REVIEW

Techniques for estimating uncertainty in climate change scenarios and impact studies

Richard W. Katz*

Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, Colorado 80307, USA

ABSTRACT: Methodology for quantifying uncertainty in global climate change studies is reviewed. The focus is on recent developments in statistics, such as hierarchical modeling and Markov chain Monte Carlo simulation techniques, that could enable more full-fledged uncertainty analyses to be performed as part of integrated assessments of climate change and its impacts. First an overview of uncertainty analysis, including its sources and how it propagates, is provided. Presently employed techniques in climate change assessments, such as sensitivity, scenario, and Monte Carlo simulation analyses, are then surveyed. Next alternative approaches, based on more formal statistical theory (especially the Bayesian statistical paradigm), are described. Finally, some tentative recommendations on strategies for achieving the goal of more reliably quantifying uncertainty in global climate change are made.

KEY WORDS: Aggregation/scaling · Bayesian statistics · Extremes · Monte Carlo simulation · Scenario analysis · Sensitivity analysis

1. INTRODUCTION

Cicero (Roman orator, 106–43 BC): 'Probabilities direct the conduct of the wise man.'

Pliny the Elder (Roman scholar, AD 23–79): 'The only certainty is uncertainty.'

Socrates (Greek philosopher, 470–399 BC): 'As for me, all I know is I know nothing.'

From these well-known quotes, it is evident that eminent scholars, long ago, had a genuine appreciation of the need to take into account uncertainty. Today, about the only thing that differs is the size and complexity of the situation, especially with respect to global problems such as climate change. Although large and complex models may well be a blessing in better approximating reality, they can be a curse when it comes to quantifying uncertainty (Cipra 2000).

In the spirit of Pliny the Elder, if any consensus about the present state of knowledge exists, it is that 'uncer-

tainties pervade all levels of a climate impact assessment' (Carter et al. 1994, p. 31). In climate modeling [especially, using general circulation models (GCMs)], 'significant reduction of key uncertainties will require a decade or more' (Mahlman 1997, p. 1416). In considering how to enhance integrated assessment methodology, 'the issue of uncertainty is … of crucial importance' (Rotmans & Dowlatabadi 1998, p. 333). From a social science perspective, 'we cannot avoid uncertainty and subjectivity in integrated assessment' (van Asselt et al. 1996, p. 88), and 'uncertainties in climate change are so pervasive and far reaching that the tools for handling uncertainty provided by decision analysis are no longer sufficient' (Jaeger et al. 1998, p. 142). Finally, Morgan et al. (1999, p. 277) questioned the 'assumption that uncertainty is modest and manageable' for global change problems.

Despite these strong statements about the importance of uncertainty, its present treatment is viewed as quite inadequate. For example, 'uncertainty analysis … is overdue in relation to GCMs' (Shackley et al. 1998, p. 180), and 'at present, apart from only a few

*E-mail: rwk@ucar.edu
**Sponsored by the National Science Foundation

exceptions, uncertainties are treated poorly in the current generation of integrated assessment models' (Rotmans & Dowlatabadi 1998, p. 370). Although the authors of the Third Assessment Report of the Intergovernmental Panel on Climate Change (Houghton et al. 2001, McCarthy et al. 2001, Metz et al. 2001) did receive guidance on dealing with uncertainty, Reilly et al. (2001, p. 430) asserted that 'much remains to be done to adequately treat uncertainty in those conclusions that are most important for policy decision-making.' Has the exercise of quantifying uncertainties not been given a high enough priority in the past or is it just too difficult to carry out in practice?

Often the term 'risk' appears when considering consequences that are uncertain, with risk sometimes being synonymous with probability, other times referring to related concepts such as expected loss (i.e. loss multiplied by probability). Although attempts have been made to distinguish between risk and uncertainty, no such distinction is drawn here. In this regard, the title of the text by Morgan & Henrion (1990), 'Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis,' is apt. The appearance of at least rudimentary applications of risk analysis to climate change impacts certainly is a welcome development (e.g. Jones 2000), but in no way diminishes the need for uncertainty analysis.

My perspective is that of a statistician with experience in applications to both climate and climate impacts. I attempt a broad review of the technology of quantitative uncertainty analysis, encompassing not just 'objective' aspects but 'subjective' ones as well. Rather than comprehensive coverage of the literature, just enough references are cited to achieve systematic treatment of the methodology employed. Only a very limited attempt is made to address qualitative issues. Moreover, recognizing the prevalent view in the social science community that the 'mainstream social science tools are presently incapable of providing a reliable basis for rational goal setting and policy implementation' (Rayner & Malone 1998, p. xxxix), no claim is made that improved quantification of uncertainty would constitute a panacea.

The present paper was originally drafted as a companion to Hulme & Carter (1999). They focused more specifically on how climate change scenarios have been constructed for actual use in impact studies. In a prototypical example, they employed Monte Carlo simulation techniques to demonstrate how uncertainty in GCM output cascades (via downscaling by a regional climate model) to regional climate change and its impacts, including resultant probability distributions for changes in streamflow (New & Hulme 2000). The approach adopted in their example is closely related to that in Jones (2000).

2. OVERVIEW OF UNCERTAINTY ANALYSIS

In this section, uncertainty analysis is described in generic terms. Explicit examples of the techniques employed in climate change assessments will be deferred until the next section. Before considering the various sources of uncertainty in detail, it is important to set the stage by thinking about why uncertainty matters. Clearly, uncertainty can directly affect policy, with the appropriate actions differing from those taken when only the 'best estimates' (e.g. mean values, output from deterministic models) are provided to decision makers. Among other things, the 'best' output from one model does not necessarily constitute the 'best' input to another model, because of non-normal distributions, nonlinear relationships, optimization criteria, etc. (Morgan & Henrion 1990, p. 319–320).

A simplistic answer to why uncertainty matters would be that the expression of uncertainty quantitatively in the form of probabilities is required as input to any decision or risk analysis. Yet, even if one is compelled to resort to alternative methods for modeling social choice, it is clear that information about uncertainty still is needed. Moreover, uncertainty analysis should play a diagnostic role in any integrated assessment efforts, highlighting factors whose uncertainty is pivotal in policy considerations. A recurrent theme in uncertainty analysis is 'the value of knowing how little you know' (Morgan & Henrion 1990, Chap. 12), as uncertainty is unavoidable, but the degree of uncertainty about uncertainty can be minimized (recall the quote from Socrates).

2.1. Sources of uncertainty

'Uncertainty' is a somewhat nebulous term, with attention in the present paper being restricted to those types of uncertainty that can be expressed in terms of probabilities (following the dictum of Cicero). Such a condition is not as restrictive as it first might appear, because the various interpretations of the concept of probability include subjective considerations. Motivated by the taxonomy given in Morgan & Henrion (1990, Chap. 4), the following sources of uncertainty are described: measurement error, variability, and model structure (see Table 1 for a synopsis).

2.1.1. Measurement error

Conceptually the simplest source of uncertainty, measurement error arises even when attempting to measure an unknown physical constant (e.g. the mean density of the earth or the speed of light). It is inevitable

that such errors arise, both those of a random nature whose magnitude reflects the precision of the instrument on which the measurements are based and those due to systematic error (e.g. due to the instrument being 'miscalibrated,' sometimes termed a source of 'bias'). It is generally straightforward to estimate precision, whereas the detection of systematic error can be quite difficult and this component of measurement error commonly is ignored. Yet 'even the greatest scientists, exercising every ounce of their ingenuity, are unable to eliminate all bias' (Stigler 1977, p. 1067).

2.1.2. Variability

All environmental or geophysical processes exhibit variations that reflect not just measurement error, but real systematic differences over space and time as well as inherent randomness (e.g. the limits to the predictability of future weather conditions). Not to minimize measurement errors, but variability is dominant for climate elements. Although some would make a distinction between uncertainty and variability, such a differentiation would be artificial in my view. For sometimes variability is only of interest insofar as it enables the determination of uncertainty estimates (e.g. for mean values), yet often in the climate context such variability is important for its own sake. The somewhat related issues of spatial and temporal aggregation and scaling will be discussed in Section 2.2.

2.1.3. Model structure

A more problematic source of uncertainty concerns the assumed form of model, with Morgan & Henrion (1990, p. 67) having observed that this source is 'usually more important, and more likely to have a substantial effect on the results of the analysis.' For example, important variables may have been omitted from a model (e.g. a model predicting crop yield may be missing some important climate variables). Moreover, the functional form of relationship may be misspecified (e.g. taking the relationship between crop yield and a climate variable to be linear when it is actually quadratic). Of course, such simplifications are inevitable, and any model can only be viewed as an approximation at best. In Section 2.5, the closely related issue of model size and complexity will be treated.

2.2. Scaling/aggregation

Environmental and geophysical phenomena exhibit variations over a wide range of spatial and temporal scales (this source also is included in Table 1). Not surprisingly, most models involving such phenomena have been 'tuned' to specific scales, and should not necessarily be viewed as being universal (e.g. a crop-climate model might be fitted using yields for individual plots at an experiment station). Yet it is inevitable that a model ends up being applied to different scales than that for which it has been developed. In some cases, what is termed in statistics the 'support' of a model (or of data) is not even clear. For instance, a dilemma that has been glossed over until recently is whether grid point output of a GCM represents a point or a spatial average (Skelly & Henderson-Sellers 1996). This question of GCM support will be explored in Sections 3.5.1 and 4.2.1.

What sort of uncertainties are introduced by this mismatch of scales? The parameter values and even the functional form may differ for empirically determined models, depending on the scale or level of aggregation. One might anticipate that mean values should scale in a simple way across space and time, because averaging is a linear operation. But variances would not be expected to behave in such a simple manner under different scales or levels of aggregation. An example of how moments can depend on the degree of aggregation will be provided in Section 4.2.1.

Table 1. Sources of uncertainty

Source	Description	Examples
Measurement error	Random error (precision), systematic error (bias)	Bias in historical measurements of speed of light or mean density of earth
Variability	Systematic differences (e.g. over space or time), inherent randomness	Spatial variations in precipitation, unpredictability of weather
Model structure	Function form of relationship (linear or nonlinear), specified predictor variables	Crop-climate model (linear or quadratic, missing climate variables)
Scaling/aggregation	Spatial or temporal scale (or level of aggregation) of model or data	Tuning model parameters to spatial or temporal scale, support of GCM output

One could make the case that the issue of the level of aggregation is artificial and avoidable in principle. As such, any uncertainty attributable to this source should not be regarded as inherent. But, in practice, the size and complexity of the problems (e.g. in integrated assessment) requires a substantial degree of aggregation to make implementation feasible (Rotmans & Dowlatabadi 1998).

2.3. Propagation of uncertainty

Despite the size or complexity of a given model, perhaps it might still be thought that it is relatively simple to quantify how uncertainty propagates. A number of simple heuristics do exist, but their applicability is quite limited. In particular, the models involved in global climate change studies are far too large and complex for such heuristics to be directly relevant. Nevertheless, these rules provide an intuitive feeling for how rapidly uncertainty can increase.

One heuristic is the 'multiplication rule' for the joint probability of occurrence of n independent events, $E_1, E_2, ..., E_n$:

$$\Pr\{E_1, E_2, ..., E_n\} = \Pr\{E_1\} \Pr\{E_2\} ... \Pr\{E_n\} \quad (1)$$

Here '$\Pr\{E\}$' denotes the probability of an event, E. As an illustration, suppose that at each of n stages of a sequential process an error could independently occur with a fixed probability, p. Even if p is quite small, the probability of at least 1 error is virtually 1 for a sufficiently large enough number of stages [i.e. $1 - (1 - p)^n$ by Eq. 1]. In the real world, events are usually far from independent, and this heuristic can be extremely unreliable for uncertainty analysis. By replacing unconditional with conditional probabilities, the multiplication rule can be extended to dependent events (see Eq. 5 in Section 4.1.3).

Another heuristic is the 'addition rule' for the variance of the sum of n independent random variables, $X_1, X_2, ..., X_n$:

$$\text{var}(X_1 + X_2 + ... + X_n) = \text{var}(X_1) + \text{var}(X_2) + ... + \text{var}(X_n) \quad (2)$$

Here '$\text{var}(X)$' denotes the variance of a random variable, X. In reality, some pairs of random variables are highly correlated, and this heuristic can be somewhat undependable. Still the addition rule commonly is applied in the case of measurement error, generally assumed to be additive and uncorrelated with the variable being measured. By incorporating a second term involving covariances, the addition rule can be generalized to correlated random variables.

2.4. Techniques for uncertainty analysis

Some of the primary techniques employed in uncertainty analysis now are described, with examples of their application to global climate change being deferred until Section 3. These techniques (summarized in Table 2) include sensitivity and scenario analyses, as well as a formal probabilistic approach (i.e. Monte Carlo simulation analysis). In the application of these techniques, only 1 source of uncertainty at a time typically is treated. Most commonly, the model itself is viewed, in effect, as being free of any uncertainties, only the inputs (e.g. parameter values) are permitted to vary.

2.4.1. Sensitivity analysis

In a traditional sensitivity analysis, the rate of change in the output of the model is determined as a single input is varied by an infinitesimal amount, holding the other inputs constant (technically, a partial derivative). Sometimes, the absolute rate of change is converted into a relative measure to make comparisons among different inputs more meaningful. Being a local linear approximation, sensitivity analysis often is unsuited for highly nonlinear models. In climate change and impact research, the term

Table 2. Techniques for uncertainty analysis

Type	Description	Advantages	Disadvantages
Sensitivity analysis	Partial derivative of model response with respect to input	Straightforward, intuitive appeal	Ignores uncertainty in input
Scenario analysis	Model output for trial values of inputs	Straightforward, pedagogical value	Inadequate mechanism for determining uncertainty
Monte Carlo simulation analysis	Simulated distribution for output as function of input distribution	Formal probabilistic approach	Computational burden, model uncertainty ignored

'sensitivity' is applied in a broader context (e.g. when determining whether the model output is relatively insensitive to a particular input over a wider range of possible values). Recent developments in sensitivity analysis include attempts to deal with more than one input simultaneously and to take into account uncertainty in inputs (Saltelli et al. 2000).

2.4.2. Scenario analysis

Scenario analysis is a natural extension of sensitivity analysis. Instead of varying only one input at a time, all the inputs are simultaneously changed, typically by much more than an infinitesimal amount. In this way, changes in output can be produced that correspond to realistic variations in inputs. The choice of individual scenarios sometimes is reflected in labels being attached, such as 'baseline,' 'most likely,' or 'worst case.' Morgan & Henrion (1990, p. 179) stated that 'by itself it is not an adequate mechanism for finding the uncertainty about the outcome.' Yet some very compelling arguments have been made about the value of scenario analysis, and this technique is quite popular in the global climate change arena (see Section 3.3).

2.4.3. Monte Carlo simulation analysis

It would be natural to question whether an uncertainty analysis should be limited to just a few scenarios. Plus the fact that all scenarios are not equally likely should be taken into account. Thus one is drawn to think about having effectively an infinite number of scenarios and weighting each scenario by its likelihood; that is, a formal probabilistic approach. One could conceive of performing such a probabilistic calculation analytically, but the computational requirements of numerical integration would be prohibitive for all but the simplest models.

To circumvent these computational obstacles, Monte Carlo simulation analysis often is employed (Morgan & Henrion 1990, p. 198–209), a method originally devised by physicists with the advent of electronic computers. In essence, the inputs are randomly drawn from probability distributions for each individual input and the corresponding output determined (in practice, more efficient sampling schemes can be devised). Through this process a probability distribution for the output is produced. The approach still requires the determination of distributions for the inputs, and it usually is the case that joint distributions are needed to permit dependence among inputs. In Section 4.1, some variants of the Monte Carlo approach, based on recent developments in statistics, are described.

2.5. Model size and complexity

The issue of determining the appropriate size or complexity of a model has already been alluded to in the discussion about model structure. It is natural to think that uncertainty automatically can be reduced by increasing the size or complexity of a model. Under many circumstances, reductions in uncertainty can be achieved, but beyond a certain point increased size or complexity may actually be a hindrance. Hence some guidance is needed about how to determine whether a model is sufficiently large or complex, given the information available. In the present paper, the issue of how to define or measure model 'complexity' is not addressed (but see Morgan & Henrion 1990, p. 289–291), only focusing on model size (e.g. number of parameters, spatial and temporal resolution).

Some hints about the form of guidance can be gleaned from what is known in the field of statistics about the selection of the 'best' model from a set of candidates. For criteria such as minimizing prediction error, it is often preferable to select a model that is smaller than the 'true' one, especially if the true model requires an effectively infinite number of parameters. Basically, there is a tradeoff between (1) improved goodness-of-fit through incorporating additional variables in the model and (2) increased sampling error due to estimating more parameters from a limited amount of data. To achieve parsimony, it turns out that a penalty that is a function of the size of a model (e.g. number of parameters estimated) needs to be imposed to avoid overfitting (Schwarz 1978). In a more general context, this principle of parsimony is sometimes termed 'Occam's razor' (e.g. Shackley et al. 1998).

Especially subtle is the danger that linearizing approximations, intended to simplify a complex model, neglect system nonlinearities that turn out to be pivotal (Morgan et al. 1999). Still there is a tendency to become mesmerized by very large and complex models (e.g. GCMs), automatically assuming that they are preferable to simpler ones. At least insofar as uncertainty is concerned, this need not be the case. 'Without thorough and systematic modeling and analysis of the uncertainty of the problem, we can not be sure that the results of a model, especially a very large and complex one, mean anything at all' (Morgan & Henrion 1990, p. 304). Of course, large numerical models exist for other purposes, not necessarily compatible with the desire for parsimony.

2.6. Communication and interpretation of uncertainty

Besides quantifying uncertainties, the issue remains of how to communicate such uncertainties to decision

and policy analysts. So a few comments are in order concerning information about uncertainty in the form of probabilities. Not surprisingly, some resistance to reliance on probabilities exists, and it has proved necessary to make the case that the statistics of climate change themselves ought to be expressed in probabilistic terms (Katz 1993).

It is often argued that it is preferable to provide decision makers with only qualitative information about uncertainty, instead of probabilities. Sometimes the argument is couched in terms of translating probabilistic information into qualitative terms so that the recipients would better understand it, other times it amounts to claiming that it is not worth the extra effort to produce quantitative information about uncertainty. The problem with such arguments is that studies consistently show that expressions of uncertainty in qualitative terms (e.g. 'unlikely' or 'quite likely') are subject to a wide range of interpretations (Beyth-Marom 1982). For seasonal climate forecasts, Nicholls & Kestin (1998) cited a recent case in which confusion arose over the meaning of the term 'likely.' In one of the few studies to go beyond relying on anecdotal evidence, Murphy et al. (1980) indicated that much confusion about probability of precipitation forecasts indeed exists, but about the event to which the probability is assigned, not anything inherently to do with probability itself. The issue also has been raised about whether it is preferable to disseminate absolute frequencies instead of probabilities (Hoffrage et al. 2000).

Besides 'ambiguity' about either probabilities or events, Fischhoff (1994) identified 3 other issues that arise in the communication of uncertainty: (1) 'irrelevance' ('a forecast is of little use unless it provides information that potential consumers need,' Fischhoff 1994, p. 391); (2) 'immodesty' ('recipients may still be left guessing whether the forecasters have overstated or understated how much they know,' Fischhoff 1994, p. 393); (3) 'impoverishment' (a 'tendency to view what is customarily modelled as being what really matters,' Fischhoff 1994, p. 397–398).

It is safe to say that all of these issues have arisen in the context of global climate change.

3. PRESENT USE OF TECHNIQUES

In the present section, how techniques actually have been employed in uncertainty analysis as part of assessments of global climate change are surveyed, with a synopsis of the examples treated being given in Table 3. For the most part, the techniques already have been introduced in the overview of uncertainty analysis (Section 2). Some of the discussion concerns related methods that are not necessarily a part of formal uncertainty analysis (i.e. model validation and intercomparison). Also, some other issues requiring specialized treatment have been postponed until now (e.g. extremes/surprises).

3.1. Model validation and intercomparison

The topic of model validation and intercomparison was not explicitly treated in the overview of uncertainty analysis. Yet, in discussions about the limitations of climate change assessments, much attention is devoted to the need for adequate validation and intercomparison of models (especially GCMs, Mearns 1997, Gates et al. 1999). Certainly, the careful evaluation of GCMs is important from a research perspective, potentially providing useful diagnostic information. What connection validation has with uncertainty analysis is explored here. I will avoid the philosophical minefield about whether the 'validation' of numerical models is, in any sense, possible (but be aware of the sentiment in the social sciences that 'validation of numerical models of natural systems is impossible,' Oreskes et al. 1994, p. 641; 'verification of a model is logically impossible,' Petersen 2000, p. 268).

The evaluation of GCMs involves a comparison of the output from a GCM control run with the observed

Table 3. Examples of use of uncertainty analysis in climate change studies

Type (source)	Application	Results
Sensitivity analysis (Mearns et al. 1996)	CERES-Wheat model used to determine sensitivity of crop yields to climate variability	Demonstrated importance of effects of climate variability on crop yields
Scenario analysis (Alcamo & Kreileman 1996)	IMAGE global change model used to evaluate effectiveness of emission pathways	Identified 'safe emission corridors,' relating emission pathways to climate protection goals
Monte Carlo simulation analysis (Shackley et al. 1998)	EL global carbon cycle model used to evaluate effects of parameter uncertainty on carbon dioxide levels	Previous studies underestimated uncertainty in projected carbon dioxide levels

climate (or with control runs for other GCMs). At first, the evaluation techniques focused on statistical tests of significance, in which the null hypothesis is that the output from the control run and observations (or the outputs from control runs for 2 or more GCMs) are drawn from the same statistical population. As pointed out by Katz (1992), such an approach (with its 0–1 categorization of a model as either 'passing' or 'failing' the significance test) does not necessarily provide much information about how well the GCM is actually able to approximate the observed climate (also see Nicholls 2001). Of course, statistics such as error variances that are sometimes determined as byproducts of evaluation exercises constitute one measure of the ability of a GCM to reproduce the observed climate. Recent model intercomparison activities, such as the Atmospheric Model Intercomparison Project (AMIP, Gates et al. 1999), have produced such byproducts. Ideally, what would be needed for purposes of uncertainty analysis are posterior probability distributions (or at least confidence intervals), instead of just descriptive statistics such as error variances (Katz 1992). For more discussion of GCM validation, see von Storch & Zwiers (1999, p. 129–130).

3.2. Sensitivity analysis

Quite a few uses of sensitivity analysis can be found in the literature on impacts of climate change. Just one example will suffice to illustrate what sort of details are involved in implementing this technique in the context of climate change. The Mearns et al. (1996) study (Table 3) was motivated by the possibility that climate change may be reflected in more than mean values of climate variables (i.e. in increases or decreases in variations about the, possibly changing, mean). Although there is no dispute that climate variability does have an effect on agricultural production (Parry & Carter 1985), 'there is considerable quantitative uncertainty concerning how agricultural crops respond to changes in climate variability' (Mearns et al. 1996, p. 257). The Mearns et al. research focused on the sensitivity of wheat yields to the variability of temperature and precipitation, with the response of yield to climate being determined by the CERES-Wheat model (Ritchie & Otter 1985). Based on a 'mechanistic' approach in which yield is represented as a function of several factors, including the weather, the Mearns et al. study can be viewed as a sensitivity analysis of the CERES-Wheat model itself.

The range of change in climate variability considered in Mearns et al. (1996) was selected to be relatively broad and to reflect shifts thought possible under a doubling of concentration of greenhouse gases.

These variability changes are imposed through the use of a particular stochastic weather generator, the Richardson model (or WGEN) (Richardson 1981). For a given set of parameter values, this stochastic model can generate synthetic time series of daily weather, including precipitation amount and maximum and minimum temperature, for any desired number of years. To achieve the intended changes in climate variability, the values of the appropriate parameters in WGEN are adjusted. For each level of climate variability, a probability distribution of yield is produced via the CERES-Wheat model (i.e. one yield value corresponding to each year of simulated weather). In this way, the effects can be measured in terms of shifts in both the mean and variance of yield.

Mearns et al. (1996, p. 257) concluded that 'this study demonstrates the importance of taking into account change in … variability when analyzing the effect of climate change on crop yields.' As such, this work certainly constitutes a potentially valuable contribution to the climate impacts literature. Still, the extent to which uncertainty in wheat yield has actually been quantified is quite limited, as the likelihood of the prescribed changes in variability has not been taken into account. In other words, it is unclear how to end up with a single probability distribution for yield, as opposed to the several distributions produced by the sensitivity analysis. They also cautioned about uncertainties in the CERES-Wheat model itself. Besides a continuation of this work (including an extension to scenario analysis, Mearns et al. 1997), several other sensitivity analyses of crop yields to climate variability and change have been performed (e.g. Riha et al. 1996, Wolf et al. 1996). To take into account crop model uncertainty, one strategy would be to replace the deterministic crop model with one based on a combination of mechanistic and statistical modeling (Landau et al. 2000).

3.3. Scenario analysis

Scenario analysis is the technique most relied on currently in climate impact research. In practice, the distinction between sensitivity and scenario analyses is somewhat blurred, being more related to how the information is used than to the tools involved. Scenario analysis is intended to be policy relevant, whereas the goals of sensitivity analysis are more limited. A case in point is the 'viewpoint' expressed by Alcamo et al. (1996) in introducing a special issue of the journal *Global Environmental Change* dealing with scenarios for policy making. They related the lessons gleaned in attempting to communicate results of the IMAGE model for integrated assessment of global change to policy advisors.

To be more specific, let us focus on one of the companion articles in the special issue (Alcamo & Kreileman 1996) that makes use of the IMAGE model to produce scenarios for policy makers (Table 3). The IMAGE 2 model includes 3 coupled subsystems: energy/industry, terrestrial environment, and atmosphere-ocean. Alcamo & Kreileman identified 'safe emission corridors' (i.e. minimum and maximum acceptable pathways) that would comply with long-term goals of climate protection. No discussion of uncertainty really is undertaken until the concluding section, where it finally is acknowledged that 'results presented in this paper have many sources of uncertainties, including the uncertainties inherent in the IMAGE 2 model' (p. 330). No quantification of uncertainty is attempted, rather only a caveat ('keeping in mind the uncertainties of the analysis,' p. 331) is given when the findings are listed.

In discussing the Alcamo & Kreileman (1996) study and some others, Alcamo et al. (1996, p. 259) made a compelling case for the usefulness of scenario analysis to policy makers, concluding that it is 'an effective form for communicating a large amount of technical information.' This approach helps the user understand how large and complex models work and can demonstrate whether particular results are 'robust' to certain assumptions (really a variant of sensitivity analysis). Yet Alcamo et al. (1996) were aware of the shortcomings of scenario analysis, having included in a list of 'methodological improvements' in integrated assessment modeling that 'uncertainty needs to be quantified' (p. 259). In other words, scenario analysis is no substitute for uncertainty analysis. Often the warning is issued about 'scenarios' not being 'predictions,' but this distinction is somewhat artificial and ineffective, with policy makers not necessarily feeling constrained by terminology alone.

3.4. Monte Carlo simulation analysis

Perhaps both the sensitivity and scenarios analyses performed to date, as part of climate change assessments, should best be regarded as necessary precursors to full-fledged uncertainty analysis. Unfortunately, few examples exist in which formal probabilistic uncertainty analysis has been applied to global climate change models. The existing analyses have focused on parameter/input uncertainty (i.e. ignoring any imperfections in model structure). Because of complexity (including the 'curse of dimensionality' that arises in problems requiring optimization), these analyses generally have dealt with simpler versions of global change models (e.g. ones that have a high degree of aggregation). Let us focus on a recent study

by Shackley et al. (1998) that includes a Monte Carlo simulation analysis of uncertainty in modeling the global carbon cycle (Table 3).

Shackley et al. (1998) considered the Enting-Lassey (EL) model of the global carbon cycle (Enting & Lassey 1993), a deterministic nonlinear model (for more details about the EL model, also see Parkinson & Young 1998). It was selected because it is smaller than a GCM, although still large enough to be challenging from the perspective of uncertainty analysis. Taking a Bayesian approach, prior probability distributions are assigned to the various parameters of the EL model. For want of other information, these priors are assumed to be independent normal distributions. The parameter values are randomly varied and the EL model repeatedly run, as in a standard Monte Carlo simulation analysis (see Section 2.4.3). Despite the limitations in this study (e.g. ad hoc choice of prior distributions, no treatment of uncertainty in model structure), Shackley et al. (1998, p. 179–180) found that the estimated uncertainties in future atmospheric levels of carbon dioxide are 'significantly greater than those gained by deterministic model studies alone.'

Henderson-Sellers & McGuffie (1999) took some issue with Shackley et al. (1998), asserting that 'most, if not all, climate scientists are only too aware of the uncertainties both in "their" models and throughout policy development' (p. 597). In their rebuttal, Shackley et al. (1999, p. 613) maintained that "climate model developers" … are still predominantly deterministic in their orientation, even thought they acknowledge the uncertainty in the global climate system.' They went on to state that 'the need for a stochastic analysis is not emphasized … despite the fact that the models are being used to forecast the behaviour of a highly uncertain system' (p. 613).

Applying Monte Carlo simulation analysis directly to GCMs would be of considerable value. Although GCMs have been used to generate ensembles, it is not yet computationally feasible to produce ensemble sizes large enough for uncertainty analysis. Nevertheless, recent developments in 'ensemble forecasting' (Palmer 1993, Wilks 1995, p. 210–221), now becoming prevalent in short- to medium-range weather forecasting, suggest this approach will eventually become practicable (an important limitation to most ensemble forecasting efforts is the assumption that the numerical prediction model is perfect). As at least a temporary expedient, one approach would be to approximate such a large and complex deterministic model by a smaller and simpler statistical one. For example, Shackley et al. (1998) used 'dominant mode analysis' to show that a relatively simple linear statistical model closely resembles the original nonlinear EL model. Before performing Monte Carlo simulation analysis,

Webster & Sokolov (2000) applied 'deterministic equivalent modeling' to approximate statistically a 2-dimensional, zonally averaged climate model (i.e. still not as large or complex as a GCM).

Rotmans & Dowlatabadi (1998, p. 334) recognized that 'advances in both computational methods and heuristic approaches to limiting the number of uncertain parameters are needed to keep pace with the growing complexity of integrated assessment models of global change.' Some recent developments in computational statistics that could help to alleviate this difficulty will be described in Section 4.1. Other recent studies in climate change and impact modeling that make use of Monte Carlo simulation analysis include Dowlatabadi (1998) and Wigley & Raper (2001), as well as those by Jones (2000) and New & Hulme (2000) already cited.

3.5. Aggregation and downscaling

In this subsection, 2 topics related to uncertainty in scaling and aggregation that arise in research on global climate change are treated; namely, the interpretation of GCM grid point output and statistical downscaling.

3.5.1. GCM grid point output

First, the issue of uncertainty attributable to scaling or aggregation is considered for the example of GCM grid point output introduced in Section 2.2. It is somewhat surprising that this question has only been explicitly considered relatively recently (Skelly & Henderson-Sellers 1996, Osborn & Hulme 1997). It has implications for the appropriate comparison with observations in the evaluation of GCMs, as well as in the interpretation of GCM scenarios by climate impact researchers. This issue is especially problematic for precipitation, with its high degree of variation on very small spatial and temporal scales. GCM numerically simulated precipitation sometimes is criticized as having the property of 'drizzling too much.' Although inadequate model physics is part of the problem, it may be unfair to compare an areal quantity with point observations (i.e. based on only a few rain gauges).

Clearly, GCM output is intended to represent a grid box. But Skelly & Henderson-Sellers (1996) made the case for GCM grid point output being, in effect, a 'hybrid' between a point and an areal quantity; in some respects (e.g. numerical computations) resembling a point, whereas in others (e.g. parameterizations) a spatial average. On the other hand, Osborn & Hulme (1997) started with the assumption that GCM output

does represent quantities averaged over the entire box surrounding the grid point, with their goal being 'more faithful' (p. 1885) evaluation of GCM simulated precipitation. Using observations of daily precipitation based on rain gauges, they derived empirical relationships that predict precipitation statistics averaged over an entire grid box, given only data for a limited number of sites within the box. How to exploit stochastic modeling and scaling theory in evaluating GCM output will be considered in Section 4.2.1.

3.5.2. Statistical downscaling

An approach known as 'statistical downscaling' is becoming increasing popular for dealing with mismatches in scale (von Storch et al. 1993, Lettenmaier 1995, Hewitson & Crane 1996, Wilby & Wigley 1997). For instance, empirical relationships between local precipitation and patterns in large-scale atmospheric circulation are utilized. The basic idea being that it is preferable to obtain climate change scenarios for precipitation indirectly from these downscaling relationships than to rely directly on GCM simulated precipitation (perhaps interpolated to a higher resolution).

Most of the methods employed in statistical downscaling are 'regression-like' procedures (i.e. methods that only specify a conditional mean, including multiple regression analysis and other techniques such as 'artificial neural nets,' Hewitson & Crane 1996), and are reminiscent of the 'model output statistics' approach often used in the production of operational weather forecasts (i.e. in which the output of numerical weather prediction models is statistically adjusted; Wilks 1995, p. 201–210). Specifically, such procedures can be expressed in terms of the conditional mean as:

$$Y = E(Y \mid X_1 = x_1, X_2 = x_2, …, X_p = x_p) + \varepsilon \qquad (3)$$

Here Y denotes the variable to be predicted via statistical downscaling, the X_1, X_2, …, X_p covariates, perhaps based on GCM output, and ε a random error term. Other statistical characteristics of the variable Y are either only implicitly modeled (i.e. its conditional variance or the variance of the error term in Eq. 3) or not modeled at all (e.g. its autocorrelation). A refinement in this approach to downscaling, more amenable both to scenario and uncertainty analysis, will be described in Section 4.2.2.

3.6. Extremes/surprises

'Surprise' is a topic that periodically receives at least limited attention in the global climate change area. 'It is usually the extreme, surprising events that cause the

most damage to human health and property,' and 'while some kinds of surprises are truly unpredictable, there are other types that could be anticipated and their adverse effects forestalled' (Streets & Glantz 2000, p. 97). Almost by definition, genuine unpredictable surprises will remain outside the scope of uncertainty analysis. Hammitt & Shlyakhter (1999) examined the role of surprise in risk assessment, including an application to global climate change.

In the present paper, the discussion is limited to what Streets & Glantz (2000, p. 101) referred to as 'surprises that probably should not be surprising, but often are;' in particular, to extreme events. Although much discussion in recent years has been devoted to the importance of extreme events in climate change, it is quite rare that modern statistical theory for extremes is applied (notable exceptions are Zwiers & Kharin 1998, Kharin & Zwiers 2000). Yet the statistics of extremes has found extensive application in engineering design problems (e.g. for hydrologic purposes) that are directly relevant to climate impacts (Faragó & Katz 1990). When the fit of an assumed form of distribution to the data is examined, it is not unusual to discover that the frequency of extreme high or low values is underestimated. To cite a climate example, Wilks (1999) found that the gamma distribution underestimates the probability of high precipitation amounts at many locations. As an example related to impacts, Shlyakhter et al. (1994) showed that the errors in energy projections have a distribution whose lower and upper tails are longer than the normal.

The types of probability distribution typically assumed for climate and climate impact variables (e.g. normal, exponential, gamma, lognormal) all possess what is termed a 'medium' right-hand tail. That is, the tail of the distribution (i.e. $1 - F(x)$, sometimes called the 'survival function,' where $F(x) = \Pr\{X \le x\}$ denotes the cumulative distribution function) satisfies the approximation:

$$\ln[1 - F(x)] \approx L(x) \qquad (4)$$

for large x, where L is a linear function (for the exponential distribution, Eq. 4 holds exactly instead of approximately). Yet there is evidence that the distributions of climate variables such as precipitation and of related variables like streamflow, as well as of the associated economic damage from extreme events such as floods or hurricanes, do not have a medium, but a 'heavy' right-hand tail (incidentally, no distributions with heavy tails are included in Morgan & Henrion 1990, Chap. 5). If so, then events are being labeled as more surprising than they should be. Heavy-tailed distributions, evidence for their existence, and their implications for uncertainty analysis will be discussed in Section 4.3.

4. ALTERNATIVE APPROACHES

Modern statistical methodology, potentially relevant to uncertainty analysis of global climate change models, is outlined in this section. Earlier in the present paper, the case has been made that anything less than a fully probabilistic approach to uncertainty analysis is inadequate.

Having been advocated for use in uncertainty analysis of global climate change (Paté-Cornell 1996, Hobbs 1997), Bayesian statistical methods for probabilistic analysis now are addressed. More generally, the Bayesian statistical paradigm is making a 'comeback' in statistical applications (Malakoff 1999).

4.1. Probabilistic analysis

In this subsection, first the Bayesian statistical paradigm is reviewed and the class of hierarchical stochastic models, well suited for Bayesian analysis, introduced. Then consideration is devoted to recent computational developments that make this approach more feasible in applications (i.e. so-called Markov chain Monte Carlo techniques) and to uncertainty about model structure (taken into account through model averaging).

4.1.1. Bayesian statistical paradigm

Bayesian statistical methods are not at all new (in many respects, they predate 'classical' techniques), and are philosophically attractive for a number of reasons. Yet, until recent years, computational limitations served to hinder the application of Bayesian methodology to large or complex problems. But now 'the tide has turned,' and pragmatic statisticians routinely take advantage of Bayesian computational methods even when their ultimate aim is classical estimation.

First, the basic steps in the Bayesian statistical paradigm are briefly reviewed (for an introduction to the use of Bayesian statistics in climatology, see Epstein 1985): (1) Elicit the joint 'prior probability' distribution for the parameters of the statistical model under consideration (the prior is formulated before 'seeing' the data, and could be 'diffuse' or non-informative). (2) For the statistical model, obtain an expression for the joint conditional distribution of the observations as a function of the parameters (termed the 'likelihood function' in classical statistics). (3) Determine the 'posterior probability' distribution for the parameters given the data by combining (via Bayes' rule) the prior distribution with the likelihood function evaluated at the sample values.
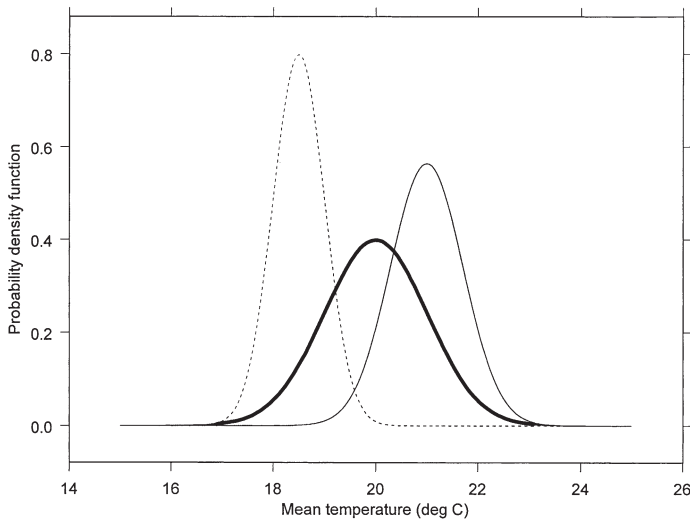
Fig. 1. Example of Bayesian statistical paradigm, showing posterior distributions of mean temperature for 2 sample outcomes (thin solid line: one with smaller sample size; dashed line: other with larger difference from prior), along with prior distribution (thick solid line) (see Appendix 1)

The computational difficulty already alluded to arises in step (3), as Bayes' rule generally requires the numerical evaluation of integrals, until recently too complex if more than a very small number of parameters are involved.

How these 3 steps in the Bayesian statistical paradigm work is demonstrated by a simple example (Fig. 1), with the details about its construction being given in Appendix 1. In this example, prior information about the mean temperature (e.g. based on physical theory or on previous observations) is combined with sample information to produce a posterior distribution. The figure illustrates how the posterior distribution for the mean temperature involves an adjustment to the prior, depending both on the number of temperature measurements and on how much the information about the mean contained in the sample differs from the prior. Although one posterior distribution is shifted by a smaller amount from the prior than is the other posterior, it has a higher variance (see Appendix 1).

The Bayesian approach commonly has been criticized for requiring the subjective assessment of prior probability distributions for model parameters, but this aspect actually is viewed by some as an advantage. Although it may be difficult to elicit the opinions of experts (e.g. Morgan & Keith 1995), it is not difficult to incorporate such subjective judgments into a Bayesian analysis. The Bayesian approach also makes it straightforward to combine different sources of information (updating the posterior distribution through taking the previous posterior distribution to be the present prior, etc.).

Despite its overwhelming reliance in applications, classical statistical inference has a number of limitations that hinder its utility in uncertainty analysis. To highlight just one, the so-called 'plug-in' method is usually employed to obtain 'standard errors' to attach to any estimates. This technique requires that the corresponding parameter estimates be substituted, in place of any unknown parameters, into the expression for the standard deviation of the estimator. So one important source of uncertainty, that of the parameter estimates themselves, typically is ignored. For instance, the plug-in approach is commonly used by climatologists to estimate the variance of a time average (or the 'equivalent sample size'), ignoring the uncertainty in the sample autocorrelation coefficients (von Storch & Zwiers 1999, p. 115–116). Despite being too difficult for classical statistics to handle in all but the simplest situations, this source of uncertainty routinely is taken into account in the Bayesian approach.

Before treating recent developments in Bayesian computing, 2 rudimentary examples of the application of Bayesian statistics that are relevant for uncertainty analysis of climate and climate impacts modeling are briefly reviewed. One study concerns how the uncertainty in estimating yield response functions (i.e. relating crop yield to ozone exposure) affects economic benefits for various levels of pollutant standards (Adams et al. 1984). Based on a limited amount of experimental data, the estimated yield response function was derived from a regression analysis. The Bayesian analysis takes into account parameter uncertainty in the fitted regression model, assuming a diffuse/non-informative prior for the parameters involved (i.e. as if no prior information were available). Posterior distributions of the economic benefits associated with each level of pollutant standard were obtained, indicating to what extent the benefit distributions for the different options overlap. Although several important sources of uncertainty were ignored, this study still illustrates how the uncertainty, in this case associated with experimental data, propagates through the system and ultimately affects the calculation of economic benefits. An uncertainty analysis of a crop-climate model would be somewhat analogous.

A more recent study includes a fully Bayesian analysis of the uncertainty in estimates of the 'climate sensitivity' (namely, the change in global mean temperature corresponding to a doubling in the concentration of atmospheric carbon dioxide 'equivalents') (Tol & de Vos 1998). Based on statistical modeling of time series of global mean temperature and covariates (including atmospheric carbon dioxide concentration), the study makes use of expert assessments of the likely values of climate sensitivity as prior distributions (Morgan & Keith 1995). Although the posterior distribution of

climate sensitivity turned out to be sensitive to the assumed form of prior information, Tol & de Vos (1998, p. 105) concluded that 'the data heavily discount the likelihood of a small climate sensitivity' and that it 'may well be very large.' This study can be criticized for being oversimplified (e.g. the influence of sulphate aerosols is ignored). Still it demonstrates how subjective information from experts can be readily combined with observations in the Bayesian approach.

4.1.2. Hierarchical stochastic models

A powerful approach to the stochastic modeling of large and complex systems is known as 'hierarchical' modeling. Such models are not inherently Bayesian, but their form makes the computations involved in the Bayesian approach amenable. The concept of a hierarchical model makes use of the 'power of conditioning,' generally involving several stages (or layers), but with conditional independence assumed between stages. In this way, complex dependencies are produced unconditionally, but the advantages of independence can be exploited conditionally at each stage of the analysis.

First the term 'conditional independence' is more fully explained. An example is shown in Fig. 2, with the details being given in Appendix 2. Suppose that the temperature at 2 sites is conditionally independent given the state of large-scale atmospheric-oceanic circulation (e.g. indicating whether or not an El Niño event occurs). This assumption implies that the temperature at the one site could be stochastically simu-

lated independently of the temperature at the other site, just alternating which conditional distribution is used (depending on El Niño). Then, unconditionally, dependence in the temperature at the 2 locations is induced by virtue of conditioning on the same circulation state. In this example, the induced correlation between the temperature at the 2 sites is about 0.4 (see Appendix 2).

Next a recent example is described in which Bayesian hierarchical modeling of climate-related phenomenon has been attempted. Wikle et al. (1998) proposed a multi-stage model for environmental space-time data. The first stage represents the data in terms of an underlying process of interest and a measurement error component, whereas the second stage separates the underlying process into 'climate' and 'weather' components. In the third stage, spatial dependence in the climate component is modeled, as well as the space-time dependence in the weather component (with relatively simple structure possible because of the power of conditioning, as well as a provision for the incorporation of nonlinear dynamics or 'physics'). The remaining stages involve the specification of prior distributions. As all of the stages are intentionally designed to be conditionally independent, this approach is a relatively straightforward way to develop space-time stochastic models (i.e. circumventing the need to model large and complex space-time covariance matrices).

As a prototypical application, Wikle et al. (1998) applied this approach to a space-time field of temperature data. More recent work concerns the prediction of the space-time field of sea surface temperatures in the El Niño region of the Pacific Ocean. In this innovative application, uncertainty is attached to the El Niño forecasts as a routine byproduct of the Bayesian analysis (Berliner et al. 2000). Thus modern Bayesian statistical methods are making rapid advances toward the time when the stochastic modeling of GCM output could be realistically contemplated (as advocated by Dempster 1998). Such a stochastic model would not make GCMs obsolete. Rather, it would enable extensive uncertainty analysis of a GCM to be conducted, not presently possible through conventional Monte Carlo simulation analysis. This approach is only conceivable because of recent developments in Bayesian computation, a topic to be dealt with in the next subsection.
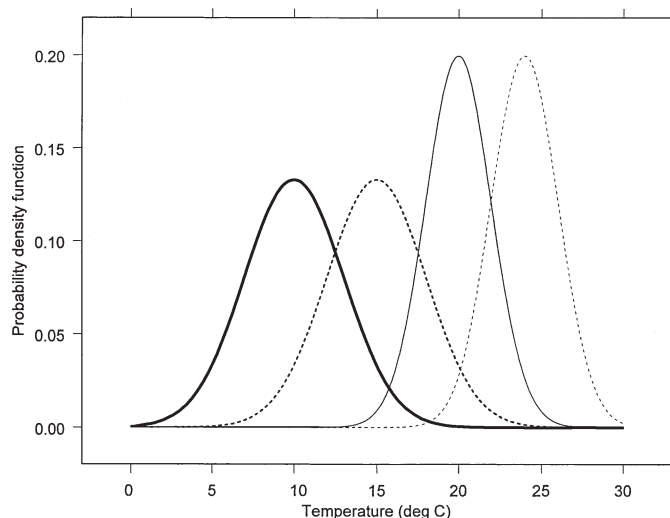

Fig. 2. Example of induced dependence, showing conditional distribution of temperature given El Niño event (dashed lines) or non-El Niño event (solid lines) for 2 sites (thick lines for one site, thin lines for other site) (see Appendix 2)

4.1.3. Markov chain Monte Carlo

Although the idea originated in statistical physics more than 40 yr ago, only in the present decade has the computational device known as Markov chain Monte Carlo (MCMC) been extensively exploited in Bayesian

statistics. One particular implementation of MCMC is termed 'Gibbs sampling' (for an expository treatment of this algorithm, see Casella & George 1992), the method used in the Wikle et al. (1998) study just reviewed. Lu & Berliner (1999) gave another geophysical application of MCMC, fitting a nonlinear ('regime-switching') time series model to rainfall-runoff data by a Bayesian approach.

MCMC is a method that indirectly simulates a joint probability distribution of any desired dimension (e.g. the joint posterior distribution of a set of parameters in a Bayesian analysis). To be applied, tractable expressions are only needed for the 'full conditionals;' that is, the conditional distribution of an individual parameter, given all of the remaining ones (the hierarchical models just discussed are designed with this property in mind). In particular, the joint probability of n (possibly dependent) events, E_1, E_2, …, E_n, can be expressed as:

$$\Pr\{E_1, E_2, \ldots, E_n\} = \\ \Pr\{E_1\}\, \Pr\{E_2 \mid E_1\} \ldots \Pr\{E_n \mid E_1, E_2, \ldots, E_{n-1}\} \quad (5)$$

which is a generalization of Eq. (1). The main point is that all of the probabilities on the right-hand side of Eq. (5) are full conditionals. A purely artificial Markov chain can be constructed (i.e. the Markov dependence has no connection to whether the data being modeled are dependent) with the property that its limiting distribution is precisely the desired one. Although these developments have led to a virtual 'revolution' in Bayesian statistics, it should be kept in mind that further computational breakthroughs will be required before problems as large and complex as a fully stochastic space-time model for GCM output could be routinely tackled by this approach.

4.1.4. Model uncertainty

Most aspects of uncertainty about model structure are not routinely taken into account in climate change and impact studies. Yet 'any approach that selects a single model and then makes inference conditionally on that model ignores the uncertainty involved in model selection, which can be a big part of overall uncertainty,' and 'this leads to underestimation of the uncertainty about quantities of interest, sometimes to a dramatic extent' (Kass & Raftery 1995, p. 784). In principle, the Bayesian approach can account for model uncertainty through 'averaging' (Hoeting et al. 1999). The posterior probabilities of all candidate models given the data are determined, and then these probabilities are employed as weights to combine estimates based on the various models in the final analysis. These posterior probabilities are closely related to 'Bayes factors,' as well as the Bayesian information cri-

terion (or BIC) popular for model selection (Kass & Wasserman 1995).

In practice, it still is quite difficult to implement the Bayesian approach to account for model uncertainty, as having more than a few candidate models makes the calculations infeasible. Kass & Raftery (1995) proposed 'Occam's window,' a variant of Occam's razor (mentioned in Section 2.5), in which only the strongest candidates enter into the model averaging. It should be noted that there is some opposition to model averaging. Specifically, Morgan & Henrion (1990, p. 68) argued:

'*Every model is definitely false*. Although we may be able to say that one model is *better* than another, in the sense that it produces more accurate predictions, we cannot say it is more probable. Therefore, it is inappropriate to try to assign probabilities to models.'

Perhaps for this reason, Bayesian model averaging has rarely if ever been attempted in the climate change context.

4.2. Stochastic models for scaling/aggregation

In this subsection, the issues of the scale of GCM grid point output and of downscaling are revisited from a more statistical perspective.

4.2.1. Stochastic scaling

Extensive research on statistical scaling properties in the environmental and geophysical sciences has been performed, especially motivated by interest in 'fractal' behavior. In part, the goal is to develop a universal principle that would enable statistical characteristics for different scales to be related. Although no such universal rule has been found, some simple heuristics have been obtained. For example, Gupta & Waymire (1993) considered the 'log-log law' that arises for fractals or 'multifractals,' governing the relationship between the moments of a spatially averaged quantity (e.g. mean or second moment) and the corresponding spatial scale (i.e. the size of region over which the quantity is averaged).

In the case of precipitation, conditional on its occurrence, the hth-order moment is assumed to satisfy:

$$\ln E\{[X(s)]^h \mid X(s) > 0\} = \alpha_h + \beta_h s \quad (6)$$

where $h > 0$, $X(s)$ denotes a variable averaged over an area of size s, and α_h and β_h are constants possibly depending on h. An analogous log-log relationship is assumed for the probability of occurrence of precipitation. Gupta & Waymire (1993) found that such a law is approximately satisfied for rainfall derived from radar

reflectivity, and proposed a stochastic model known as a 'random cascade' to probabilistically represent such behavior.

Brown et al. (1995) applied this log-log law as a heuristic in an exploratory study on the spatial scale of grid point precipitation simulated by a numerical model. Scaling characteristics for precipitation simulated by a regional climate model (i.e. a mesoscale model nested within a GCM) were compared with those of aggregated hourly precipitation observations for a dense network of rain gauges. Although the results were somewhat inconclusive, they suggest that GCM simulated precipitation has a similar spatial structure to but a different 'scale' from the observations (i.e. somewhere between a point and the entire grid box). More systematic studies of this type, in combination with theoretical developments on stochastic scaling, would help make the uncertainty attributable to aggregation more amenable to analysis.

4.2.2. Statistical downscaling

The focus of statistical downscaling on conditional means has been pointed out in Section 3.5.2. Usually, even the error term in Eq. (3) is neglected, with a consequence that the variance of the downscaled variable is underestimated. Only quite recently has 'randomization' (i.e. explicit inclusion of the error term) in downscaling been advocated (von Storch 1999). Although the underestimation of variance is not an inherent defect in the regression-like approach to downscaling, it has other shortcomings that are not necessarily remedied by simple adjustments such as randomization.

If downscaling were formulated in terms of conditional stochastic processes instead, then a framework for full-fledged uncertainty analysis would be created. It still is rare to regard downscaling as involving stochastic processes (as opposed to just static statistics), and rarer yet to study its statistical properties (for one exception, see Katz & Parlange 1996). Formally, the stochastic approach calls for modeling the conditional distribution of the present state of the variable to be downscaled, say denoted by $Y(t)$ at time t, given not only the p covariates X_1, X_2, …, X_p (using the same notation as in Section 3.5.2), but its own past history, say denoted by $Y(t-1)$, $Y(t-2)$, …, $Y(1)$; that is:

$$\Pr\{Y(t) \le y \mid Y(t-1) =$$
$$y_{t-1}, …, Y(1) = y_1, X_1 = x_1, …, X_p = x_p\} \quad (7)$$

One way to implement this approach involves stochastic weather generators, with the parameters of the generator being varied conditional on covariates (see Wilks & Wilby 1999 for a review of weather generators,

including their use in downscaling). The formal treatment of weather generators as stochastic processes usually is lacking as well (one exception is Katz 1996). Ordinarily, parameter uncertainty is ignored in climate change scenarios produced by downscaling, whether through weather generators or regression-like approaches, with a Bayesian approach again being the most natural way to take this source of uncertainty into account. It should be noted that one alternative to a stochastic weather generator is to employ 'resampling' instead, generating synthetic times series of weather (through random draws from the original sample) conditional on atmospheric variables (Buishand & Brandsma 2001).

Although the performance of different methods for statistical downscaling is starting to be evaluated (Wilby et al. 1998), so far the inherent advantage of this approach with respect to uncertainty analysis has apparently not been recognized. Even if statistical downscaling does not necessarily improve upon some alternative (e.g. direct reliance on GCM output), it still would be useful in enabling uncertainty to be estimated. In other words, there is value in adding a downscaling layer to the analysis, even if no 'downscaling' literally needs to take place.

4.3. Extreme value theory

A well-developed statistical theory of extreme values exists, but texts which connect this theory to applications are lacking (one recent attempt is Reiss & Thomas 1997). Without describing this theory in any detail, the most important point is that it differs in fundamental ways from the corresponding statistical theory for averages. Instead of the normal distribution, ubiquitous for averages or sums, skewed distributions arise for extremes (e.g. the maximum value over some time period), in some cases having heavy tails. This theory is 'robust,' in the sense that it does not require making very specific parametric assumptions about the shape of the extreme tails of distributions (for which not much data is ordinarily available).

Formally (using the same notation as in Section 3.6), the survival function of a heavy-tailed distribution satisfies the approximation:

$$\ln[1 - F(x)] \approx L(\ln x) \quad (8)$$

for large x. Note that Eq. (8) is satisfied exactly for a Pareto distribution (Reiss & Thomas 1997, Chap. 1; also see Appendix 3). A distribution satisfying Eq. (8) has fundamentally different properties from one satisfying Eq. (4), assigning so much probability mass to extreme values that higher moments are infinite; in some instances, even the mean or variance are not finite.

In a recent study, Zwiers & Kharin (1998) used extreme value theory to compare extremes numerically simulated in a GCM control run with observations. They also compared GCM-simulated extremes under doubled carbon dioxide with those for a control run. These comparisons were generally made in terms of 'return' (or 'design') values (i.e. the value of a climate variable corresponding to the desired 'return period,' such as 10, 20, or 50 yr). In a follow-up study, Kharin & Zwiers (2000) applied essentially the same approach to transient climate simulations by a GCM. Among other things, they detected some cases in which GCM-simulated variables apparently have heavy-tailed distributions. The statistics of extremes remains under-utilized in searching for comparable trends in observed climate extremes.

Evidence of heavy-tailed distributions is starting to be uncovered, especially for geophysical variables, such as streamflow, that integrate climate or for certain impact variables, such as monetary damage, that relate to income or wealth. For example, Anderson and Meerschaert (1998) found evidence of heavy tails in modeling streamflow, and Katz (2002) treated damage associated with hurricanes (data from Pielke & Landsea 1998). Both of these variables are traditionally modeled by the lognormal distribution, and only if the extreme right-hand tail is closely examined are any substantial discrepancies found. Fig. 3 shows a comparison of the fit of the lognormal (Fig. 3a) and the generalized Pareto distribution (Fig. 3b), in terms of their survival functions (see Appendix 3 for more details about the construction of this figure). Although the evidence is not completely conclusive about whether the fit of the heavy-tailed

Pareto is superior to the medium-tailed lognormal for extremely high damages, the 2 forms of fitted distributions have quite different implications concerning what should constitute a surprising event.

Most statistical methods employed in global climate change assessments make no provision for heavy-tailed distributions. For example, all the popular stochastic weather generators (including WGEN) assume medium-tailed distributions for precipitation amounts. Under this type of assumption, the maximum value of precipitation simulated by such a weather generator cannot have a heavy tail (Katz 1999), despite evidence to the contrary for observed precipitation extremes (e.g. Buishand 1989). Although resampling techniques might be thought to circumvent this issue (Lall & Sharma 1996, Rajagopalan & Lall 1999), they actually treat any variable as being bounded by the largest value in the original sample; in effect, neither medium-tailed nor heavy-tailed.

5. RECOMMENDATIONS

One of the main points that comes out of this review is that uncertainty analysis should not be viewed as a minor component that can be 'added on' once a model has been developed. Instead, its consideration should be an integral part of the development of any model (this point has been made before, e.g. Dowlatabadi & Morgan 1993). For example, in determining what level of aggregation is appropriate in an integrated assessment, explicit treatment of uncertainty should be one of the factors taken into account. Ideally, uncertainty
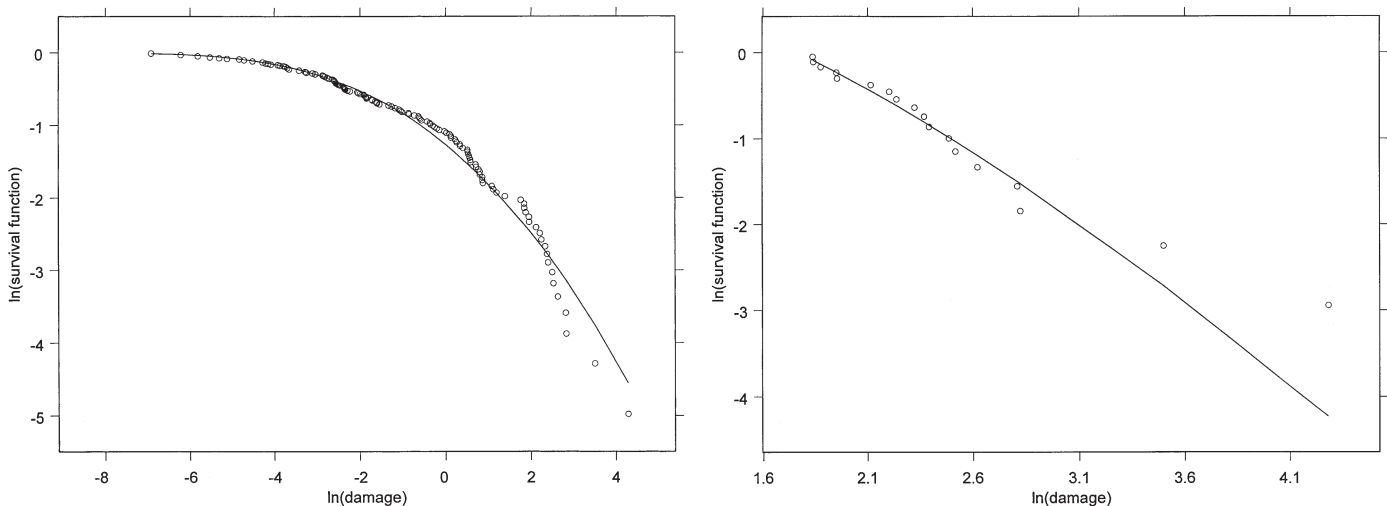

Fig. 3. Example of heavy-tailed distribution, showing logarithm of survival function vs logarithm of damage from individual hurricanes (a) for all 144 values (circles = data; line = fitted lognormal distribution), (b) for only the 18 highest damage values (circles = data; line = fitted generalized Pareto distribution) (see Appendix 3)

should be addressed even when deciding whether to increase the size or complexity of a GCM.

Uncertainty analysis that is fully probabilistic has been advocated in the present paper, taking advantage of recent computational developments in Bayesian statistics. But it is important to acknowledge that sensitivity and scenario analyses still are needed, as such analyses can serve as catalysts for further examination of uncertainties. The question remains of what strategy to adopt, both to provide useful guidance about uncertainty in the interim and to make progress as efficiently as possible toward the ultimate goal of implementing fully probabilistic analyses for global climate change models. A number of different approaches come to mind, each of which has serious limitations if taken in isolation, but when pursued in parallel might be relatively robust (akin to the 'strategic cyclical scaling' research paradigm of Root & Schneider 1995).

Broad recommendations include:

(1) *'Disintegrated' uncertainty analysis.* Although integrated assessment is the natural framework to account for uncertainty, it would be worthwhile to start by assessing uncertainty in individual model components separately.

(2) *Uncertainty analysis for simpler models.* The systematic treatment of uncertainty for simpler models may be preferable to incomplete uncertainty analysis for more complex models. Only in this way can the mechanisms by which uncertainty propagates begin to be understood.

More specific recommendations concerning uncertainty analysis, already mentioned in the present paper, include: (1) make use of scaling theory in quantifying uncertainty due to aggregation; (2) take advantage of statistical downscaling; (3) exploit statistical theory for extreme events; and (4) incorporate subjective/qualitative information via a fully Bayesian approach.

What are the prospects for the global climate change arena as uncertainty analysis is increasingly attempted and more realistically performed? In the short-term, an apparent increase in uncertainty should be anticipated, not any reduction (Morgan 1998). But recall the theme of 'the value of knowing how little you know' (Section 2). The field of climate change impact assessment will be better off in the long run the sooner it is recognized how severely underestimated uncertainty presently is. In the meantime, value of information studies (i.e. quantifying the economic value of eventual reductions in uncertainty), predicated on decision making under uncertainty, can be used as one criterion to help prioritize research (Hobbs 1997, Katz & Murphy 1997). Still it is important to be aware that this method can sometimes produce misleading results (Hammitt & Shlyakhter 1999), and that 'research prioritization cannot be based on value of information alone' (Dowlatabadi & Morgan 1993, p. 1932).

The aforementioned difficulties will hinder any attempt to perform a full-fledged uncertainty analysis of models for global climate change. Nevertheless, I close on a relatively optimistic note, with a quote from Francis Bacon (English philosopher, AD 1561–1626) giving the 'raison d'être' for uncertainty analysis:

'If a man will begin with certainties he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties.'

Appendix 1. Example of the Bayesian paradigm

The example of the Bayesian statistical paradigm follows Epstein (1985, Chap. 5). Suppose that observations (say, of temperature) $Y(t)$, $t = 1, 2, \ldots$, are generated independently and normally distributed, with unknown mean μ but known variance σ^2 [denoted by $N(\mu, \sigma^2)$]. Assume that the prior probability distribution for the mean temperature μ is $N(m', \sigma^2/n')$. Further, suppose that a sample of size n with sample mean $\bar{y}$ subsequently becomes available. Then the posterior distribution of the mean μ is $N(m'', \sigma^2/n'')$, where:

$$n'' = n + n', \quad m'' = (n/n'')\bar{y} + (n'/n'')m' \qquad (A1)$$

In Fig. 1, $\sigma = 5$, $m' = 20°C$, and $n' = 25$, so that the prior distribution of the mean temperature μ is $N(20, 1)$. For one sample outcome of $n = 25$ and $\bar{y} = 22°C$, Eq. (A1) yields a posterior distribution of μ that is $N(21, 0.5)$; for the other sample outcome of $n = 75$ and $\bar{y} = 18°C$, a posterior distribution of μ that is $N(18.5, 0.25)$

Appendix 2. Example of induced dependence

The example of how unconditional dependence can be induced from conditional independence makes use of the relationships between conditional and unconditional moments (i.e. means, variances, and covariances) as described, for example, in Katz & Parlange (1996). Consider 2 variables, X and Y, say representing the temperature at 2 sites, and let another variable I indicate the occurrence (i.e. $I = 1$) or non-occurrence (i.e. $I = 0$) of, for instance, an El Niño event, with $w = \Pr\{I = 1\}$. In addition, let the conditional means and variances of X given the state I be denoted by $\mu_X(i) = E(X \mid I = i)$ and $\sigma_X^2(i) = \mathrm{var}(X \mid I = i)$, $i = 0, 1$ [similarly, denoted by $\mu_Y(i)$ and $\sigma_Y^2(i)$ for Y].

Assume that X and Y are conditionally independent given I; in particular, conditionally uncorrelated [i.e. corr$(X, Y \mid I)$ = 0]. Then X and Y are unconditionally dependent, with unconditional correlation coefficient that can be expressed as:

corr$(X, Y) =$
$\{w(1 - w)[\mu_X(1) - \mu_X(0)][\mu_Y(1) - \mu_Y(0)]\}/\{[\mathrm{var}(X)\mathrm{var}(Y)]^{1/2}\}$
$\mathrm{var}(X) = (1 - w)\sigma_x^2(0) + w\sigma_x^2(1) + w(1 - w)[\mu_X(1) - \mu_X(0)]^2$

$$(A2)$$

[var(Y) can be expressed similarly].

In Fig. 2, the conditional distributions of X and Y given I are normal with parameters $\mu_X(0) = 20$ºC, $\mu_X(1) = 24$ºC, $\sigma_X(0) = \sigma_X(1) = 2$ºC, $\mu_Y(0) = 10$ºC, $\mu_Y(1) = 15$ºC, and $\sigma_Y(0) = \sigma_Y(1) = 3$ºC. If $w = 0.25$, then Eq. (A2) gives corr$(X, Y) = 0.383$

Appendix 3. Example of a heavy-tailed distribution

The example of evidence of a heavy-tailed distribution is taken from Katz (2002), with the original source of the data being Pielke & Landsea (1998). A lognormal distribution is one for which the logarithmically transformed data are normally distributed. Its survival function, $1 - F(x)$, satisfies Eq. (4); that is, a medium-tailed distribution. A generalized Pareto has cumulative distribution function of the form:

$$F(x) = 1 - [1 + \gamma(x - \mu)/\sigma]^{-1/\gamma} \quad x > \mu \quad \sigma > 0 \quad (A3)$$

Its survival function satisfies Eq. (8), a heavy-tailed distribution, provided the shape parameter $\gamma > 0$. Taking the limit as $\gamma \to 0$, Eq. (A3) reduces to the distribution function for the exponential, the prototypical, medium-tailed distribution (Reiss & Thomas 1997, Chap. 1).

In Fig. 3, the data consist of the damage from 144 North Atlantic tropical cyclones (primarily hurricanes) making landfall in the US during 1925–1995. The damages have been adjusted for inflation, wealth, and population at risk, and are expressed in 1995 US billion \$ (see Pielke & Landsea 1998). For Fig. 3b, the maximum likelihood estimate of the shape parameter of the generalized Pareto distribution, fitted only to the highest 18 damage values, is $\hat{\gamma} = 0.471$

LITERATURE CITED

Adams RM, Crocker TD, Katz RW (1984) Assessing the adequacy of natural science information: a Bayesian approach. Rev Econ Stat LXVI:568–575

Alcamo J, Kreileman E (1996) Emission scenarios and global climate protection. Global Environ Change 6:305–334

Alcamo J, Kreileman E, Leemans R (1996) Global models meet global policy. Global Environ Change 6:255–259

Anderson PL, Meerschaert MM (1998) Modeling river flows with heavy tails. Water Resour Res 34:2271–2280

Berliner LM, Wikle CK, Cressie N (2000) Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. J Clim 13:3953–3968

Beyth-Marom R (1982) How probable is probable? numerical translation of verbal probability expressions. J Forecast 1:257–269

Brown BG, Mearns LO, McDaniel L (1995) Climatological characteristics of spatial scaling of hourly precipitation. In: Preprints, Ninth Conference on Applied Climatology. American Meteorological Society, Boston, p 362–367

Buishand TA (1989) Statistics of extremes in climatology. Stat Neerlandica 43:1–30

Buishand TA, Brandsma T (2001) Multi-site simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. Water Resour Res 37:2761–2776

Carter TR, Parry ML, Harasawa H, Nishioka S (1994) IPCC technical guidelines for assessing climate change impacts and adaptations. UCL, London, and CGER, Tsukuba

Casella G, George EI (1992) Explaining the Gibbs sampler. Am Stat 46:167–174

Cipra B (2000) Revealing uncertainties in computer models. Science 287:960–961

Dempster AP (1998) Logicist statistics I. Models and modeling. Stat Sci 13:248–276

Dowlatabadi H (1998) Sensitivity of climate change mitigation estimates to assumptions about technical change. Energy Econ 20:473–493

Dowlatabadi H, Morgan MG (1993) Integrated assessment of climate change. Science 259:1813, 1932

Enting IG, Lassey KR (1993) Projections of future CO_2. Division of Atmospheric Research, CSIRO, Mordialloc, Technical Paper 27

Epstein ES (1985) Statistical inference and prediction in climatology: a Bayesian approach. Meteorological Monographs 20(42), American Meteorological Society, Boston

Faragó T, Katz RW (1990) Extremes and design values in climatology. Report No. WCAP-14, WMO/TD-No. 386. World Meteorological Organization, Geneva

Fischhoff B (1994) What forecasts (seem to) mean. Int J Forecast 10:387–403

Gates WL and 15 others (1999) An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). Bull Am Meteorol Soc 80:29–54

Gupta VK, Waymire EC (1993) A statistical analysis of mesoscale rainfall as a random cascade. J Appl Meteorol 32:251–267

Hammitt JK, Shlyakhter AI (1999) The expected value of information and the probability of surprise. Risk Anal 19: 135–152

Henderson-Sellers A, McGuffie K (1999) Concepts of good science in climate change modelling. Clim Change 42: 597–610

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. Clim Res 7:85–95

Hobbs BF (1997) Bayesian methods for analysing climate change and water resource uncertainties. J Environ Manage 49:53–72

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial (with discussion). Stat Sci 14:382–417

Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G (2000) Communicating statistical information. Science 290:2261–2262

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (2001) Climate change 2001: the scientific basis. Cambridge University Press, Cambridge

Hulme M, Carter TR (1999) Representing uncertainty in climate change scenarios and impact studies. In: Carter TR, Hulme M, Viner D (eds) Representing uncertainty in climate change scenarios and impact studies. Proceedings of the ECLAT-2 Helsinki Workshop, 14–16 April 1999, Climatic Research Unit, Norwich, p 11–37

Jaeger CC, Renn O, Rosa EA, Webler T (1998) Decision analysis and rational action. In: Rayner S, Malone EL (eds) Human choice and climate change, Vol 3: The tools for policy analysis. Battelle Press, Columbus, p 141–215

Jones RN (2000) Analysing the risk of climate change using an irrigation demand model. Clim Res 14:89–100

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90: 773–795

Kass RE, Wasserman L (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. J Am Stat Assoc 90:928–934

Katz RW (1992) Role of statistics in the validation of general circulation models. Clim Res 2:35–45

Katz RW (1993) Towards a statistical paradigm for climate change. Clim Res 2:167–175

Katz RW (1996) Use of conditional stochastic models to generate climate change scenarios. Clim Change 32:237–255

Katz RW (1999) Extreme value theory for precipitation: sensitivity analysis for climate change. Adv Water Resour 23: 133–139

Katz RW (2002) Stochastic modeling of hurricane damage. J Appl Meteorol (in press)

Katz RW, Murphy AH (eds) (1997) Economic value of weather and climate forecasts. Cambridge University Press, Cambridge

Katz RW, Parlange MB (1996) Mixtures of stochastic processes: application to statistical downscaling. Clim Res 7: 185–193

Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. J Clim 13:3760–3788

Lall U, Sharma A (1996) A nearest neighbor bootstrap for resampling hydrologic time series. Water Resour Res 32: 679–693

Landau S, Mitchell RAC, Barnett V, Colls JJ, Craigon J, Payne RW (2000) A parsimonious, multiple-regression model of wheat yield response to environment. Agric For Meteorol 101:151–166

Lettenmaier D (1995) Stochastic modeling of precipitation with applications to climate model downscaling. In: von Storch H, Navarra A (eds) Analysis of climate variability:

applications of statistical techniques. Springer-Verlag, Berlin, p 197–212

Lu ZQ, Berliner LM (1999) Markov switching time series modeling with application to a daily runoff series. Water Resour Res 35:523–534

Mahlman JD (1997) Uncertainties in projections of human-caused climate warming. Science 278:1416–1417

Malakoff D (1999) Bayes offers a 'new' way to make sense of numbers. Science 286:1460–1464

McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS (2001) Climate change 2001: impacts, adaptation, and vulnerability. Cambridge University Press, Cambridge

Mearns LO (1997) On the statistical evaluation of climate model experiments. Clim Change 37:443–448

Mearns LO, Rosenzweig C, Goldberg R (1996) The effect of changes in daily and interannual climatic variability on CERES–Wheat: a sensitivity study. Clim Change 32: 257–292

Mearns LO, Rosenzweig C, Goldberg R (1997) Mean and variance change in climate scenarios: methods, agricultural applications, and measures of uncertainty. Clim Change 35:367–396

Metz B, Davidson O, Swart R, Pan J (2001) Climate change 2001: mitigation. Cambridge University Press, Cambridge

Morgan MG (1998) Policy analysis for decisionmaking about climate change. In: Nordhaus WD (ed) Economics and policy issues in climate change. Resources for the Future, Washington, DC, p 25–50

Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge

Morgan MG, Keith DW (1995) Subjective judgments by climate experts. Environ Sci Technol 29:468A–476A

Morgan MG, Kandlikar M, Risbey J, Dowlatabadi H (1999) Why conventional tools for policy analysis are often inadequate for problems of global change. Clim Change 41: 271–281

Murphy AH, Lichtenstein S, Fischhoff B, Winkler RL (1980) Misinterpretations of precipitation probability forecasts. Bull Am Meteorol Soc 61:695–701

New M, Hulme M (2000) Representing uncertainty in climate change scenarios: a Monte-Carlo approach. Integr Assess 1:203–213

Nicholls N (2001) The insignificance of significance testing. Bull Am Meteorol Soc 82:981–986

Nicholls N, Kestin T (1998) Communicating climate. Clim Change 40:417–420

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. Science 263:641–646

Osborn TJ, Hulme M (1997) Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. J Clim 10:1885–1908

Palmer TN (1993) Extended-range atmospheric prediction and the Lorenz model. Bull Am Meteorol Soc 74:49–65

Parkinson S, Young P (1998) Uncertainty and sensitivity in global carbon cycle modeling. Clim Res 9:157–174

Parry ML, Carter TR (1985) The effect of climatic variations on agricultural risk. Clim Change 7:95–110

Paté-Cornell E (1996) Uncertainties in global climate change estimates. Clim Change 33:145–149

Petersen AC (2000) Philosophy of climate science. Bull Am Meteorol Soc 81:265–271

Pielke RA Jr, Landsea CW (1998) Normalized hurricane damages in the United States, 1925–95. Weather Forecast 13: 621–631

Rajagopalan B, Lall U (1999) A *k*-nearest neighbor simulator

for daily precipitation and other weather variables. Water Resour Res 35:3089–3101

Rayner S, Malone EL (eds) (1998) Human choice and climate change, Vol 3: the tools for policy analysis. Battelle Press, Columbus

Reilly J, Stone PH, Forest CE, Webster MD, Jacoby HD, Prinn RG (2001) Uncertainty and climate change assessments. Science 293:430–433

Reiss RD, Thomas M (1997) Statistical analysis of extreme values. Birkhäuser Verlag, Basel

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Riha SJ, Wilks DS, Simoens P (1996) Impact of temperature and precipitation variability on crop model predictions. Clim Change 32:293–311

Ritchie JT, Otter S (1985) Description and performance of CERES–wheat: a user-oriented wheat yield model. In: Willis WO (ed) ARS wheat yield project. ARS-38, USDA, Washington, DC, p 159–175

Root TL, Schneider SH (1995) Ecology and climate: research strategies and implications. Science 269:334–341

Rotmans J, Dowlatabadi H (1998) Integrated assessment modeling. In: Rayner S, Malone EL (eds) Human choice and climate change, Vol 3: the tools for policy analysis. Battelle Press, Columbus, p 291–377

Saltelli A, Tarantola S, Campolongo F (2000) Sensitivity analysis as an ingredient of modeling. Stat Sci 15:377–395

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Shackley S, Young P, Parkinson S, Wynne B (1998) Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools? Clim Change 38:159–205

Shackley S, Young P, Parkinson S (1999) Response to A. Henderson-Sellers and K. McGuffie. Clim Change 42:611–617

Shlyakhter AI, Kammen DM, Broido CL, Wilson R (1994) Quantifying the credibility of energy projections from trends in past data—the United States energy sector. Energy Policy 22:119–130

Skelly WC, Henderson-Sellers A (1996) Grid box or grid point: what type of data do GCMs deliver to climate impacts researchers? Int J Climatol 16:1079–1086

Stigler SM (1977) Do robust estimators work with *real* data?

(with discussion). Ann Stat 5:1055–1098

Streets DG, Glantz MH (2000) Exploring the concept of climate surprise. Global Environ Change 10:97–107

Tol RSJ, de Vos AF (1998) A Bayesian statistical analysis of the enhanced greenhouse effect. Clim Change 38:87–112

van Asselt MBA, Beusen AHW, Hilderink HBM (1996) Uncertainty in integrated assessment: a social scientific perspective. Environ Model Assess 1:71–90

von Storch H (1999) On the use of 'inflation' in statistical downscaling. J Clim 12:3505–3506

von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge

von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Webster MD, Sokolov AP (2000) A methodology for quantifying uncertainty in climate projections. Clim Change 46: 417–446

Wigley TML, Raper SCB (2001) Interpretation of high projections for global-mean warming. Science 293:451–454

Wikle CK, Berliner LM, Cressie N (1998) Hierarchical Bayesian space-time models. Environ Ecol Stat 5:117–154

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21:530–548

Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: a comparison of methods. Water Resour Res 34:2995–3008

Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, San Diego

Wilks DS (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93:153–169

Wilks DS, Wilby RL (1999) The weather generation game: a review of stochastic weather models. Prog Phys Geogr 23: 329–357

Wolf J, Evans LG, Semenov MA, Eckersten H, Iglesias A (1996) Comparison of wheat simulation models under climate change. I. Model calibration and sensitivity analysis. Clim Res 7:253–270

Zwiers FW, Kharin VV (1998) Changes in the extremes of the climate simulated by CCC GCM2 under CO_2. J Clim 11:2000–2222