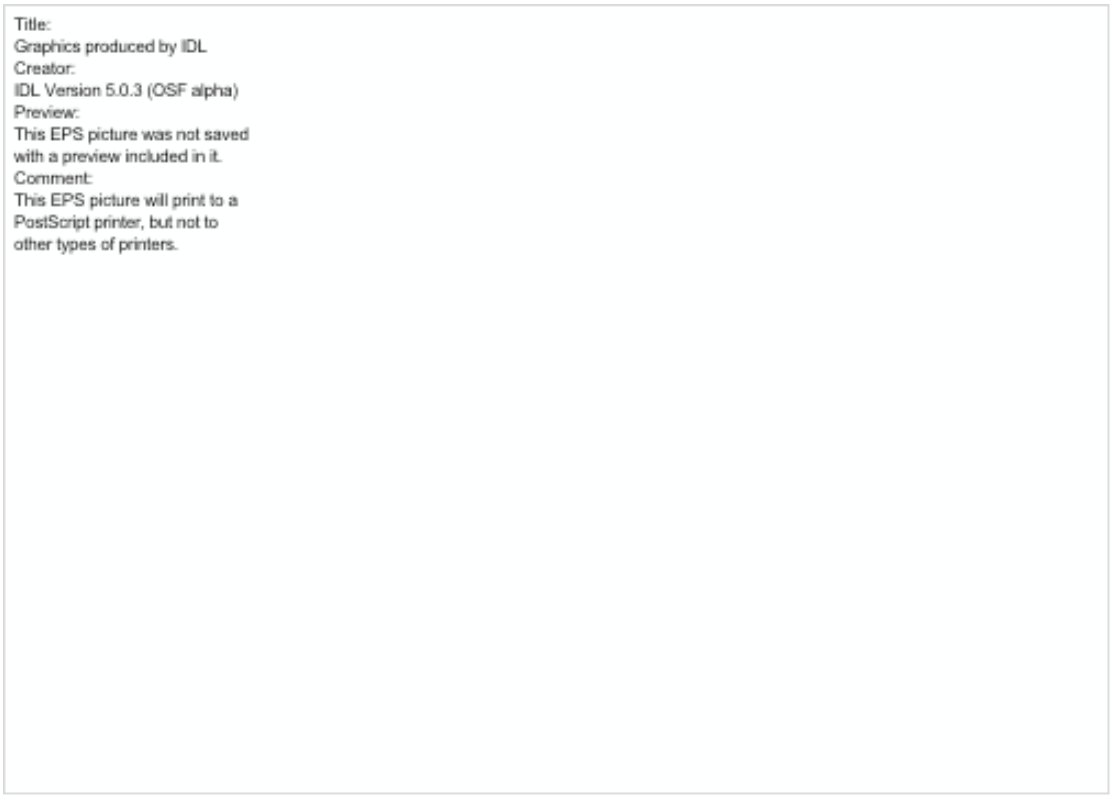
DDC range HadCM2 95% interval

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	grid-box values
patterns:	as calculated in Figures 7.6 and 7.8
DDC:	from the five GG patterns and HadCM2 Ga 1 pattern: median (**top left**), and range (**bottom left**)
95% interval:	from HadCM2 Ga 1,2,3,4 (**bottom right**, derived from Figure 7.8)
inter:intra	ratio: DDC range / HadCM2 95% interval (**top right**)

Figure 7.11

Inter- and intra-model variations in temperature response patterns.

DDC median bottom left / bottom right



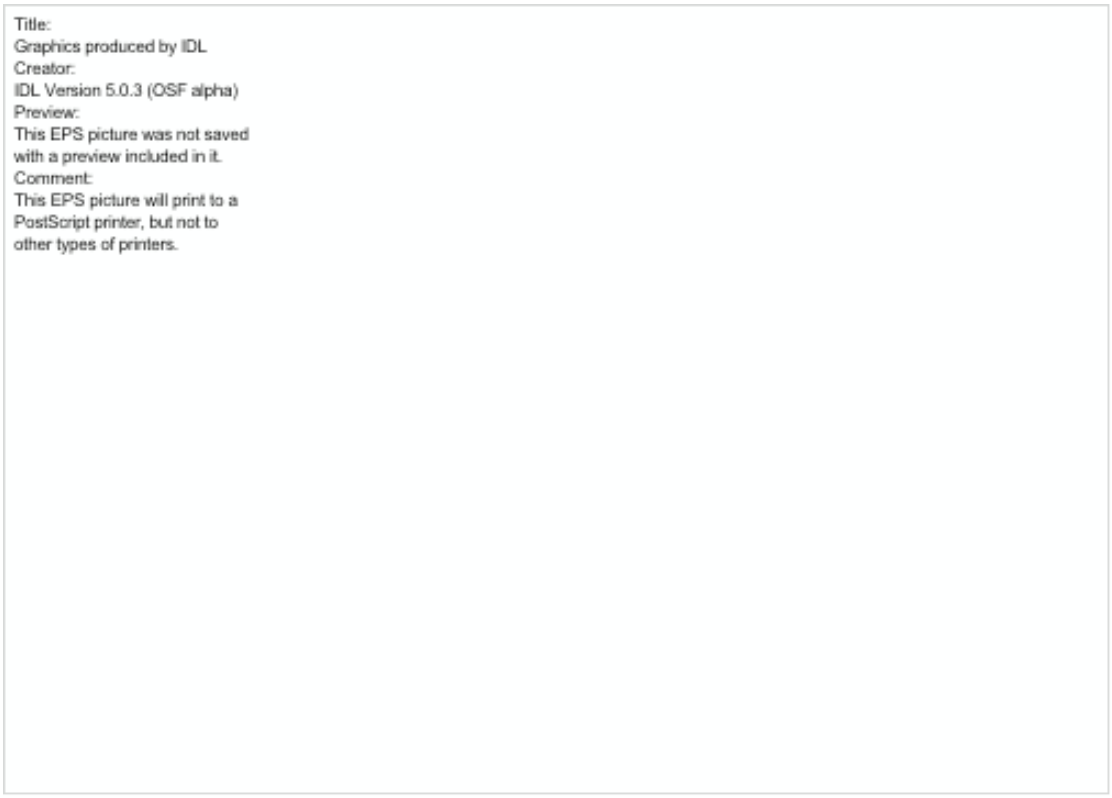
DDC range HadCM2 95% interval

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	Countries set regional means
patterns:	as calculated in Figures 7.5 and 7.7, but for the Countries set
DDC:	from the five GG patterns and HadCM2 Ga 1 pattern: median (**top left**), and range (**bottom left**)
95% interval:	from HadCM2 Ga 1,2,3,4 (**bottom right**, derived from Figure 7.7)
inter:intra	ratio: DDC range / HadCM2 95% interval (**top right**)

Figure 7.12

Inter- and intra-model variations in precipitation response patterns.

DDC median bottom left / bottom right



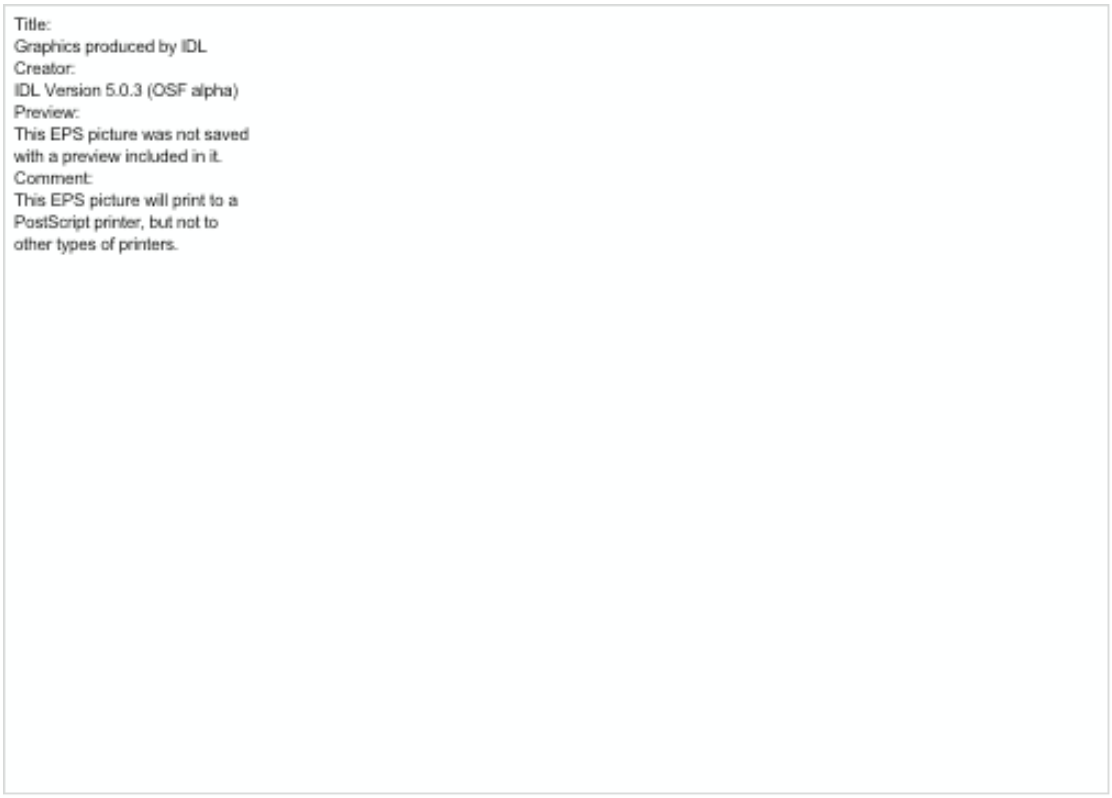
DDC range HadCM2 95% interval

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	Countries set regional means
patterns:	as calculated in Figures 7.6 and 7.8, but for the Countries set
DDC:	from the five GG patterns and HadCM2 Ga 1 pattern: median (**top left**), and range (**bottom left**)
95% interval:	from HadCM2 Ga 1,2,3,4 (**bottom right**, derived from Figure 7.8)
inter:intra	ratio: DDC range / HadCM2 95% interval (**top right**)

Figure 7.13

Significance of the differences between DDC patterns and the HadCM2 mean.

CCSR-NIES CGCM1



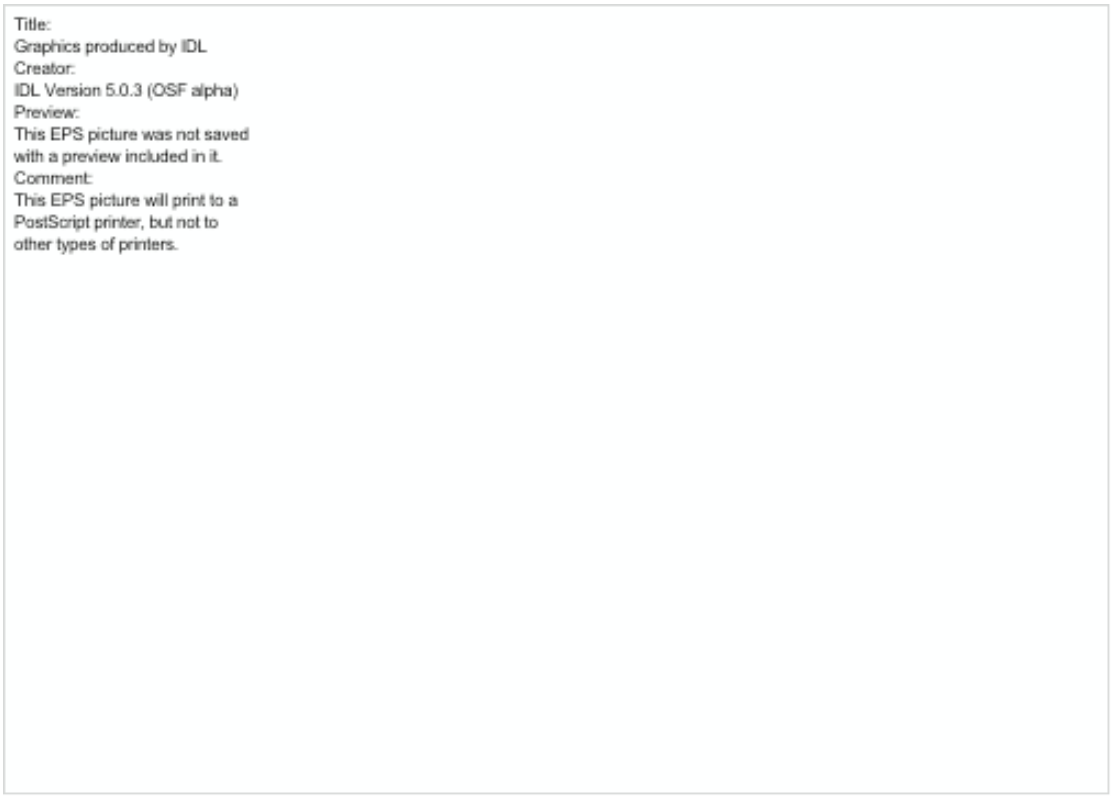
CSIRO2 ECHam4

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM2
simulations:	GG (first four models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	grid-box values
patterns:	as calculated in Figures 7.5 and 7.7
8, □:	pop. mean, sd from HadCM2 Ga 1,2,3,4 patterns (Figure 7.7)
difference:	DDC model pattern - 8
significance:	difference / □

Figure 7.14

Significance of the differences between DDC patterns and the HadCM2 mean.

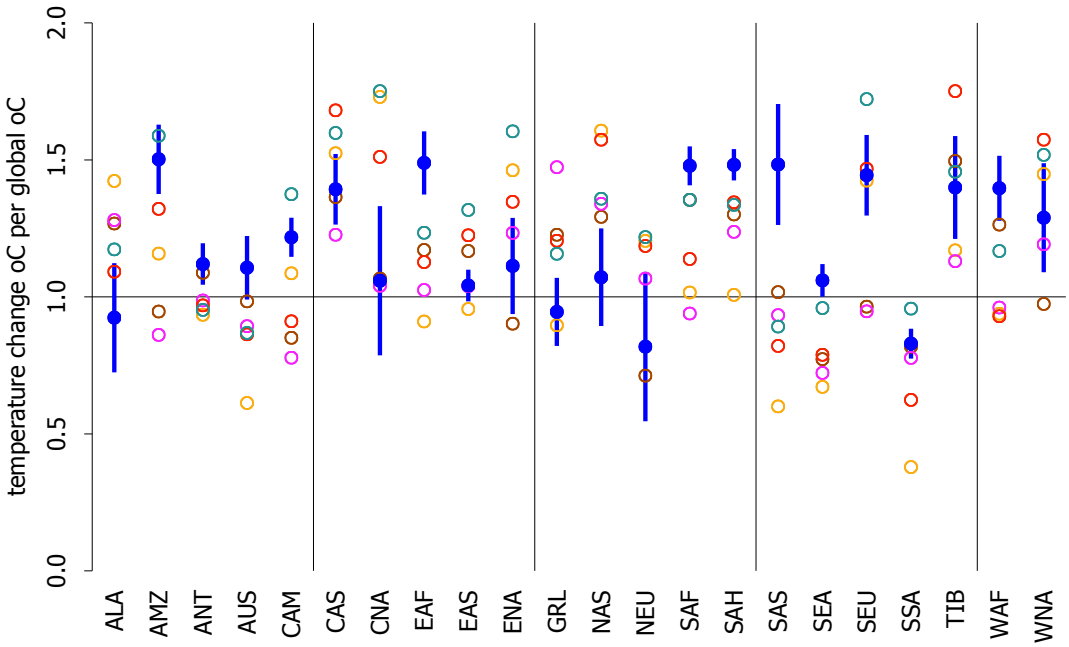
CCSR-NIES CGCM1



CSIRO2 ECHam4

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM2
simulations:	GG (first four models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	grid-box values
patterns:	as calculated in Figures 7.6 and 7.8
8, □:	pop. mean, sd from HadCM2 Ga 1,2,3,4 patterns (Figure 7.8)
difference:	DDC model pattern - 8
significance:	difference / □

Figure 7.15



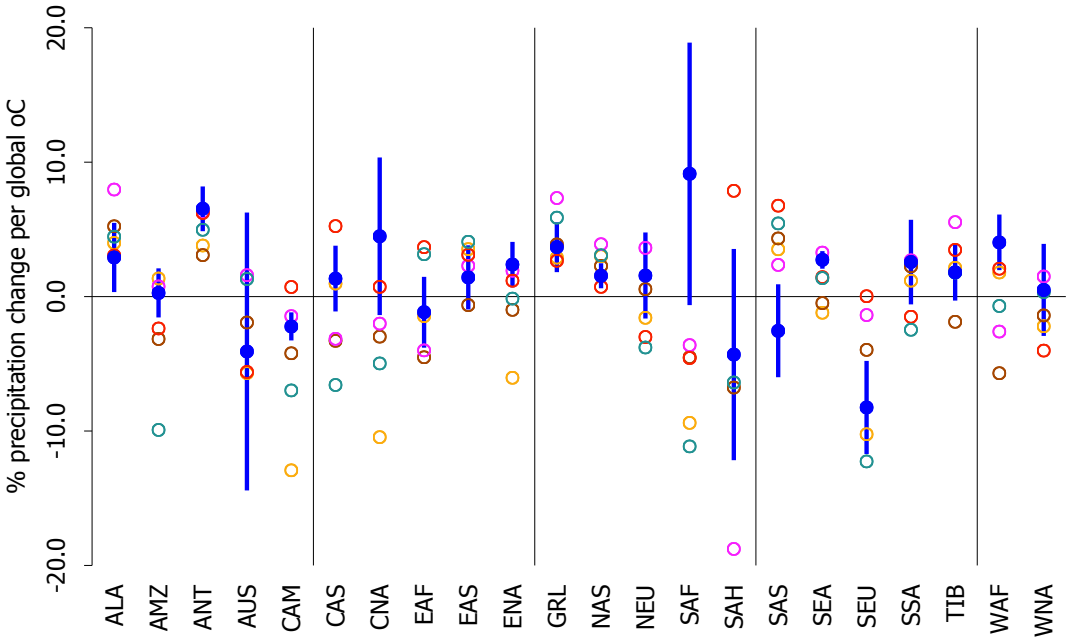
Normalised temperature responses for individual regions in the Giorgi set.

---

models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	regional means from the Giorgi set (see Table 7.3 for labels)
patterns:	calculated similarly to Figures 7.5 and 7.7, for CCSR-NIES (O), CGCM1 (O), CSIRO2 (O), ECHam4 (O), HadCM3 (O)
8, □:	pop. mean (●), sd of HadCM2 Ga 1,2,3,4 patterns (Figure 7.7)

95% interval: **limits** from 8 – 2□ to 8 + 2□

Figure 7.16



Normalised % precipitation responses for individual regions in the Giorgi set.

---

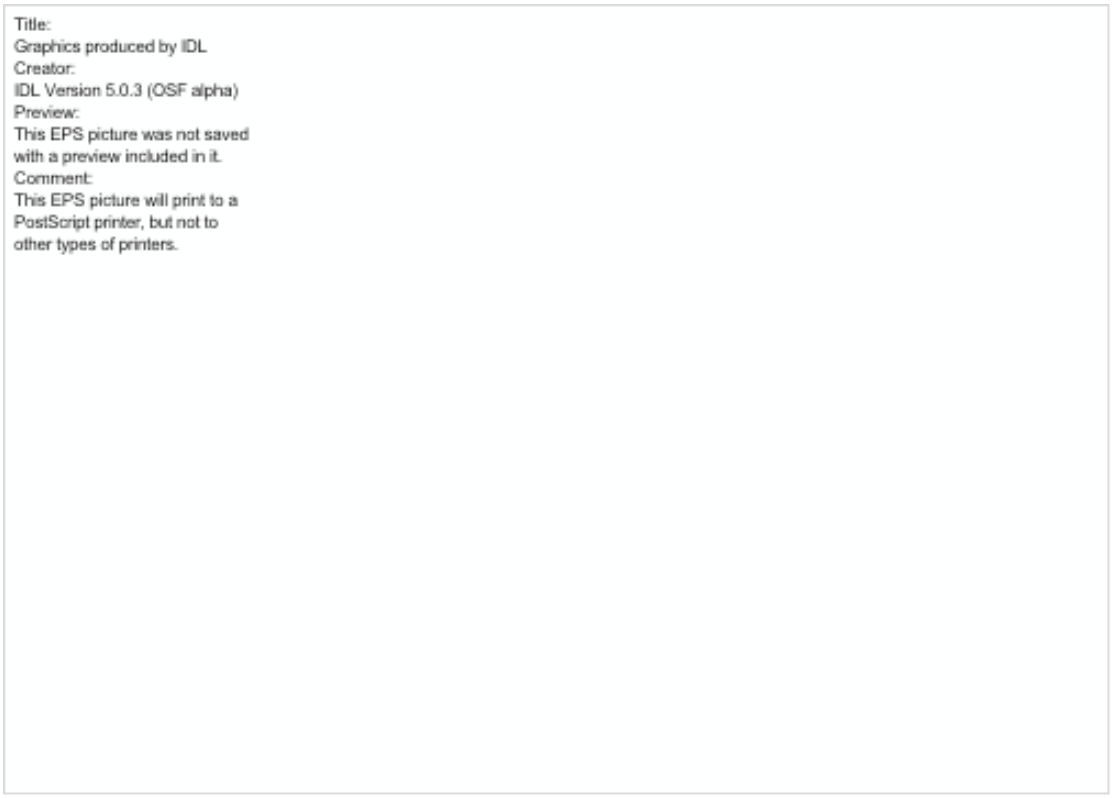
models:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models), Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (%), base 1960-1989
spatial:	regional means from the Giorgi set (see Table 7.3 for labels)
patterns:	calculated similarly to Figures 7.6 and 7.8, for CCSR-NIES (O), CGCM1 (O), CSIRO2 (O), ECHam4 (O), HadCM3 (O)
8, □:	pop. mean (●), sd of HadCM2 Ga 1,2,3,4 patterns (as Figure 7.8)

95% interval: **limits** from 8 – 2□ to 8 + 2□

Figure 7.17

Estimate of ECHam4 2080s temperatures by scaling ECHam4 pattern.

modelled estimated



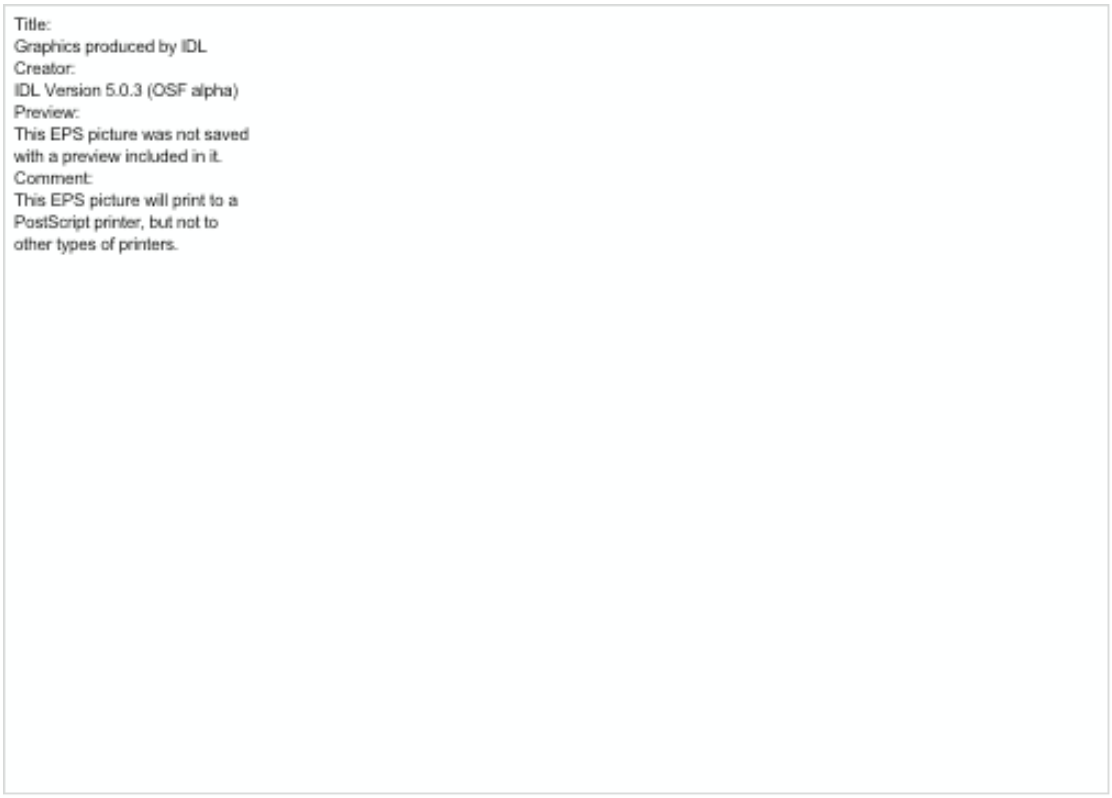
error fractional error

model:	ECHam4
simulations:	GG
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	grid-box values
modelled:	mean anomaly for the 2080s (2070-2099, **top left**)
pattern:	as calculated in Figure 7.5
estimated:	pattern * scaler in 2080s in ECHam4 GG (**top right**)
error:	estimated – modelled (**bottom left**)
fract. error:	absolute value of: error / modelled (**bottom right**)

Figure 7.18

Estimate of ECHam4 2080s precipitation by scaling ECHam4 pattern.

modelled | estimated



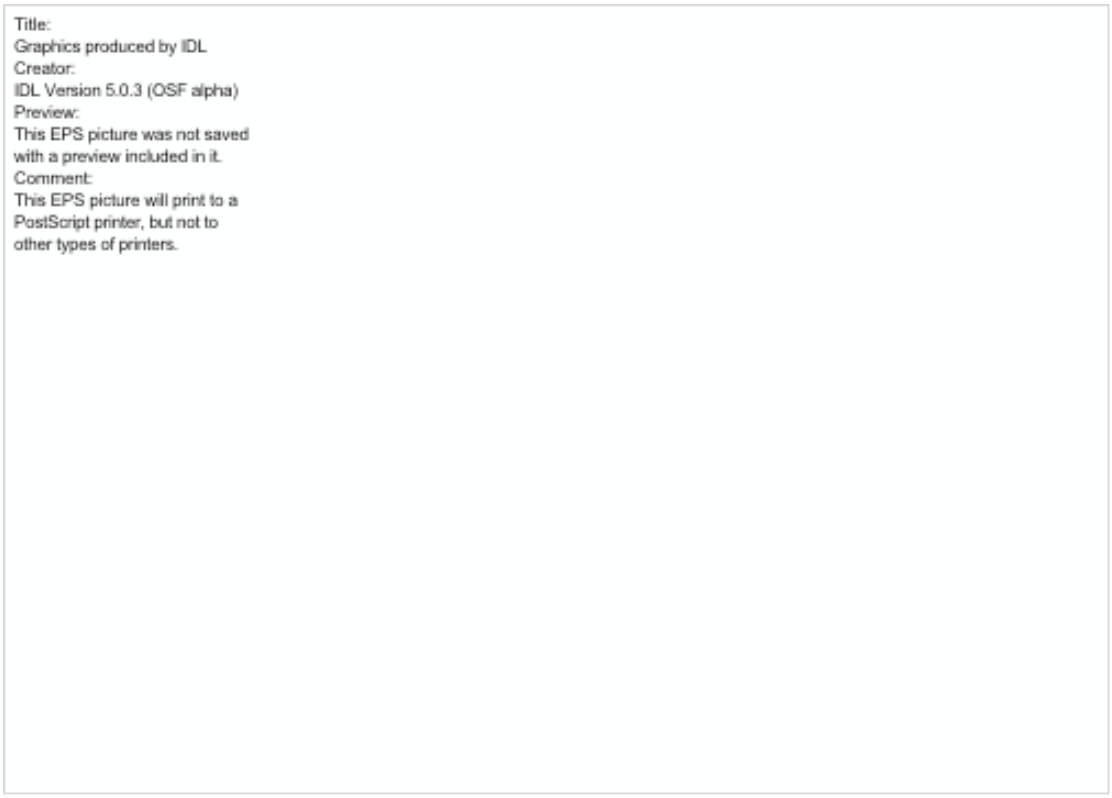
error | fractional error

model:	ECHam4
simulations:	GG
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	grid-box values
modelled:	mean anomaly for the 2080s (2070-2099, **top left**)
pattern:	as calculated in Figure 7.5
estimated:	pattern * scaler in 2080s in ECHam4 GG (**top right**)
error:	estimated – modelled (**bottom left**)
fract. error:	absolute value of: error / modelled (**bottom right**)

Figure 7.19

Estimate of ECHam4 2080s temperatures by scaling ECHam4 pattern.

modelled | estimated



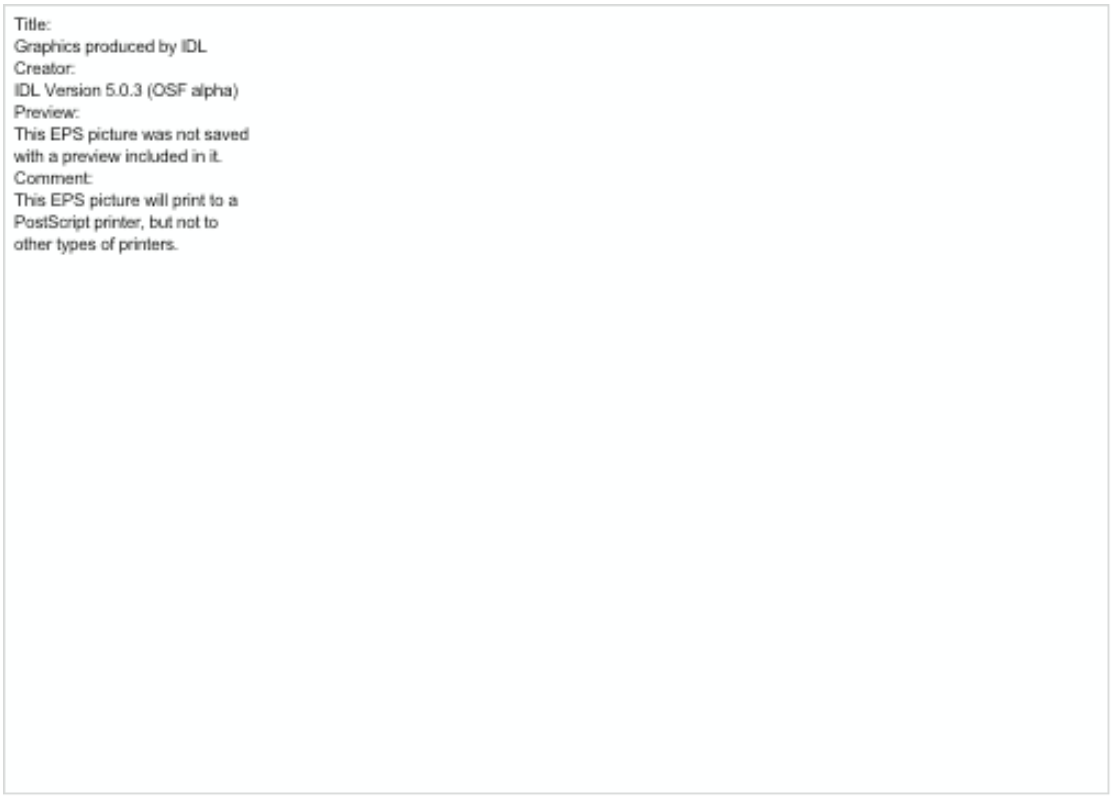
error | fractional error

model:	ECHam4
simulations:	GG
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	regional means from the Countries set
modelled:	mean anomaly for the 2080s (2070-2099, **top left**)
pattern:	similar to that calculated in Figure 7.5
estimated:	pattern * scaler in 2080s in ECHam4 GG (**top right**)
error:	estimated – modelled (**bottom left**)
fract. error:	absolute value of: error / modelled (**bottom right**)

Figure 7.20

Estimate of ECHam4 2080s precipitation by scaling ECHam4 pattern.

modelled estimated



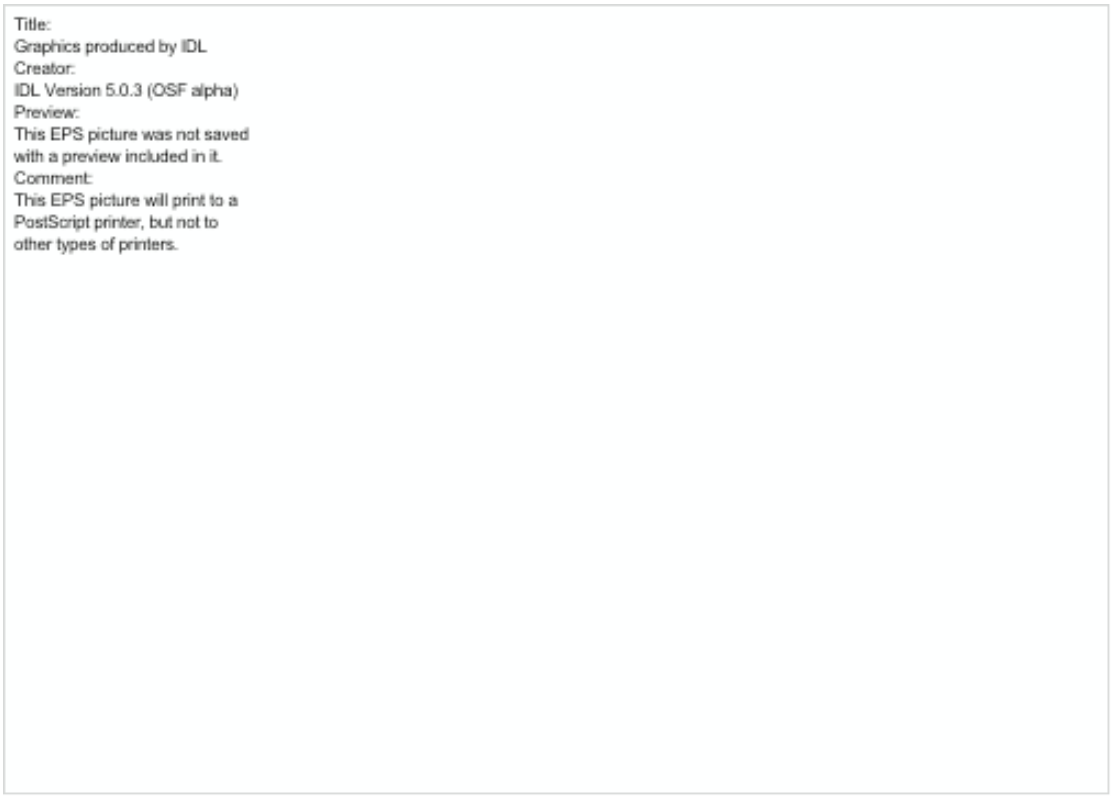
error fractional error

model:	ECHam4
simulations:	GG
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	regional means from the Countries set
modelled:	mean anomaly for the 2080s (2070-2099, **top left**)
pattern:	similar to that calculated in Figure 7.6
estimated:	pattern * scaler in 2080s in ECHam4 GG (**top right**)
error:	estimated – modelled (**bottom left**)
fract. error:	absolute value of: error / modelled (**bottom right**)

Figure 7.21

Errors in estimating 2080s temperatures by scaling model patterns.

CCSR-NIES CGCM1



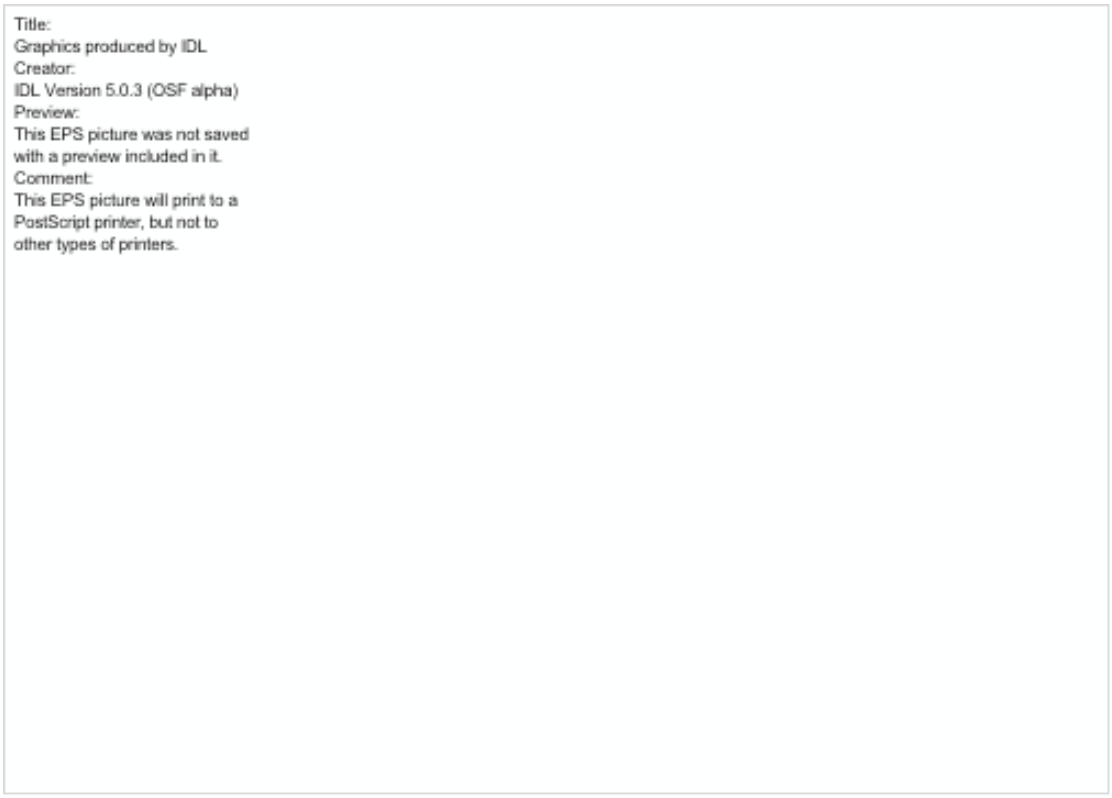
CSIRO2 HadCM3

model:	CCSR-NIES, CGCM1, CSIRO2, HadCM3
simulations:	GG
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	grid-box values
modelled:	mean anomaly for the 2080s (2070-2099)
patterns:	for each model, as calculated in Figures 7.5 and 7.7
estimated:	for each model, pattern * scaler in 2080s
error:	estimated – modelled (**plotted**)

Figure 7.22

Errors in estimating 2080s precipitation by scaling model patterns.

CCSR-NIES CGCM1



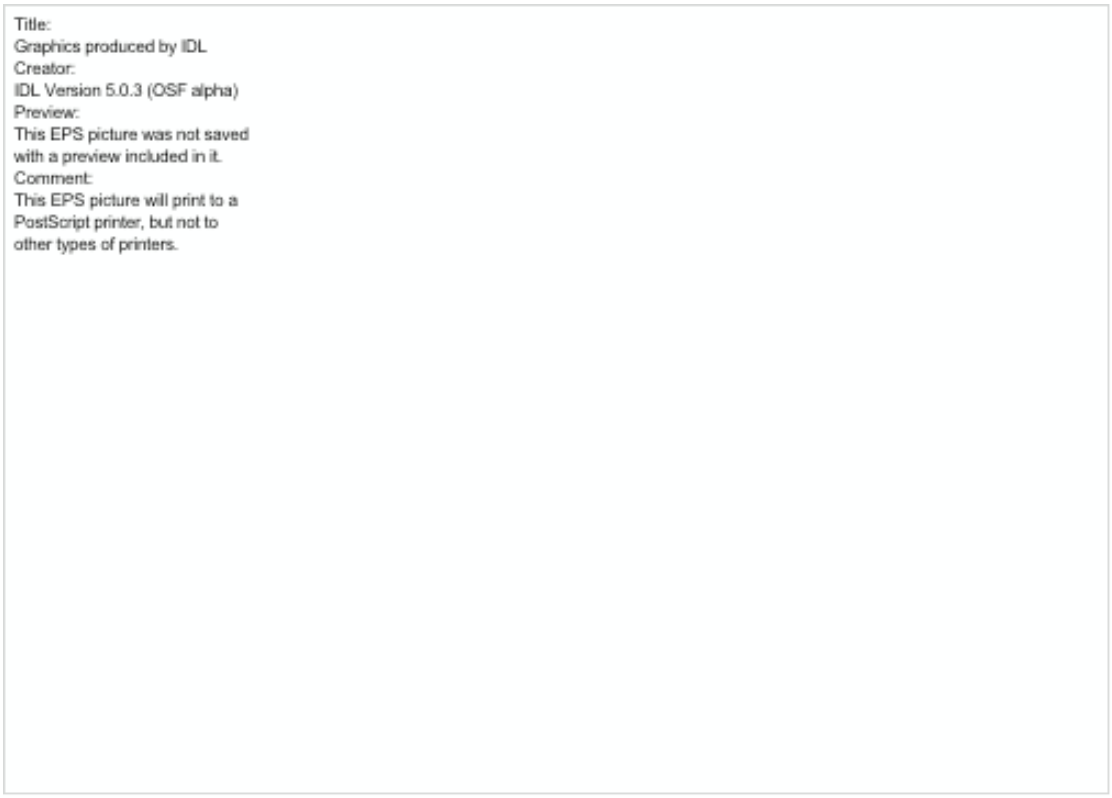
CSIRO2 HadCM3

model:	CCSR-NIES, CGCM1, CSIRO2, HadCM3
simulations:	GG
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	grid-box values
modelled:	mean anomaly for the 2080s (2070-2099)
patterns:	for each model, as calculated in Figures 7.6 and 7.8
estimated:	for each model, pattern * scaler in 2080s
error:	estimated – modelled (**plotted**)

Figure 7.23

Errors in estimating 2080s temperatures by scaling model patterns.

DDC minimum error HadCM2 error: 8+2□



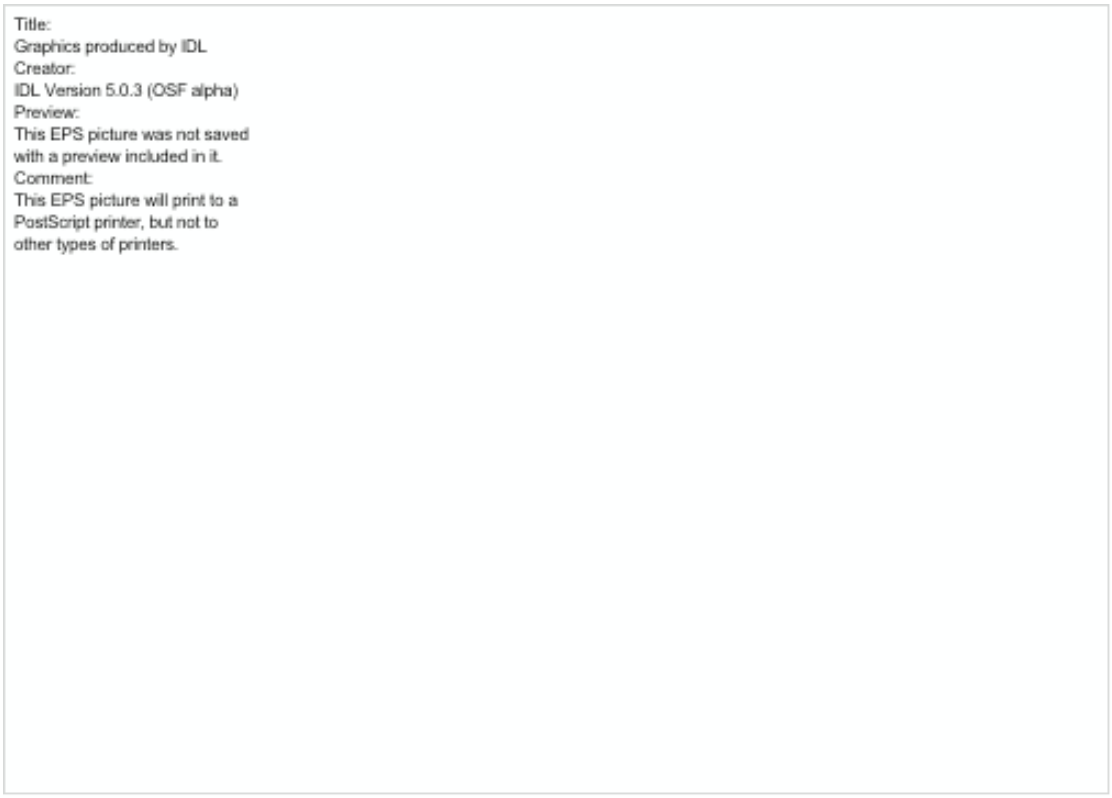
HadCM2 error: 8–2□
DDC maximum error

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	grid-box values
modelled:	for each simulation, mean anomaly for the 2080s (2070-2099)
patterns:	for each simulation, as calculated in Figures 7.5 and 7.7
estimated:	for each simulation, pattern * scaler in 2080s
error:	for each simulation, estimated – modelled
DDC range:	min (**top left**) & max (**top right**) of errors from the DDC sample
Had2 error:	mean (8) and sd (□) of errors from the HadCM2 sample
Had2 95%:	min (**bottom left**) & max (**bottom right**) of 95% confid. interval

Figure 7.24

Errors in estimating 2080s precipitation by scaling model patterns.

DDC minimum error HadCM2 error: 8+2□



HadCM2 error: 8–2□
DDC maximum error

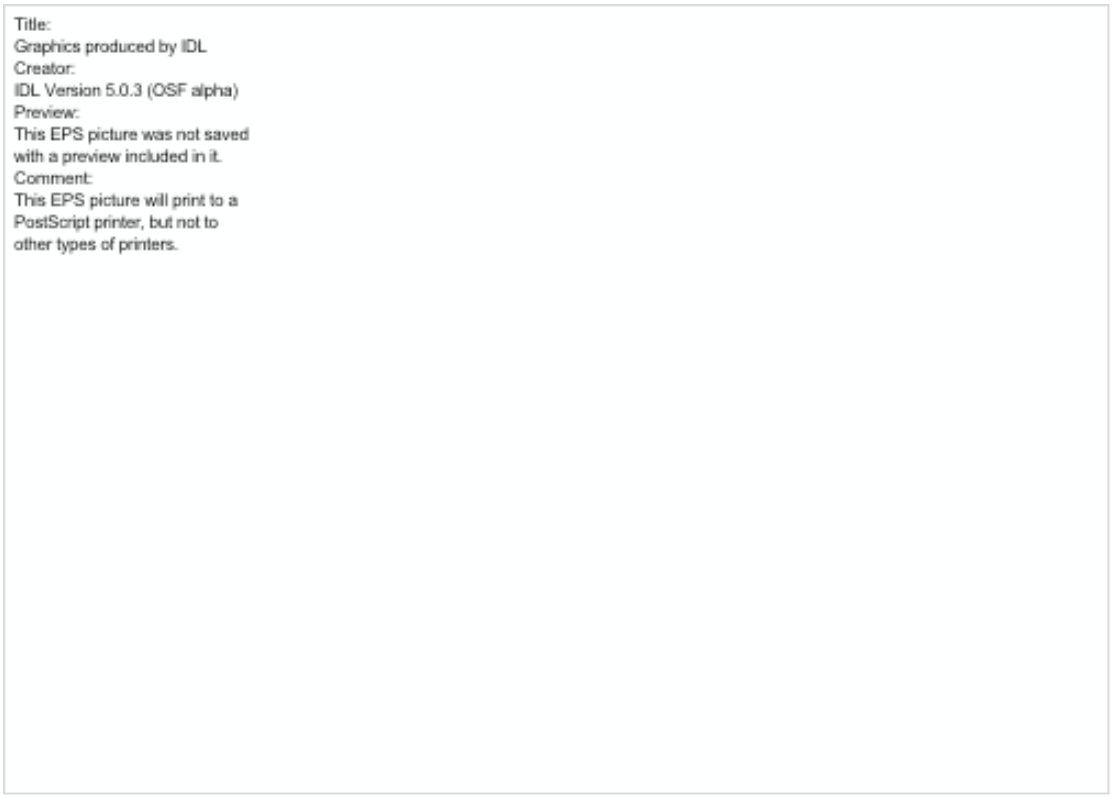
model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	grid-box values
modelled:	for each simulation, mean anomaly for the 2080s (2070-2099)
patterns:	for each simulation, as calculated in Figures 7.6 and 7.8
estimated:	for each simulation, pattern * scaler in 2080s
error:	for each simulation, estimated – modelled
DDC range:	min (**top left**) & max (**top right**) of errors from the DDC sample
Had2 error:	mean (8) and sd (□) of errors from the HadCM2 sample
Had2 95%:	min (**bottom left**) & max (**bottom right**) of 95% confid. interval

Figure 7.25

Errors in estimating 2080s temperature by scaling model patterns.

DDC error range

Had2 error 95% interval



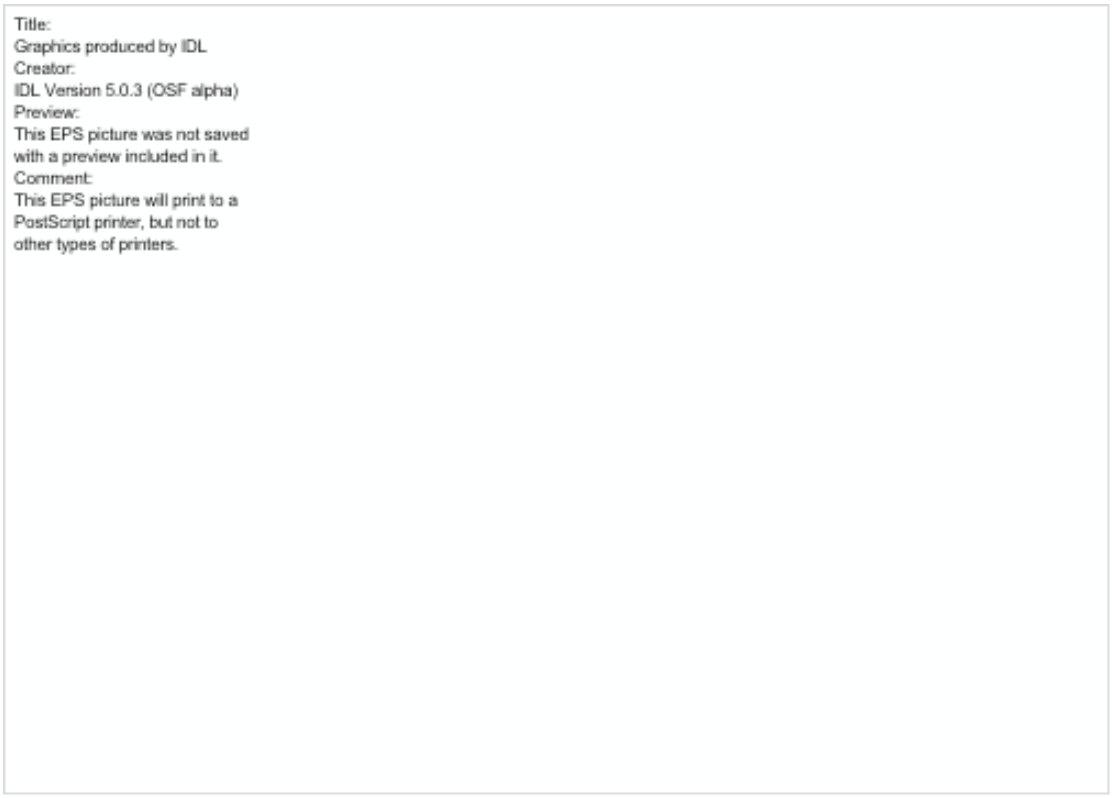
top left / top right

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	grid-box values
modelled:	for each simulation, mean anomaly for the 2080s (2070-2099)
patterns:	for each simulation, as calculated in Figures 7.5 and 7.7
estimated:	for each simulation, pattern * scaler in 2080s
error:	for each simulation, estimated – modelled
DDC range:	max – min of errors from the DDC sample (**top left**)
Had2 error:	mean (8) and sd (□) of errors from the HadCM2 sample
Had2 95%:	range of 95% confid. interval for mean error (**top right**)
ratio:	DDC range / Had2 95% (**bottom left**)

Figure 7.26

Errors in estimating 2080s precipitation by scaling model patterns.

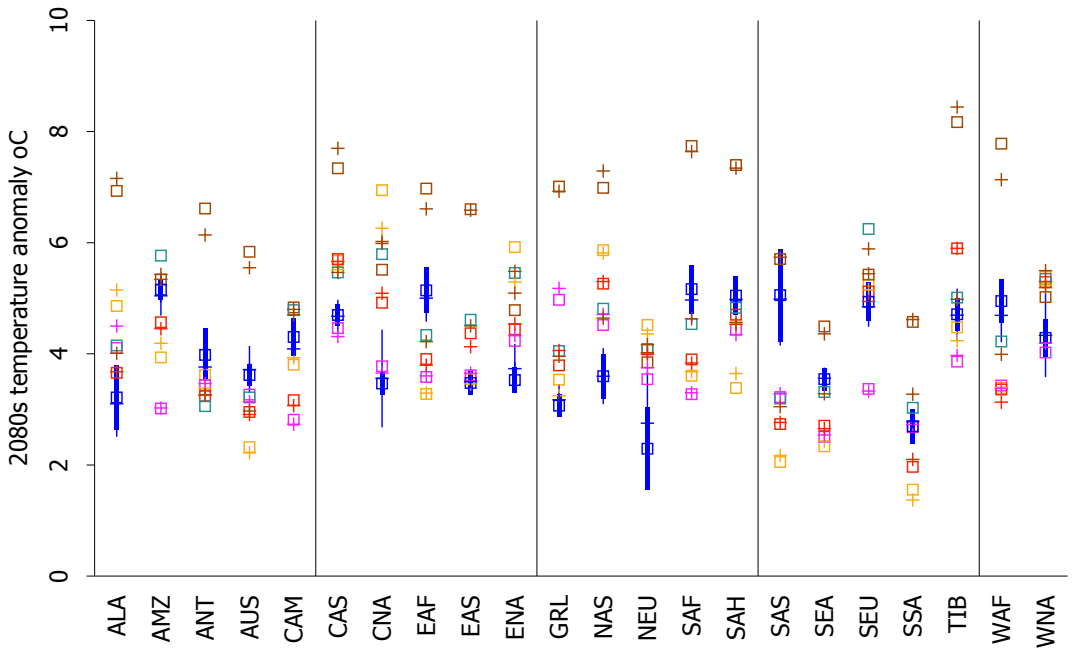
DDC error range Had2 error 95% interval



top left / top right

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	grid-box values
modelled:	for each simulation, mean anomaly for the 2080s (2070-2099)
patterns:	for each simulation, as calculated in Figures 7.6 and 7.8
estimated:	for each simulation, pattern * scaler in 2080s
error:	for each simulation, estimated – modelled
DDC range:	max – min of errors from the DDC sample (**top left**)
Had2 error:	mean (8) and sd (□) of errors from the HadCM2 sample
Had2 95%:	range of 95% confid. interval for mean error (**top right**)
ratio:	DDC range / Had2 95% (**bottom left**)

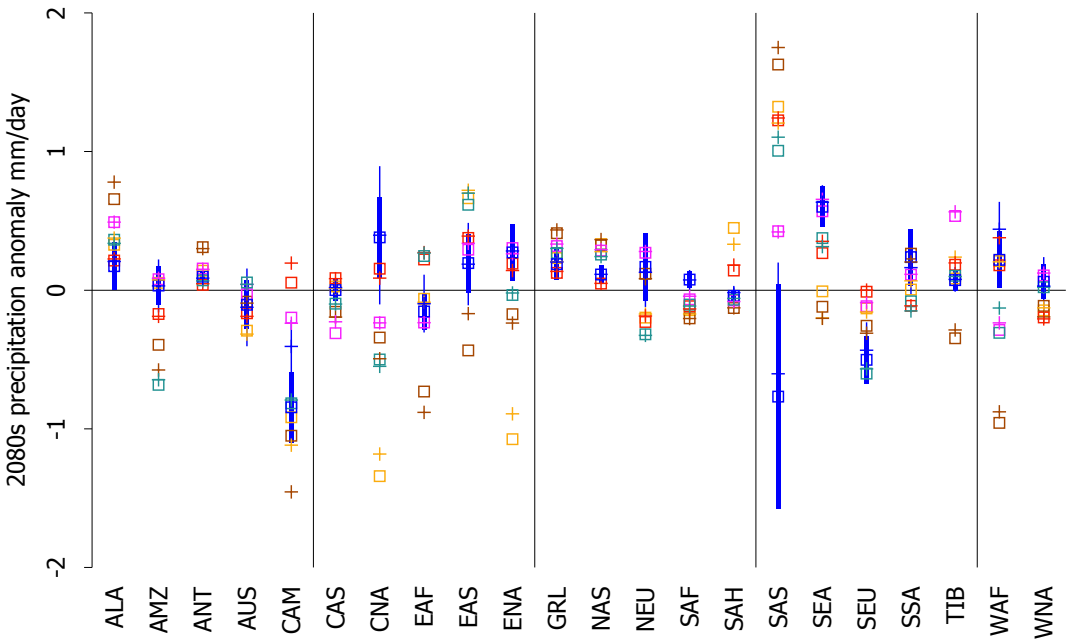
Figure 7.27



Estimating 2080s temperature by scaling model patterns.

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	regional means from Giorgi set
modelled:	mean anomaly for the 2080s: CCSR-NIES (□), CGCM1 (□), CSIRO2 (□), ECHam4 (□), HadCM3 (□), HadCM2 mean (□)
Had2 95%:	range of 95% confidence interval for mean anomaly (thick bar)
patterns:	for each simulation, similar to those in Figures 7.5 and 7.7
estimated:	pattern * scaler in 2080s: CCSR-NIES (✚), CGCM1 (✚), CSIRO2 (✚), ECHam4 (✚), HadCM3 (✚), HadCM2 mean (✚)
Had2 95%:	range of 95% confidence interval for mean estimate (thin line)

Figure 7.28

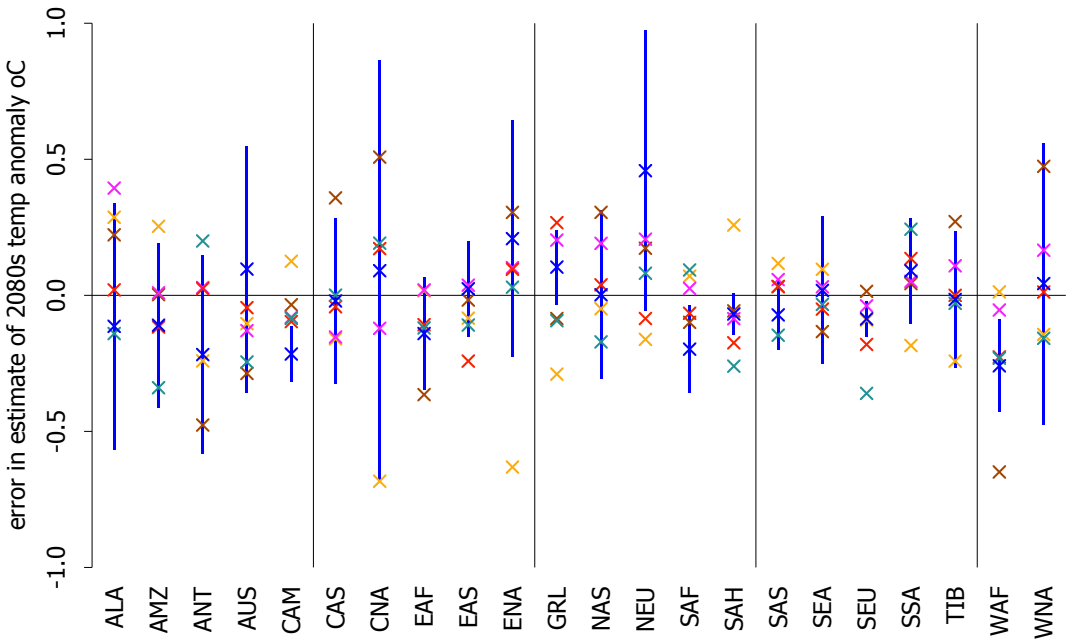


Estimating 2080s precipitation by scaling model patterns.

---

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	regional means from Giorgi set
modelled:	mean anomaly for the 2080s: CCSR-NIES (□), CGCM1 (□), CSIRO2 (□), ECHam4 (□), HadCM3 (□), HadCM2 mean (□)
Had2 95%:	range of 95% confidence interval for mean anomaly (thick bar)
patterns:	for each simulation, similar to those in Figures 7.6 and 7.8
estimated:	pattern * scaler in 2080s: CCSR-NIES (✚), CGCM1 (✚), CSIRO2 (✚), ECHam4 (✚), HadCM3 (✚), HadCM2 mean (✚)
Had2 95%:	range of 95% confidence interval for mean estimate (thin line)

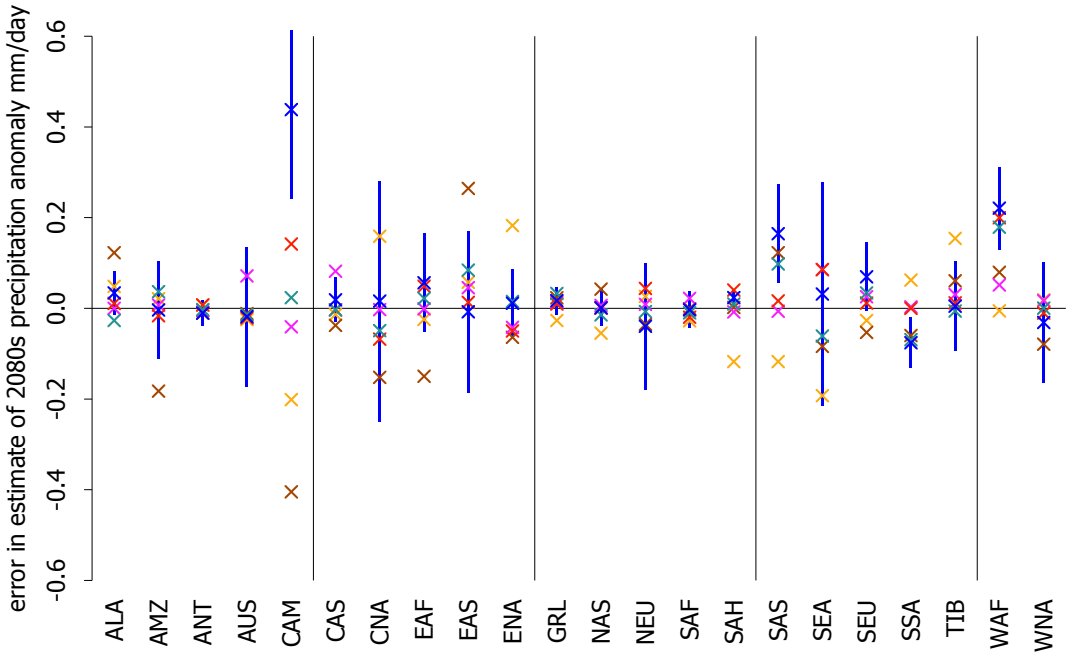
Figure 7.29



Errors in estimating 2080s temperature by scaling model patterns.

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (°C), base 1960-1989
spatial:	regional means from Giorgi set
modelled:	for each simulation: mean anomaly for the 2080s
patterns:	for each simulation, similar to those in Figures 7.5 and 7.7
estimated:	for each simulation, pattern * scaler in 2080s:
error:	estimated – modelled: CCSR-NIES (×), CGCM1 (×), CSIRO2 (×), ECHam4 (×), HadCM3 (×), HadCM2 mean (×)
Had2 95%:	range of 95% confidence interval for mean error (thick bar)

Figure 7.30



Errors in estimating 2080s precipitation by scaling model patterns.

---

model:	CCSR-NIES, CGCM1, CSIRO2, ECHam4, HadCM3, HadCM2
simulations:	GG (first five models); Ga 1,2,3,4 (HadCM2)
temporal:	JJA, 30y means 1990-2099, anomalised (mm/d), base 1960-1989
spatial:	regional means from Giorgi set
modelled:	for each simulation: mean anomaly for the 2080s
patterns:	for each simulation, similar to those in Figures 7.6 and 7.8
estimated:	for each simulation, pattern * scaler in 2080s:
error:	estimated – modelled: CCSR-NIES (×), CGCM1 (×), CSIRO2 (×), ECHam4 (×), HadCM3 (×), HadCM2 mean (×)
Had2 95%:	range of 95% confidence interval for mean error (thick bar)

---

8. Conclusions

8.1 Introduction

In this final chapter we draw together our conclusions. Our analyses have centred on the technique of pattern scaling. In chapter 2 we reviewed the historical development of climate assessments and discussed how pattern scaling has become popular in that context. In chapter 3 we gave a more detailed description of pattern scaling and we explained the reasons for our particular choices of data for evaluating the technique. Using these data we investigated four important questions concerning the use of pattern scaling that have not been dealt with fully – or at all – in the literature. We asked whether we might apply pattern scaling to:

- the multi-decadal mean? (chapter 4)
- the inter-annual probability distribution? (chapter 5)
- regions? (chapter 6)
- a variety of models? (chapter 7)

We draw together our answers to these questions in sections 8.2-8.5. We discuss some of the limitations to our study in section 8.6. In section 8.7 we present some ways in which our work might be extended. We also recognise the importance of pattern scaling to any probabilistic assessment of regional climate changes, so we pose some of the questions that any implementation of the probabilistic approach must address. We summarise our conclusions and the thesis in section 8.8.

8.2 May we use pattern scaling?

In chapter 4 we examined whether it is possible to scale a response pattern by annual global-mean temperature to estimate changes in multi-decadal means on regional scales for forcing scenarios unsimulated by a GCM. This question has been addressed before, most notably by Mitchell *et al.* (1999), but in a number of respects we have broken new ground. In answer to the question posed in the title of this section, we extend the affirmative response of Mitchell *et al.* (1999):

- from temperature to precipitation;
- from annual means to seasonal (JJA) means;
- from large-scale patterns to individual grid-boxes and regions.

We have also been innovative in assessing the errors that are introduced through pattern scaling. We outlined four potential sources of errors:

(a) internal variability;
(b) the non-applicability of the spatial response under greenhouse warming to the pre-industrial world;
(c) non-linear global-mean responses to radiative forcing;
(d) non-linear grid-box responses to radiative forcing.

We found that internal variability - (a) in the list above - made a comparatively small contribution to the errors introduced by pattern scaling; when combined, sources (a) and (b) introduced errors to the grid-box estimates with an average magnitude of 0.1°C or 0.06 mm/day. However, the global-mean responses are not entirely linear – source (c) – so the average error reverses polarity over the course of the 21st century. There are also non-linearities in individual grid-boxes – source (d) – so there are changes in the rate of response in over a third of grid-boxes during the 21st century; the changes in the spatial pattern of response represented by these non-linearities may double the magnitude of the errors introduced by pattern scaling, depending on the forcing scenario and the period. It is important to note that the errors are minimised when we interpolate between forcing scenarios, rather than extrapolate. When we estimate the changes in temperature and precipitation at the end of the 21st century under the Gd scenario by scaling the Ga pattern, the statistically significant errors are limited to a tenth of the grid-boxes.

It is also worth noting that not all *statistically* significant errors may be *practically* significant. Since practical significance is context-dependent we cannot make a universal assessment of it. However, we arbitrarily defined an error that exceeds a fifth of the estimated value as practically significant, and identified the proportions of grid-boxes that were simultaneously statistically and practically significant. In practice this means that we may treat the estimated changes for all other grid-boxes as accurate to within ±20%. When we scaled a Ga response pattern to estimate changes in Gd at the end of the 21st century, the errors are both statistically and practically significant in only 1.8% (temperature) or 8.9% (precipitation) of grid-boxes.

Again, the *worth* of pattern scaling is context-dependent. However, one measure of worth may be obtained by comparing the errors from pattern scaling with the errors that would be obtained were we to assume that the global-mean change occurred at every grid-box. We found that the errors from pattern scaling were far

smaller than under that assumption. On this basis we argue that even where significant errors arise from pattern scaling, for some applications it may be better to make the estimate using pattern scaling than not to estimate at all. As has been pointed out with regard to climate change in general (*e.g.* Fowler and Hennessy, 1995), unless specific guidance is given, the norm of assuming a stationary climate may prevail. In this case, the norm may prevail of assuming that the regional change will be similar to the global-mean change.

We were hampered in our assessment of the stabilisation scenarios by the lack of any ensembles for them. We did ascertain that the global-mean errors introduced by pattern scaling are larger for stabilisation scenarios than under steadily increasing concentrations. However, it is possible that there are not substantially more grid-boxes with statistically significant errors.

8.3 May we scale the distribution?

The value of pattern scaling would be considerably enhanced if it were possible to adapt the technique to estimate changes in distributional parameters other than the mean. It might even prove possible thus to estimate changes in entire probability distributions. To our knowledge pattern scaling has not been applied in this way before. We considered inter-annual probability distributions of seasonal temperature and precipitation in the perturbed ensembles.

We found that it is practical to scale a response pattern of inter-annual variance from one forcing scenario to estimate changes under another scenario. Although the errors are much larger (in proportion to the modelled value) than was the case when estimating means, they are never significant in more than a fifth of the grid-boxes for either temperature or precipitation. The response patterns for the skewness and kurtosis parameters explained so small a proportion of the variability that we did not attempt to scale them.

For temperature the skewness and kurtosis parameters were consistent with a normal distribution, and it is reasonable to assume that the distribution of inter-annual temperature is normal. Therefore if we may accurately estimate both the mean and the inter-annual variance by pattern scaling, we may accurately identify changes in the entire probability distribution. By identifying the grid-boxes where pattern scaling error introduces errors for either the mean or the variance, we were

able to state that errors – significant in both statistical and practical senses[299] – are present in no more than a fifth of the grid-boxes.

We cannot reason so effectively for precipitation because the distribution of inter-annual precipitation appears – from the skewness and kurtosis parameters – not to be normal. Therefore further consideration must be given to the distributional form most appropriate to inter-annual precipitation if we are to apply pattern scaling to the entire distribution for precipitation. However, the accuracy of estimating changes in the variance from pattern scaling – whereas there are errors for the temperature mean or variance in a fifth of the grid-boxes, the proportion for precipitation is a third – suggests that other distributional parameters might also be accurately estimated.

Since it seems possible to estimate changes in a probability distribution using pattern scaling:

(c) We may present estimated changes in probability distributions for forcing scenarios not simulated by GCMs.

(d) We may apply Monte Carlo techniques to the generation of climate 'projections' at regional scales. We may combine probabilities from different sources of uncertainty using Bayesian statistics and exploit the highly complex integrals by random sampling. Through pattern scaling we may extend the sources of uncertainty included to regional climate changes.

8.4 May we regionalise?

In chapter 6 we examined the effects on our estimates of climate change made by our choices of region. A very wide range of choices of scale and zone have been made in climate assessments in the past, yet very little assessment has been made of the effects of these choices. We broke the problem down into two components: the scale effect and the zoning effect. We constructed a number of region sets on the HadCM2 grid that differed from each other either in spatial scale, or else in the principles by which they were zoned. We compared the region sets in terms of:

- the loss of information from regionalising;
- the signal-to-noise ratio under increased radiative forcing;
- the errors in estimates made using pattern scaling.

[299] We arbitrarily define a practically significant error as occurring where the error exceeds a fifth of the simulated value.

In answer to the question posed in the section title, we may apply pattern scaling to regions – not just individual grid-boxes – with considerable accuracy. However, there are two further major conclusions:

(a) There is little evidence that we can improve climate scenarios by defining regions using climatically-relevant information.
(b) There is no optimal spatial scale for pattern scaling, in the sense that any change in spatial scale involves a trade-off.

The region set employed by the IPCC is made up of rectangles that loosely represent a range of different climates. However, our first conclusion suggests that the use of climatic information does not improve the results; similar results would be obtained if similarly-sized regions were zoned using geometric principles. Therefore we conclude that the only conceivable way by which climatically-informed zoning might increase signal-to-noise ratios – and the accuracy of pattern scaling – is to construct the regions on the basis of the response pattern itself. However, there would be a strong element of circularity in using the HadCM2 response pattern to define a region set, examining the performance of pattern scaling for that region set and pronouncing it better at approximating the response to forcing than any other region set!

We found that while we may improve the signal-to-noise ratio (representing the response to increased radiative forcing) by increasing the spatial scale, the same increase also diminishes the information available from the spatial pattern. Similarly, we may diminish the magnitudes of the errors introduced through pattern scaling by increasing the spatial scale, but the same increase also renders those errors more statistically significant. It is in this sense that we conclude that there is no optimal spatial scale for pattern scaling. However, this conclusion ought to be qualified by the greater agreement between models and with observations that we might expect to find at larger spatial scales.

8.5 May we scale different models?

In chapter 7 we addressed a third problem in assessing regional climate change. We concluded in chapter 4 that pattern scaling may be applied to HadCM2. However, since it is not sufficient to rely on a single model in estimating future climate change, we extended our examination of pattern scaling to a sample of six fully-coupled GCMs from the DDC. This sample covered a fairly wide range of representations of the climate system. Since each model only contributed a single simulation, we also provided an accompanying sample from HadCM2, which

enabled us to estimate the influence of internal variability on the differences between the DDC models.

The continental-scale features of the various model response patterns matched the features identified in IPCC reports (Mitchell *et al.*, 1990; Kattenberg *et al.*, 1996), in which inter-model differences are highlighted. However, the additional information about internal variability in our results suggests that a large proportion of the inter-model differences may be attributable to internal variability, and might not be 'disagreements' at all.

We also presented the responses in the regions employed by the IPCC. Since our patterns were normalised, the contribution to inter-model differences from their different climate sensitivities had been removed. Moreover, by including a HadCM2 ensemble we were able to estimate the contribution to inter-model differences from internal variability. As a result, the inter-model 'disagreements' appear to be much smaller than those discussed in the IPCC Second Assessment Report. We conclude that a large proportion of the inter-model differences in estimated regional climate change for a doubling of CO_2 (*e.g.* Mitchell *et al.*, 1990; Kattenberg *et al.*, 1996) may be attributable either to differing climate sensitivities or else to internal variability.

This conclusion is supported by Räisänen's (2001) exploration of the contribution of internal variability to the different regional anomalies in simulations from 15 CMIP models in which concentrations of CO_2 were doubled. He only isolated internal variability, not climate sensitivity, and he did not have access to long controls or ensembles, but nonetheless he found that of the inter-model differences in JJA, 20% of the temperature and 42% of the precipitation (percentage) anomalies might be explained by internal variability. Since Räisänen (2001) obtained similar proportions for other seasons, it seems likely that our conclusion may also apply to seasons other than JJA.

Another interesting result was obtained when we increased the spatial scale at which the response patterns were constructed. We found that any decreases in the inter-model differences were more than matched by corresponding decreases in the internal variability of HadCM2. This suggests that if we express inter-model

agreement in terms of internal variability, the inter-model differences are less dependent upon spatial scale than is commonly thought.[300]

Since only a single simulation was available from each of the DDC models, we were unable to examine the accuracy of pattern scaling as fully as for HadCM2. However, we did consider the accuracy of estimating the climate anomalies at the end of the 21st century by scaling the response pattern from the entire 21st century in the same simulation. We found that the errors were generally much smaller than the anomalies being estimated. In many regions the range of errors in the HadCM2 sample was comparable to the range in the DDC sample, suggesting that a large proportion of the errors for the various models may not be statistically significant.

In answer to the question posed in the title to this section, we conclude that we may apply pattern scaling to GCMs other than HadCM2, since within the limits of our analysis there is no evidence that applying pattern scaling to the DDC models gives less accurate estimates than applying it to HadCM2. We recognise that we have not been able to evaluate the accuracy of estimating for a forcing scenario other than that from which the pattern was extracted. We also recognise that if this were possible, the errors from pattern scaling would almost certainly increase. However, when we considered this issue for HadCM2 (in chapter 4), we did not find that the errors were so much larger that the technique became invalid. Moreover, two other factors both suggest that the conclusions drawn in chapter 4 with respect to HadCM2 may apply to the other models also:

- the broadly linear responses to forcing evident from the small size of the errors from estimating for GG,
- the similarity between the errors found for HadCM2 and the other models.

The practical importance of our conclusion may be noted in the context of the considerable uncertainties in regional climate assessments that stem from differences between models. Our analysis suggests that we may apply pattern scaling to a variety of models and obtain estimates of change from each model. It is reasonable to suggest that we might treat such estimates as equally plausible, and present all of them in an attempt to represent the uncertainty arising from differences between models.

---

[300] Note that our results only extend to the smaller spatial scales, never including hemispheric or global scales. It may still be the case that inter-model differences are smaller at global and hemispheric scales, even when we take internal variability into account.

8.6 Limitations

We recognise that there are limitations to the analysis that we have conducted and the conclusions that we have drawn. Since many of these limitations are common to more than one chapter, we present them together. Among the more important omissions in our analyses are the following:

(a) no seasons other than JJA;
(b) no variables other than precipitation and temperature;
(c) no sulphate aerosols in the forcing scenarios.

However, we do not believe that these limitations seriously compromise the conclusions we have drawn. Although we have only considered a single season, other examinations of pattern scaling have found considerable accuracy on both annual (Mitchell *et al.*, 1999) and monthly (Huntingford and Cox, 2000) time scales. Therefore we would expect to obtain similar results for seasons other than JJA.

It is a considerable advance on Mitchell *et al.* (1999) to have confirmed the accuracy of pattern scaling for precipitation, since it is a much 'harder' problem to represent the precipitation pattern than the temperature pattern. If pattern scaling may be applied to precipitation, we might expect pattern scaling to be applicable to many similarly 'difficult' variables. Indeed, Huntingford and Cox (2000) had some success with a wide range of land surface-related variables.

The lack of any sulphate aerosols in our forcing scenarios does not compromise them as much as might at first be thought.

- During the 1990s the estimates of sulphur emissions for the 21st century were greatly diminished, so that recent scenarios (IPCC, 2000) contain much lower emissions of sulphur than the IS92 scenarios (Alcamo *et al.*, 1995).
- Moreover, an important issue concerns the reference period for climate scenarios: if – as may well be true for Europe – the level of sulphate emissions has already peaked, any sulphate response pattern for the 21st century will have very little magnitude relative to present day levels.

If sulphate emissions are relatively small for much of the world, we might argue that the sulphate contribution to changes in radiative forcing is so small relative to the contribution from greenhouse gases that the large-scale effects of sulphate aerosols may be ignored. This is the argument presented by Mitchell *et al.* (1999) following their evaluation of two ensembles with differing levels of sulphate

emissions. If this argument cannot be sustained for certain parts of the world, it may be possible to:

i. break the response pattern down into greenhouse and sulphate elements (Ramaswamy and Chen, 1997),
ii. construct (perhaps region-specific) sulphate response patterns,
iii. scale both greenhouse and sulphate patterns individually, and
iv. combine them into a single climate scenario (Schlesinger *et al.*, 1997).

8.7 Ways forward

Although the limitations that we have described above do not seriously compromise our conclusions, they do offer ways in which we might further develop pattern scaling. Here we present a number of other ways forward.

1. We found in chapter 4 that many regions exhibited non-linear responses to increased radiative forcing over the course of the 21st century. Many of the errors introduced by pattern scaling stem from the assumption that the regional responses are linear. Therefore we suggest that it might be possible to modify the technique by assuming instead a quadratic response, which would enable us to incorporate many of the non-linear changes. The critical factor in the accuracy of this method might be the extra amount of information (*i.e.* ensemble members) required to construct stable estimates of the constants in the quadratic equations.

2. We also found in chapter 4 that stabilisation introduced additional differences between the response patterns obtained from different forcing scenarios. Stabilisation tended to smooth out the regional contrasts in the Ga response pattern, as recognised by Stouffer *et al.* (1989) and specifically for HadCM2 (Mitchell *et al.*, 2000; Senior and Mitchell, 2000). This effect appeared to be consistent between the stabilisation forcing scenarios and the scenarios with steadily increasing radiative forcing. Therefore we suggest that a modified form of pattern scaling might employ two parameters: we might use one parameter to estimate the unstabilised response as a function of the scaler (as we have done here), and an additional parameter to estimate the stabilising component of the response, perhaps as a function of the gradient of the scaler.

3. We examined distributional parameters for the *inter-annual* probability distribution of seasonal means in chapter 5. However, changes in variability

on time scales of weeks, days, and even hours may have greater impacts than changes on the seasonal time scales we have considered. The models have been evaluated far more for seasons than on shorter time scales, but a comparison of mid-latitude variability in HadCM2 with observations has shown promising results for a daily time scale (Osborn *et al.*, 1999b), and responses to increased radiative forcing have been found (Carnell and Senior, 1998). We suggest that the usefulness of pattern scaling would be enhanced if it were possible to accurately estimate probability distributions for much shorter time scales.

4. We briefly considered skewness and kurtosis parameters in chapter 5, but we concentrated on the variance. However, changes in extremes are very important, and Gaussian forms may be insufficient to describe the distribution tails of climate variables (*e.g.* Palutikof *et al.*, 1999; Brabson and Palutikof, 2000). Again, the usefulness of pattern scaling would be enhanced if it were possible to accurately estimate changes in the distribution tails.

5. In chapter 6 we considered the possibility of applying pattern scaling to regions made up of a number of grid-boxes aggregated together. However, climate change assessments often make use of statistical downscaling or regional climate models (RCMs) to represent climate change at smaller scales than the GCM grid-box (Kattenberg *et al.*, 1996; Wilby and Wigley, 1997). It is conceivable that statistical downscaling might be conducted on scaled patterns from GCMs in order to represent local changes under forcing scenarios for which no GCM simulations are available. An alternative approach might be to extract patterns from RCMs and scale them.

6. In chapter 7 we concluded that pattern scaling may be applied to models other than HadCM2. We suggested that we might represent a number of GCMs as equally plausible representations of the climate system, and apply pattern scaling to each of them in a comprehensive assessment of future climate change. However, in chapter 2 we pointed out that a sample of models from different modelling centres is not a random sample from the population of plausible models that could be constructed. There are plans for experiments in which parameters are varied to create a large number of models (Allen, 1999; UKMO/DETR, 2000), but these models are likely to bear

a strong resemblance to each other,[301] and may well be more similar to each other than they are to nature (Palmer, 2000). In this context we may imagine an experiment in which we would like to represent the changes that would be simulated by a model – with defined parameters – that had never been constructed. A novel use for pattern scaling might be to scale a response pattern from one model in an attempt to estimate changes in another model.

We were concerned at the start of this thesis to place it within the context of the need for a probabilistic approach to the assessment of future climate change. Pattern scaling belongs within that wider context, and there is considerable scope for further investigation of a number of issues that must be addressed if pattern scaling is to be successfully employed within the probabilistic framework.

(a) If the SRES scenarios (IPCC, 2000) are adopted as the anthropogenic emissions scenarios, how should they be treated? Should each of the 40 scenarios be treated as equi-probable? Should scenarios be included that represent 'surprises' or mitigation?

(b) How should the emissions scenarios be combined with possible future changes in solar irradiance and volcanic aerosols, which also affect the climate system (*e.g.* Stott *et al.*, 2000)?

(c) Further uncertainty is introduced when we estimate future changes in radiative forcing or in the climate sensitivity. How should we cope with that uncertainty? Should we use a single SCM to represent all changes in atmospheric chemistry, radiative forcing, and global-mean climate change? Or should we use a number of task-specific SCMs?

(d) How should we represent uncertainty over the consequent regional climate changes (assuming that pattern scaling is accurate)? Should we use a single GCM, a multi-model mean, or treat all available GCMs as equi-probable? Should we incorporate 'conceivable GCMs' (*i.e.* possible representations of the climate system) that have not been modelled?

(e) If more localised climate changes must be assessed, how should we represent uncertainty over downscaling techniques (*e.g.* Murphy, 1999)? Should we use particular statistical or dynamical downscaling techniques, a selection, or all available techniques?

8.8 Summary

[301] "Even when models have been developed to their final form at different institutions, they often have a family relationship in the form of some similar if not identical components" (Räisänen, 2001).

In this thesis we have assessed the accuracy of pattern scaling within the context of a probabilistic approach to the evaluation of future climate change. We have assessed the accuracy with which we may reconstruct the pattern of change that would be simulated by a given GCM, not the accuracy of the GCM itself. We conclude that pattern scaling does enable us to accurately estimate the regional changes in seasonal temperature and precipitation that would be simulated by a given GCM under different radiative forcing. The changes to which we refer may be either a multi-decadal mean or the inter-annual variance. The spatial domain may be either a grid-box or a regional mean. Therefore we conclude that pattern scaling facilitates a probabilistic approach to the assessment of future regional climate changes, and the confirmation of the accuracy of pattern scaling opens the way for this approach to be fully implemented.

References

Alcamo, J., Bouwman, A., Edmonds, J., Grubler, A., Morita, T. and Sugandhy, A. 1995: An Evaluation of the IPCC IS92 Emission Scenarios. In Houghton, J. T., Meira Filho, L. G., Bruce, J., Hoesung Lee, B. A., Callander, B. A., Haites, E., Harris, N., and Maskell, K., editors: *Climate Change 1994: Radiative forcing of Climate Change and an evaluation of the IPCC IS92 Emission Scenarios*. Cambridge University Press.

Alcamo, J., Kreileman, E., and **Leemans, R.**, 1996: Global models meet global policy. *Global Environmental Change* 6: 255-259.

Allen, M. 1999: Do-it-yourself climate prediction. *Nature* 401: 642.

Arrhenius, S. 1896a: Ueber den Einfluss des Atmosphärischen Kohlensäuregehalts auf die Temperatur der Erdoberfläche. *Proceedings of the Royal Swedish Academy of Sciences* 22.

Arrhenius, S. 1896b: On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* 41, 237-276.

Barnett, T. P., Hegerl, G. C., Santer, B., and **Taylor, K.**, 1998. The potential effect of GCM uncertainties and internal atmospheric variability on anthropogenic signal detection. *Journal of Climate* 11: 659-675.

Barrow, E. M., Hulme, M., Semenov, M. A., and **Brooks, R. J.**, 2000: Climate change scenarios. In Downing, T. E., Harrison, P. A., Butterfield, R. E., and Lonsdale, K. G., (eds.) *Climate Change, Climatic Variability and Agriculture in Europe: An Integrated Assessment*. ECI, Oxford.

Boer, G. J., Flato, G., and **Ramsden, D.,** 2000: A transient climate change simulation with greenhouse gas and aerosol forcing: projected climate to the twenty-first century. *Climate Dynamics* 16: 427-450.

Bolin, B., Watson, R. T., Zinyowera, M. C., Sundararaman, N., and **Moss, R. H.,** 1998: Preface to Watson, R. T., Zinyowera, M. C., Moss, R. H., and Dokken, D. J. (editors): *The regional impacts of climate change: an assessment of vulnerability*, Cambridge University Press.

Brabson, B. B., and **Palutikof, J. P.**, 2000: Tests of the generalized pareto distribution for predicting extreme wind speeds. *Journal of Applied Meteorology* 39: 1627-1640.

Bradley, J. V., 1968. *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, NJ.

Carnell, R. E., and **Senior, C. A.**, 1998: Changes in midlatitude variability due to increasing greenhouse gases and sulphate aerosols. *Climate Dynamics* 14: 369-383.

Carter, T. R., 1996: Developing scenarios of atmosphere, weather and climate for northern regions. *Agricultural and Food Science in Finland* 5: 235-249.

Carter, T. R., 1998: Changes in the thermal growing season in Nordic countries during the past century and prospects for the future. *Agricultural and Food Science in Finland* 7: 161-179.

Carter, T. R., Porter, J. H., and **Parry, M. L.**, 1991: Climatic warming and crop potential in Europe – prospects and uncertainties. *Global Environmental Change* 1: 291-312.

Carter, T. R., Parry, M. L., Nishioka, S., and **Harasawa, H.**, 1992: *Preliminary Guidelines for Assessing Impacts of Climate Change.* Environmental Change Unit and Center for Global Environmental Research.

Carter, T. R., Saarikko, R. A., and **Niemi, K. J.**, 1996: Assessing the risks and uncertainties of regional crop potential under a changing climate in Finland. *Agricultural and Food Science in Finland* 5: 329-350.

Charney, J., 1979: *Carbon Dioxide and Climate: a scientific assessment*. National Academy of Sciences, National Academy Press, Washington DC.

Climate Change Impacts Review Group (CCIRG), 1991: *The potential effects of climate change in the United Kingdom*. Department of the Environment, HMSO, London.

Corti, S., Molteni, F., and **Palmer, T. N.**, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature* 398: 799-802.

Croley, T. E. II, 1990: Laurentian Great Lakes Double-CO_2 Climate Change Hydrological Impacts. *Climatic Change* 17: 27-47.

Cullen, M. J. P., 1993: The unified forecast/climate model. *Meteorological Magazine* 122: 81-94.

DETR, 2000: *Climate Change: the UK programme: summary*. Department of the Environment, Transport, and the Regions, London.

Edgington, E. S., 1987. *Randomization Tests* (2nd edition). Marcel Dekker, New York.

Emori, S., Nosawa, T., Abe-Ouchi, A., Numaguti, A., Kimoto, M., and **Nakajima, T.**, 1999: Coupled ocean-atmosphere model experiments of future climate change with an explicit representation of sulfate aerosol scattering. *Journal of the Meteorological Society of Japan* 77: 1299-1307.

Epstein, E. S., 1985: *Statistical inference and prediction in climatology: a Bayesian approach*. American Meteorological Society, Boston, MA.

Fisher, R. A., 1935. *Design of Experiments*. Oliver and Boyd, Edinburgh.

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. *Climate Dynamics* 16: 451-467.

Fotheringham, A. S., and **Wong, D. W.-S.**, 1991: The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 27: 715-729.

Fotheringham, A. S., Brunsdon, C., and **Charlton, M.**, 2000: *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London.

Fowler, A. M., and **Hennessy, K. J.**, 1995: Potential impacts of global warming on the frequency and magnitude of heavy precipitation. *Natural Hazards* 11: 283-303.

Francis, A., 1979: *Advanced level statistics*. Stanley Thornes, Cheltenham.

Gaffen, D. J., Barnett, T. P., and **Elliott, W. P.**, 1991: Space and time scales of global tropospheric moisture. *Journal of Climate* 4: 989-1008.

Gates, W. L., Mitchell, J. F. B., Boer, G. J., Cubasch, U., and **Meleshko, V. P.**, 1992: Climate Modelling, Climate Prediction and Model Validation. In *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment*. Cambridge University Press.

Giorgi, F. and **Francisco, R.**, 2000: Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dynamics* 16: 169-182.

Giorgi, F., Meehl, G. A., Kattenberg, A., Grassl, H., Mitchell, J. F. B., Stouffer, R. J., Tokioka, T., Weaver, A. J., and **Wigley, T. M. L.**, 1998: Simulation of Regional Climate Change with Global Coupled Climate Models and Regional Modeling Techniques. In Watson, R. T., Zinyowera, M. C., Moss, R. H., and Dokken, D. J. (editors): *The regional impacts of climate change: an assessment of vulnerability*. Cambridge University Press.

Glantz, S. A., 1980: Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation* 61: 1-7.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and **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. *Climate Dynamics* 16: 147-168.

Gordon, H. B., and **O'Farrell, S. P.**, 1997: Transient Climate Change in the CSIRO Coupled Model with Dynamic Sea Ice. *Monthly Weather Review* 125: 875-907.

Gregory, J. M., Mitchell, J. F. B., and **Brady, A. J.**, 1997: Summer drought in Northern Midlatitudes in a time-dependent climate experiment. *Journal of Climate* 10: 662-686.

Grübler, A., 1999: Uncertainties in social and economic projections. In Carter, T. R., Hulme, M., and Viner, D. (eds.) *Representing uncertainty in climate change scenarios and impact studies*. ECLAT Report Number 1. CRU, Norwich, UK.

Harvey, D., Gregory, J., Hoffert, M., Jain, A., Lal, M., Leemans, R., Raper, S., Wigley, T., and **de Wolde, J.**, 1997: *An introduction to simple climate models used in the IPCC Second Assessment Report*. IPCC.

Henderson-Sellers, A., 1996: Can we integrate climatic modelling and assessment? *Environmental Modelling and Assessment* 1: 59-70.

Henderson-Sellers, A., and **McGuffie, K**., 1999: Concepts of good science in climate change modelling. *Climatic Change* 42: 597-610.

Hennessy, K. J., Suppiah, R., and **Page, C. M.**, 1999: Australian rainfall changes, 1910-1995. *Australian Meteorological Magazine* 48: 1-13.

Hewitt, C., and **Mitchell, J. F. B.**, 1998: A fully coupled GCM simulation of the climate of the mid-Holocene. *Geophysical Research Letters* 25: 361-364.

Hulme, M. and **Carter, T. R.**, 1999: Representing uncertainty in climate change scenarios and impact studies. In Carter, T. R., Hulme, M., and Viner, D. (eds.) *Representing uncertainty in climate change scenarios and impact studies*. ECLAT Report Number 1. CRU, Norwich, UK.

Hulme, M., and Jenkins, G. J., 1998: *Climate change scenarios for the UK: scientific report*. UKCIP Technical report number 1, CRU, Norwich.

Hulme, M., Osborn, T. J., and **Johns, T. C.,** 1998: Precipitation sensitivity to global warming: comparison of observations with HadCM2 simulations. *Geophysical Research Letters* 25(17): 3379-3382.

Hulme, M., Arntzen, J., Downing, T., Leemans, R., Malclom, J., Reynard, N., Ringrose, S., and **Rogers, D.**, 1996: *Climate Change and Southern Africa: an exploration of some potential impacts and implications for the SADC region*. CRU/WWF.

Hulme, M., Wigley, T. M. L., Barrow, E. M., Raper, S. C. B., Centella, A., Smith, C., and **Chipanshi, A. C.**, 2000: *Using a Climate Scenario Generator*

for Vulnerability and Adaptation Assessments: MAGICC and SCENGEN Version 2.4 Workbook. Climatic Research Unit, Norwich.

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

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation regional temperatures and precipitation. *Science* 269: 676-679.

IPCC, 1996a: Policymakers' Summary. In Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. and Maskell, K., editors, *Climate Change 1995: the Science of Climate Change*. Cambridge University Press.

IPCC, 1996b: Technical Summary. In Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. and Maskell, K., editors, *Climate Change 1995: the Science of Climate Change*. Cambridge University Press.

IPCC, 1996c: Summary for Policymakers. In Watson, R. T., Zinyowera, M. C., Moss, R. H., and Dokken, D. J., (eds.): *Climate Change 1995: impacts, adaptations and mitigation of climate change: scientific-technical analyses*. Cambridge University Press.

IPCC, 1996d: Appendix 5: Acronyms. In Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. and Maskell, K., editors, *Climate Change 1995: the Science of Climate Change*. Cambridge University Press.

IPCC, 2000: Summary for Policymakers. In Nakicenovic, N., and Swart, R., (eds.): *Emissions Scenarios*. Cambridge University Press.

IPCC, 2001: *Summary for Policymakers* . Published on the internet prior to the publication of the full WGI Third Assessment Report. http://www.meto.gov.uk/sec5/CR_div/ipcc/wg1/WGI-SPM.pdf

IPCC-TGCIA, 1999: *Guidelines on the use of scenario data for climate impact and adaptation assessment*. Version 1. Prepared by Carter, T. R., Hulme, M., and Lal, M.. IPCC Task Group on Scenarios for Climate Impact Assessment.

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, spinup, and validation. *Climate Dynamics* 13: 103-134.

Jonas, M., Fleischmann, K., Ganopolski, A. V., Krabec, J., Sauer, U., Olendrznski, K., Petoukhov, V. K., and **Shaw, R. W.**, 1996: A fast

turnaround: combining the results of IMAGE and a GCM. *Climatic Change* 34: 479-512.

Jones, R. N., 2000: Analysing the risk of climate change using an irrigation demand model. *Climate Research* 14: 89-100.

Joubert, A. M. and **Hewitson, B. C.** 1997: Simulating present and future climates of southern Africa using general circulation models. *Progress in Physical Geography* 21, 51-78.

Karl, T. R., and **Knight, R. W.**, 1998: Secular trends of precipitation amount, frequency, and intensity in the USA. *Bulletin of the American Meteorological Society* 79: 231-242.

Kattenberg, A., Giorgi, F., Grassl, H., Meehl, G. A., Mitchell, J. F. B., Stouffer, R. J., Tokioka, T., Weaver, A. J., and **Wigley, T. M. L.,** 1996: Climate Models - Projections of Future Climate. In Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. and Maskell, K., editors, *Climate Change 1995: the Science of Climate Change*. Cambridge University Press.

Katz, R. W., 1999: Techniques for estimating uncertainty in climate change scenarios and impact studies. In Carter, T. R., Hulme, M., and Viner, D. (eds.) *Representing uncertainty in climate change scenarios and impact studies*. ECLAT Report Number 1. CRU, Norwich, UK.

Katz, R. W., and **Brown, B. G.**, 1992: Extreme events in a changing climate: variability is more important than averages. *Climatic Change* 21: 289-302.

Kellogg, W. W., 1987: Mankind's impact on climate: the evolution of an awareness. *Climatic Change* 10: 113-136.

Kittel, T. G. F., Giorgi, F., and **Meehl, G. A.**, 1998: Intercomparison of regional biases and doubled CO_2 sensitivity of coupled atmosphere-ocean general circulation model experiments. *Climate Dynamics* 14: 1-15.

Kuusisto, E., **Kauppi, L.**, and **Heikinheimo, P.** (editors) 1996: *Ilmastonmuutos ja Suomi* (Climate Change and Finland). Helsinki University Press.

Lambert, S. J., and **Boer, G. J.**, 2000: CMIP1 Evaluation and Intercomparison of Coupled Climate Models. submitted

Leemans, R., 1999: Uncertainties in integrated assessment models. In Carter, T. R., Hulme, M., and Viner, D. (eds.) *Representing uncertainty in climate change scenarios and impact studies*. ECLAT Report Number 1. CRU, Norwich, UK.

Livezey, R. E., 1995: Field intercomparison. In von Storch, H., and Navarra, A. (eds.): *Analysis of climate variability*. Springer, Berlin.

Lorenz, E. N. 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20, 130-141.

McCracken, M. C., and **Luther, F. M.**, 1985: Executive Summary. In McCracken, M. C., and Luther, F. M. (eds.) *Projecting the Climatic Effects of Increasing Carbon Dioxide*. US Department of Energy. Washington DC.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and **Stouffer, R. J.**, 1997: Intercomparison makes for a better climate model. *EOS, Transactions of the American Geophysical Union* 78: 445-446, 451.

Mitchell, J. F. B., and **Gregory, J. M.**, 1992: Climatic consequences of emissions and a comparison of IS92a and SA90. In Houghton, J. T., Callander, B. A., and Varney, S. K. (eds.): *Climate Change 1992: Supplement to the IPCC scientific assessment*. Cambridge University Press.

Mitchell, J. F. B., and **Johns, T. C.,** 1997: On modification of global warming by sulfate aerosols. *Journal of Climate* 10: 245-267.

Mitchell, J. F. B., Manabe, S., Tokioka, T., and **Meleshko, V.**, 1990: Equilibrium Climate Change. In Houghton, J. T., Jenkins, G. J., and Ephraums, J. J., 1990: *Climate Change: The IPCC Scientific Assessment*. Cambridge University Press.

Mitchell, J. F. B., Johns, T. C., and **Senior, C. A.**, 1998: Transient response to increasing greenhouse gases using models with and without flux adjustment. *Hadley Centre Technical Note* 2. UK Meteorological Office.

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

Mitchell, J. F. B., Johns, T. C., Ingram, W. J., and **Lowe, J. A.**, 2000: The effect of stabilising atmospheric carbon dioxide concentrations on global and regional climate change. *Geophysical Research Letters* 27: 2977-2980.

Mitchell, T. D. and **Hulme, M.**, 1999: Predicting regional climate change: living with uncertainty. *Progress in Physical Geography* 23: 57-78.

Mortsch, L., 2000: Methods – 'Artificial Scenarios'. In Beersma, J., Agnew, M., Viner, D., and Hulme, M., (eds.): *Climate scenarios for water-related and coastal impacts*. ECLAT-2 Workshop Report Number 3.

Moss, R. H., 2000: Ready for IPCC-2001: Innovation and change in plans for the IPCC Third Assessment Report. *Climatic Change* 45: 459-468.

Murphy, J. M., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *Journal of Climate* 12: 2256-2284.

Murphy, J. M., and **Mitchell, J. F. B.**, 1995: Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide.

Part II: spatial and temporal structure of response. *Journal of Climate* 8: 57-80.

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

Nicholls, N., Gruza, G. V., Jouzel, J., Karl, T. R., Ogallo, L. A., and **Parker, D. E.**, 1996: Observed climate variability and change. In Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. and Maskell, K., editors, *Climate Change 1995: the Science of Climate Change*. Cambridge University Press.

Obasi, **G. O. P.** and **Dowdeswell**, **E.** 1998: Foreword to Watson, R. T., Zinyowera, M. C., Moss, R. H., and Dokken, D. J. (editors): *The regional impacts of climate change: an assessment of vulnerability*. Cambridge University Press.

Oglesby, R. J., and **Saltzman, B.**, 1992: Equilibrium climate statistics of a GCM as a function of atmospheric carbon dioxide. Part I: geographic distributions of primary variables. *Journal of Climate* 5: 66-92.

Openshaw, S., 1984: *The Modifiable Areal Unit Problem*. CATMOG 38. Norwich.

Openshaw, S., and **Rao, L.**, 1995: Algorithms for Re-engineering 1991 Census Geography. *Environment and Planning A* 27: 425-426.

Osborn, T. J., Briffa, K. R., Tett, S. F. B., Jones, P. D., and **Trigo, R. M.**, 1999a: Evaluation of the North Atlantic Oscillation as simulated by a coupled climate model. *Climate Dynamics* 15: 685-702.

Osborn, T. J., Conway, D., Hulme, M., Gregory, J. M., and **Jones, P. D.**, 1999b: Air flow influences on local climate: observed and simulated mean relationships for the UK. *Climate Research* 13: 173-191.

Palmer, T. N., 1993: A nonlinear dynamical perspective on climate change. *Weather* 48: 314-326.

Palmer, T.N., 2000: A nonlinear dynamical perspective on model error: a proposal for nonlocal stochastic-dynamic parameterisation in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, in press.

Palutikof, J. P., Brabson, B. B., Lister, D. H., and **Adcock, S. T.**, 1999: A review of methods to calculate extreme wind speeds. *Meteorological Applications* 6: 119-132.

Phillips, T., 2000: *Summary Documentation: CMIP I Model Features and Experimental Implementation (Version 1.2)*. PCMDI, LLNL, CA, USA. http://www-pcmdi.llnl.gov/modeldoc/cmip/

Pitman, E. J. G., 1937a. Significance tests which may be applied to samples from any populations. *J. R. Statist. Soc. B.* 4: 119-130.

Pitman, E. J. G., 1937b. Significance tests which may be applied to samples from any populations. II. The correlation coefficient. *J. R. Statist. Soc. B.* 4: 225-232.

Pitman, E. J. G., 1937c. Significance tests which may be applied to samples from any populations. III. The analysis of variance test. *Biometrika,* 29: 322-335.

Pittock, A. B., 1993: Climate scenario development. In Jakeman, A. J., Beck, M. B., and McAleer, M. J. (eds.), *Modelling change in environmental systems*. Wiley, Chichester.

Prescott, J., 1999: Foreword in UK Meteorological Office / DETR: *Climate change and its impacts: Stabilisation of CO_2 in the atmosphere*. Hadley Centre.

Ramaswamy, V., and **Chen, C-T.**, 1997: Linear additivity of climate response for combined albedo and greenhouse perturbations. *Geophysical Research Letters* 24: 567-570.

Räisänen, J., 2001: CO_2-induced climate change in CMIP2 experiments: quantification of agreement and role of internal variability. *Journal of Climate*, in press.

Raper, S. C. B., and **Cubasch U.**, 1996: Emulation of the results from a coupled general circulation model using a simple climate model. *Geophysical Research Letters* 23: 1107-1110.

Raper, S. C. B., Gregory, J. M., and **Osborn, T. J.**, 2001: Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. *Climate Dynamics* submitted.

Roeckner, E., Oberhauer, J. M., Bacher, A., Christoph, M., and **Kirchner, I.**, 1996: ENSO variability and atmospheric response in a global atmosphere-ocean GCM. *Climate Dynamics* 12: 737-754.

Roeckner, E., Bengtsson, L., Feichter, J., Lelieveld, J., and **Rodhe, H.**, 1999: Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. *Journal of Climate* 12: 3004-3032.

de Ronde, J, 2000: Climate needs and availability of climate scenarios for coastal impact assessments. In Beersma, J., Agnew, M., Viner, D., and Hulme, M., (eds.): *Climate scenarios for water-related and coastal impacts*. ECLAT-2 Workshop Report Number 3.

Ropelewski, C. F. and **Halpert, M. S.**, 1987: Global and regional scale precipitation patterns associated with the El Niño / Southern Oscillation. *Monthly Weather Review* 115: 1606-1626.

Ropelewski, C. F. and **Halpert, M. S.**, 1989: Precipitation patterns associated with the High Index Phase of the Southern Oscillation. *Journal of Climate* 2: 268-284.

Rotmans, J., Hulme, M., and **Downing, T. E.**, 1994: Climate-change implications for Europe – an application of the ESCAPE model. *Global Environmental Change* 4: 97-124.

Santer, B. D., 1988: *Regional validation of GCMs*. CRU Publication No.9

Santer, B. D., and **Wigley, T. M. L.**, 1990: Regional validation of Means, Variances, and Spatial Patterns in GCM Control Runs. *Journal of Geophysical Research* 95 (D1): 829-850.

Santer, B. D., Wigley, T. M. L., Schlesinger, M. E., and **Mitchell, J. F. B.**, 1990: *Developing climate scenarios from equilibrium GCM results*. MPI Report Number 47, Hamburg.

Schlesinger, M. E., Andronova, N., Ghanem, A., Malyshev, S., Reichler, T., Rozanov, E., Wang, W., and **Yang, F.**, 1997: *Geographical scenarios of greenhouse gas and anthropogenic aerosol induced climate changes, report of the Climate Research Group*. University of Illinois at Urbana, Champaign.

Schulze, R. E. 1997: Impacts of global climate change in a hydrologically vulnerable region: challenges to South African hydrologists. *Progress in Physical Geography* 21, 113-136.

Shackley, S., Young, P., Parkinson, S., and **Wynne, B.**, 1998: Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools? *Climatic Change* 38: 159-205.

Smagorinsky, J., 1982: *Carbon Dioxide and Climate: a second assessment*. National Academy of Sciences, National Academy Press, Washington DC.

Sprent, P., 1998. *Data Driven Statistical Methods*. Chapman and Hall, London.

Stott, P. A. and **Tett, S. F. B.**, 1998: Scale-Dependent Detection of Climate Change. *Journal of Climate* 11: 3282-3294.

Stott, P. A., Tett, S. F. B., Jones, G. S., Allen, M. R., Ingram, W. J., and **Mitchell, J. F. B.**, 2000: Anthropogenic and natural causes of twentieth century temperature change. *Space Science Reviews* 94: 337-344.

Stouffer, R. J., Manabe, S., and **Bryan, K.**, 1989: Interhemispheric asymmetry in climate response to a gradual increase in CO_2. *Nature* 342: 660-662.

Tegart, W. J., Sheldon, G. W., and **Griffiths, G. C.,** (eds.) 1990: *Climate Change: the IPCC impacts assessment*. Australian Government Publication Service, Canberra.

Tett, S. F. B., Mitchell, J. F. B., Parker, D. E., and **Allen, M. R.,** 1996: Human influence on the atmospheric vertical temperature structure: detection and observations. *Science* 274: 1170-1173.

Tett, S. F. B., Johns, T. C., and **Mitchell, J. F. B.**, 1997: Global and regional variability in a coupled AOGCM. *Climate Dynamics* 13: 303-323.

Titus, J. G., and **Narayanan, V.**, 1996: The risk of sea level rise. *Climatic Change* 33: 151-212.

Tol, R. S. J., and **de Vos, A. F.**, 1998: A Bayesian statistical analysis of the enhanced greenhouse effect. *Climatic Change* 38: 87-112.

Trenberth, K. E., 1998: Atmospheric moisture residence times and cycling: implications for rainfall rates and climate change. *Climatic Change* 39: 667-694.

Trenberth, K. E., 1999: Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climatic Change* 42: 327-339.

Tyndall, J., 1863: On radiation through the earth's atmosphere. *Phil. Mag.* 4: 200-207.

UKMO / DETR, 1997: *Climate change and its impacts: a global perspective*. Hadley Centre.

UKMO / DETR, 1998: *Climate change and its impacts*. Hadley Centre.

UKMO / DETR, 1999: *Climate change and its impacts: Stabilisation of CO_2 in the atmosphere*. Hadley Centre.

UKMO / DETR, 2000: *Climate change: an update of recent research from the Hadley Centre*. Hadley Centre.

Vaughan, D. G., and **Spouge, J. R.**, 2001: Risk estimation of collapse of the West Antarctic Ice Sheet. *Climatic Change* in press.

Watson, R. T., **Zinyowera, M. C.**, **Moss, R. H.**, and **Dokken, D. J.** (editors), 1998: *The regional impacts of climate change: an assessment of vulnerability*. Cambridge University Press.

Wigley, T. M. L., and **Santer, B. D.**, 1990: Statistical Comparison of Spatial Fields in Model Validation, Perturbation, and Predictability Experiments. *Journal of Geophysical Research* 95 (D1): 851-865.

Wigley, T. M. L., Raper, S. C. B., Smith, S., and **Hulme, M.,** 2000: *The MAGICC/SCENGEN Climate Scenario Generator: Version 2.4: Technical Manual*. CRU, UEA, Norwich, UK.

Wilby, R. L. and **Wigley, T. M. L.** 1997: Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography* 21, 530-548.