Patrick Frank
5 March 2016

For Editorial Review Only

Cover Letter Supplemental

On the Understanding of Error Analysis As Exhibited by Climate Modelers

The uniform experience of the author is that climate modelers uniformly made very serious mistakes concerning physical error analysis.

The mistakes are so fundamental as to indicate no understanding of the subject.

"Propagation of Error ..." concerns physical error analysis as applied to climate models.

The manuscript has gone through submission and review at four journals. Of fifteen total reviewers, thirteen clearly were climate modelers. Not one of them evidenced any understanding of propagated error or the meaning of physical uncertainty.

In the Sections below, the following is demonstrated concerning these reviewers:

- They evinced neither respect for or understanding of the distinction between accuracy and precision.
- They showed no understanding of the meaning or method of propagated error.
- They supposed the uncertainty bars of propagated error imply increasingly wild oscillations within the model.
- They showed no understanding of physical error itself.
- They showed no understanding of a unique result or of its importance.

The body of evidence supports a general conclusion that training in climate modeling does not include physical error analysis. On that basis, it is requested that no climate modelers be included among the reviewers.

All comments below are quoted to reveal the full meaning intended by the reviewer.

1. Accuracy vs. Precision

The distinction between accuracy and precision is central to the argument presented in the manuscript. The terms are defined in the Introduction as:

- model accuracy: the difference between predictions and observations.
- model precision: the variance among predictions without reference to observations.

These are conventional definitions. (Bevington and Robinson 2003; JCGM 100:2008; Roy and Oberkampf 2011)

Physical evaluation of a model requires an accuracy metric. Therefore, the critical distinction between accuracy and precision is fundamental.

Reviewer comments on accuracy vs. precision:

1.1 "*Too much of this paper consists of philosophical rants (e.g., accuracy vs. precision) ...*"

1.2 "*[T]he author thinks that a probability distribution function (pdf) only provides information about precision and it cannot give any information about accuracy. This is wrong, and if this were true, the statisticians could resign.*"

1.3 "*The best way to test the errors of the GCMs is to run numerical experiments to sample the predicted effects of different parameters...*"

1.4 "*The author is simply asserting that uncertainties in published estimates* [i.e., model precision - P] *are not 'physically valid'* [i.e., not accuracy - P]- *an opinion that is not widely shared.*"

The first reviewer scorned the distinction between accuracy and precision. The remaining statements are alternative declarations that model variance, i.e., precision, is identical to physical accuracy. This confusion is entirely incorrect. See the distinction between "accuracy" and "sharpness" in (Wilks 1995), for example, as well as (Bevington and Robinson 2003; JCGM 100:2008; Roy and Oberkampf 2011).

The further evidence below implies these mistakes in understanding are general among climate modelers.

2. No understanding of propagated error

2.1 "*The authors claim that published projections do not include 'propagated errors' is fundamentally flawed. It is clearly the case that the model ensemble may have structural errors that bias the projections.*"

Here, the reviewer incorrectly supposed that model precision is identical to propagated physical error.

2.2 "*The repeated statement that no prior papers have discussed propagated error in GCM projections is simply wrong (Rogelj (2013), Murphy (2007), Rowlands (2012)).*"

The reviewer examples are taken in order:

Rogelj (2013) concerns the economic costs of mitigation. Their Figure 1b includes a global temperature projection plus uncertainty ranges. The uncertainties, "*are based on a*

600-member ensemble of temperature projections for each scenario..." (Rogelj et al. 2013)

That is, Rogeli (2013) present deviation from a model mean, which defines precision. Propagation of error appears nowhere in Rogelj (2013). In supposing that model precision is identical to propagated error, a lack of critical understanding was revealed.

From (Murphy et al. 2007), "*In order to sample the effects of model error, it is necessary to construct ensembles which sample plausible alternative representations of earth system processes.*"

Murphy (2007) again present model precision, which the reviewer again supposed is identical to propagated error.

From (Rowlands et al. 2012), "*Here we present results from a multi-thousand-member perturbed-physics ensemble of transient coupled atmosphere–ocean general circulation model simulations.*" They go on to state that, *"Perturbed-physics ensembles offer a systematic approach to quantify uncertainty in models of the climate system response to external forcing, albeit within a given model structure.*"

Perturbed physics ensembles present deviations around a model mean due to variation of parameter values; a measure of model precision. Perturbed physics projections are not a measure of physical accuracy, because the divergence metric is only meaningful to the internal dynamics of the model. Comparison with observations does not reveal model physical fidelity and so is not an accuracy measure.

Not one of this reviewer's examples of propagated error includes any propagated error, or even mentions propagated error.

Further, not one of the examples discusses physical error at all. They all present model precision.

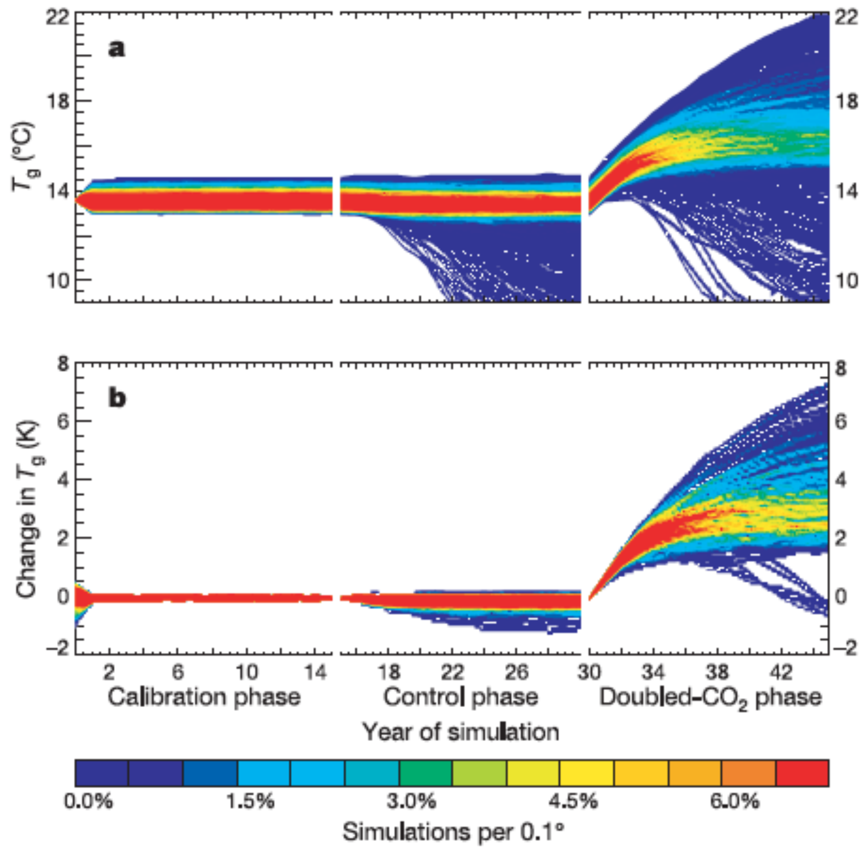
This reviewer showed no understanding of propagated error, or of what it means, or of how to identify it. This reviewer also showed no evidence of an ability to recognize physical error itself.

2.3 "*Examples of uncertainty propagation: Stainforth, D. et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433, 403-406.*

"M. Collins, R. E. Chandler, P. M. Cox, J. M. Huthnance, J. Rougier and D. B. Stephenson, 2012: Quantifying future climate change. Nature Climate Change, 2, 403-409."

These are taken in turn: (Stainforth et al. 2005) includes three Figures; Every single one of them presents error as projection variation about a model mean, i.e., model precision.

Stainforth, et al., 2005, Figure 1:



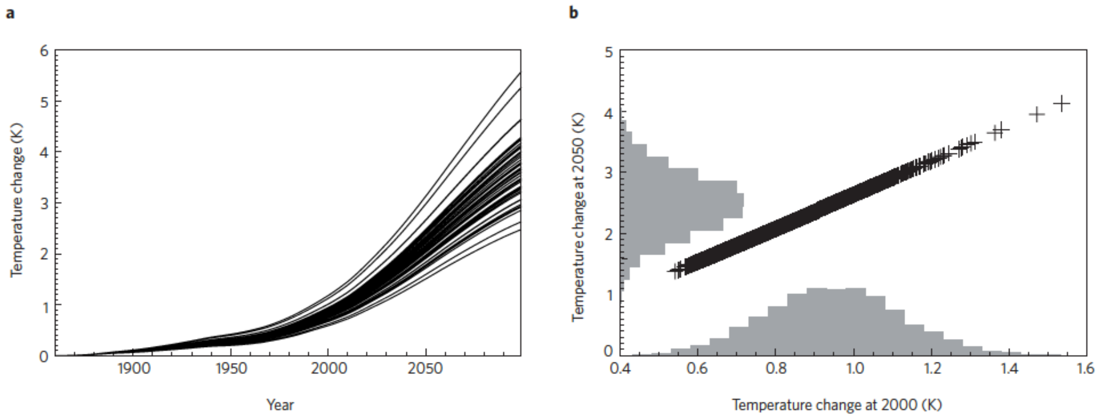
Original Figure Legend: "*Figure 1 Frequency distributions of T g (colours indicate density of trajectories per 0.1 K interval) through the three phases of the simulation.* ***a****, Frequency distribution of the 2,017 distinct independent simulations.* ***b****, Frequency distribution of the 414 model versions. In* ***b****, T g is shown relative to the value at the end of the calibration phase and where initial condition ensemble members exist, their mean has been taken for each time point.*"

Concerning uncertainty: "*[W]e have carried out a grand ensemble (an ensemble of ensembles) exploring uncertainty in a state-of-the-art model. Uncertainty in model response is investigated using a perturbed physics ensemble in which model parameters are set to alternative values considered plausible by experts in the relevant parameterization schemes.*"

That is, uncertainty is directly represented as model variability (*density of trajectories; perturbed physics ensemble*); both parenthetical metrics represent model precision.

The remaining figures in Stainforth, (2005) derive from Figure 1. Propagated error appears nowhere in Stainforth (2005) and is nowhere mentioned.

(Collins et al. 2012) state that adjusting model parameters so that projections approach observations is enough to “hope” that a model has physical validity. Propagation of error is never mentioned. Collins Figure 3, reproduced below, shows physical uncertainty as model variability about an ensemble mean.



Original Legend: "*Figure 3 | Global temperature anomalies.* ***a****, Global mean temperature anomalies produced using an EBM forced by historical changes in well-mixed greenhouse gases and future increases based on the A1B scenario from the Intergovernmental Panel on Climate Change’s Special Report on Emission Scenarios. The different curves are generated by varying the feedback parameter (climate sensitivity) in the EBM.* ***b****, Changes in global mean temperature at 2050 versus global mean temperature at the year 2000, ... The histogram on the x axis represents an estimate of the twentieth-century warming attributable to greenhouse gases. The histogram on the y axis uses the relationship between the past and the future to obtain a projection of future changes.*"

Collins 2012, Figure 3 part a: model variability itself; part b: model variability (precision) represented as physical uncertainty (accuracy). Propagated error is nowhere to be found.

Thus, neither of this reviewer's examples of propagated error actually includes any propagated error, or even mentions propagated error.

In this section, each reviewer mistook model precision as propagated error. One may safely conclude that none of these reviewers revealed any understanding of propagated error. They also apparently can not distinguish between physical error (accuracy) and projection deviance from a model mean (precision).

3. Supposition that error bars imply model oscillation

3.1 "*To say that this error indicates that temperatures could hugely cool in response to* CO_2 *shows that their model is unphysical.*"

3.2 "*[T]his analysis would predict that the models will swing ever more wildly between* *snowball and runaway greenhouse states.*"

3.3 "*Indeed if we carry such error propagation out for millennia we find that the uncertainty will eventually be larger than the absolute temperature of the Earth, a clear absurdity.*"

3.4 "*An entirely equivalent argument* [to the error bars] *would be to say (accurately) that there is a 2K range of pre-industrial absolute temperatures in GCMs, and therefore the global mean temperature is liable to jump 2K at any time - which is clearly nonsense...*"

In every case, these climate modelers proposed that "±" error bars imply the model itself is oscillating ("*swing ... wildly*," "*liable to jump"*) between the error bar extremes.

Or that the uncertainty bars resulting from propagated error represent physical temperature itself ("*hugely cool*," "*larger than the absolute temperature of the Earth, a clear absurdity.*").

There can be no more naive misunderstanding of the uncertainty bars of propagated error than these. No sophomore undergraduate in physics, chemistry, or engineering would make such a mistake.

Both the manuscript and the supporting information document explained the meaning of error bars and that they represent an ignorance width. None of these reviewers gave any evidence of having read them.

4. Unaware of the importance of a unique result

Scientific prediction requires deduction from theory of a unique (tightly bounded) observable. Tightly bounded deductions from physical theory represent predictions. Theory falsifiability requires a unique prediction. These ideas combined with that of replicable observations are the scientific method.

The comments below show no understanding of these concepts.

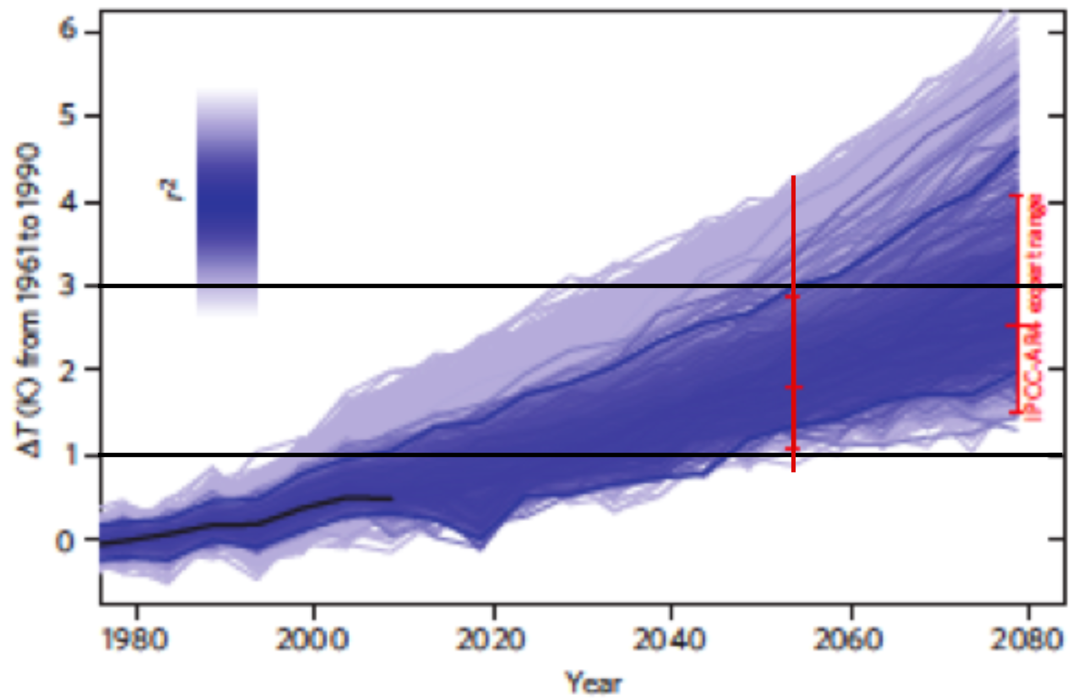
4.1 "*[L]ooking the last glacial maximum, the same models produce global mean changes of between 4 and 6 degrees colder than the pre-industrial. If the conclusions of this paper were correct, this spread (being so much smaller than the estimated errors of +/- 15 deg C) would be nothing short of miraculous.*"

4.2 "*In reality climate models have been tested on multicentennial time scales against paleoclimate data (see the most recent PMIP intercomparisons) and do reasonably well at simulating small Holocene climate variations, and even glacial-interglacial transitions. This is completely incompatible with the claimed results*."

4.3 "*The most obvious indication that the error framework and the emulation framework* *presented in this manuscript is wrong is that the different GCMs with well-known different cloudiness biases (IPCC) produce quite similar results, albeit a spread in the climate sensitivities.*"

Note that in 4.1, the reviewer again mistakenly conflates an uncertainty bar with a physical temperature.

However, (Rowlands et al. 2012) demonstrate the source of the reviewers' confidence, and that it is misplaced.



Original Legend: "*Figure 1 | Evolution of uncertainties in reconstructed global-mean temperature projections under SRES A1B in the HadCM3L ensemble.*" (Rowlands et al. 2012)

The variable black line in the middle of the group represents the observed air temperature. The horizontal black lines at 1 K and 3 K, and the vertical red line at year 2055 were added by the author. Part of the red line is the precision uncertainty bar in the original figure.

Rowlands, 2012, Figure 1 displays thousands of perturbed physics simulations of global air temperatures. Each member of the ensemble is of equivalent evidential weight. None of them are known to be physically more correct than any of the others. Closeness to observation in this case, does not imply model accuracy (see below).

The physical energy-state of the simulated climate varies systematically across the years. The horizontal black lines show that multiple physical energy states produce the same simulated 1 K or 3 K anomaly temperature.

The vertical red line at year 2055 shows that the identical physical energy-state (the year 2055 state) produces multiple simulated air temperatures.

These considerations demonstrate that climate models are not able to provide a unique solution to the problem of the climate energy-state.

Further, these wandering projections do not represent natural variability. They represent how parameter magnitudes varied across their uncertainty ranges affect the temperature simulations of the HadCM3L model itself.

The Figure fully demonstrates that climate models are incapable of producing a unique solution to any climate energy-state.

This result means that simulations close to observations are not known to accurately represent the true physical energy-state of the climate. They just happen to have opportunistically off-setting errors.

In turn, that means the projections have no informational value. They tell us nothing about possible future air temperatures.

There is no way to know which of the simulations actually represents the correct underlying physics. Or whether any of them do. And even if one of them happens to conform to future climate observables, there is no way to know it wasn't a fortuitous accident.

Even if Rowlands, et al., 2012 tuned the parameters of the HADCM3L model so that it precisely reproduced the observed air temperature line, it would not suddenly attain the ability to accurately produce a unique solution to the climate energy-state.

Tuned parameters merely obscure uncertainty. They hide the unreliability of the model. It is no measure of accuracy that tuned models produce similar projections. Or that their projections are close to observations. Tuning parameter sets merely off-sets errors and produces a false precision.

Models with large parameter uncertainties can not produce a unique prediction. The confident statements of these reviewers concerning model projections show they have no understanding of a unique prediction, or of why it is important.

The Holocene or Glacial-era temperature hindcasts referenced by the reviewers above are likewise non-unique. Not one of them validate the accuracy of a climate model. Not one of them tell us anything about any physically real global climate energy-state. Not one single climate modeler reviewer evidenced any understanding of a unique result as the basic and necessary standard of science.

5. An Example of multiplexed mistakes

A final more extended example shows how the lack of understanding of error and error analysis produced a multiplex of mistakes and a journey into the wilderness.

"*I will give (again) one simple example of why this whole exercise* [i.e., the manuscript analysis -- P] *is a waste of time. Take a simple energy balance model, solar in, long wave out, single layer atmosphere, albedo and greenhouse effect. i.e. sigma Ts^4 = S (1-a) /(1 - lambda/2) where lambda is the atmospheric emissivity, a is the albedo (0.7), S the incident solar flux (340 W/m^2), sigma is the SB coefficient and Ts is the surface temperature (288K).*

"The sensitivity of this model to an increase in lambda of 0.02 (which gives a 4 W/m2 forcing) is 1.19 deg C (assuming no feedbacks on lambda or a). The sensitivity of an erroneous model with an error in the albedo of 0.012 (which gives a 4 W/m^2 SW TOA flux error) to exactly the same forcing is 1.18 deg C.

"This the difference that a systematic bias makes to the sensitivity is two orders of magnitude less than the effect of the perturbation. The author's equating of the response error to the bias error even in such a simple model is orders of magnitude wrong. It is exactly the same with his GCM emulator."

The "*difference*" the reviewer referenced is 1.19 C - 1.18 C = 0.01 C. The reviewer supposed that this 0.01 C difference is the entire uncertainty produced by the model due to a 4 Wm^{-2} bias error in either albedo or emissivity.

However, the reviewer is multiply mistaken.

First reviewer mistake: If 1.19 C or 1.18 C are produced by a 4 Wm^{-2} bias forcing error, then 1.19 C or 1.18 C are the consequent bias temperature errors. They are not sensitivities. Their tiny difference, if anything, confirms the uniform consequence of the 4 Wm^{-2} error magnitude.

Second mistake: The reviewer mistook a bias error (a physically meaningful statistic) for a temperature (a thermodynamic magnitude). What the reviewer called "*sensitivity*" is actually the model bias error accruing to the 4 Wm^{-2} bias forcing error.

Third mistake: The reviewer has mistakenly represented the manuscript ±4 W/m^2 physical error statistic as a 4 W/m^2 energetic perturbation.

As an aside, this mistake shows that this reviewer apparently does not know to distinguish between a physical magnitude and an error statistic.

Fourth mistake: The reviewer compared a single step "sensitivity" calculation to multi-step propagated error.

Fifth mistake: The reviewer is apparently unfamiliar with the generality that root square physical uncertainties express a bounded range of ignorance; i.e., "±" about some value. Such statistical uncertainties and physical error ranges are never constant offsets.

Lemma to five: the reviewer apparently also did not know that the correct way to express the uncertainties is ±lambda range or ±albedo range.

If the uncertainties had been correctly expressed, the prescribed forcing uncertainty would inevitably have been ±4 W/m^2, not a constant 4 Wm^{-2} bias. The physical error statistic as an energy malapropism would have thus been avoided.

Following from that, the "*sensitivity*" of 1.18 C or 1.19 C would have necessarily become ±1.18 C or ±1.19 C, which obviously are not temperatures.

The very form of the error as ±4 Wm^{-2} should have been a sufficient warning of its statistical meaning, because no energetic perturbation can be simultaneously positive and negative.

When the reviewer's example is expressed using the correct ± statistical notation, 1.19 C and 1.18 C become ±1.19 C and ±1.18 C.

Finally, these are uncertainties ***for a single step calculation***. They are similar in magnitude to the single-step ±4 Wm^{-2} uncertainties presented in the manuscript.

As soon as the reviewer's forcing uncertainty enters into a multi-step linear extrapolation, i.e., a GCM projection, the ±1.19 C and ±1.18 C uncertainties would arise within every single step. They must then propagate through the steps as the root-sum-square. (Bevington and Robinson 2003; JCGM 100:2008)

After 100 annual steps (a centennial projection) the reviewer's ±1.18 C per-step error propagates to a ±11.8 C uncertainty in projected air temperature. This result is comparable with that from the manuscript analysis.

That is, correctly done, the reviewer's own analysis validates the very manuscript that the reviewer dismissed.

This reviewer:

- evidently does not know the meaning of physical uncertainty.
- did not distinguish between model response (sensitivity) and model error. This mistake amounts mistaking a physical error statistic for an energetic perturbation.
- evidently did not know how to express a physical uncertainty.
- evidently did not know the difference between single step error and propagated error.

In conclusion, the uniformity of mistakes in all the prior review comments above put the training of climate modelers on display. They evidently and uniformly:

- neither respect nor understand the distinction between accuracy and precision.
- are entirely ignorant of propagated error.
- think the ± bars of propagated error mean the model itself is oscillating.
- have no understanding of physical error.
- have no understanding of the importance or meaning of a unique result.

This evidence supports a conclusion that climate modelers are not equipped to review a manuscript concerned with physical error analysis.

For this reason, it is requested that no climate modelers be included among the reviewers.

References:

Bevington, P. R., and D. K. Robinson, 2003: *Data Reduction and Error Analysis for the Physical Sciences.* 3rd ed. McGraw-Hill, 320 pp.
Collins, M., R. E. Chandler, P. M. Cox, J. M. Huthnance, J. Rougier, and D. B. Stephenson, 2012: Quantifying future climate change. *Nature Clim. Change*, **2,** 403-409.
JCGM, 100:2008: Evaluation of measurement data — Guide to the expression of uncertainty in measurement Document produced by Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1).
Murphy, J. M., B. B. B. Booth, M. Collins, G. R. Harris, D. M. H. Sexton, and M. J. Webb, 2007: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365,** 1993-2028.
Rogelj, J., D. L. McCollum, A. Reisinger, M. Meinshausen, and K. Riahi, 2013: Probabilistic cost estimates for climate change mitigation. *Nature*, **493,** 79-83.
Rowlands, D. J., and Coauthors, 2012: Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geosci*, **5,** 256-260.
Roy, C. J., and W. L. Oberkampf, 2011: A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Comput. Methods Appl. Mech. Engineer.*, **200,** 2131-2144.
Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433,** 403-406.
Wilks, D. S., Ed., 1995: *Statistical Methods in the Atmospheric Sciences*. Vol. 59, Academic Press, 464 pp.