

Measurements, Models, and Hypotheses in the Atmospheric Sciences



David A. Randall* and Bruce A. Wielicki[†]

ABSTRACT

Measurements in atmospheric science sometimes determine universal functions, but more commonly data are collected in the form of case studies. Models are conceptual constructs that can be used to make predictions about the outcomes of measurements. Hypotheses can be expressed in terms of model results, and the best use of measurements is to falsify such hypotheses. Tuning of models should be avoided because it interferes with falsification. Comparison of models with data would be easier if the minimum data requirements for testing some types of models could be standardized.

1. Introduction

The atmospheric science community includes a large and energetic group of researchers who devise and carry out measurements in the atmosphere. This work involves instrument development, algorithm development, data collection, data reduction, and data analysis. The data by themselves are just numbers. To make physical sense of the data, some sort of model is needed. This might be a qualitative conceptual model, or it might be an analytical theory, or it might take the form of a computer program.

Accordingly, a community of modelers is hard at work developing models, performing calculations, and analyzing the results by comparison with data. The models by themselves are just “stories” about the atmosphere. In making up these stories, however, modelers must strive to satisfy a very special and rather daunting requirement: the stories must be true, as far

as we can tell; in other words, the models must be consistent with all of the relevant measurements.

This essay deals with the relationships between models and measurements in the atmospheric sciences. In view of our own backgrounds and interests, we emphasize cloud and dynamical processes. The tone of the essay is playful, and we make no attempt to be rigorous or comprehensive, but the subject is important and is not often explicitly discussed.

2. Measurements

It is useful to distinguish two kinds of measurements.

a. Measurements of universal functions

Sometimes we measure a number or function that is believed, on theoretical grounds, to be universal, so that in principle it should only have to be measured once. The surface-layer similarity functions of Monin and Obukhov (e.g., Monin and Yaglom 1971) are among the few good examples of this type of measurement in atmospheric science. Note that these functions describe certain statistics of the turbulent surface layer rather than the outcomes of measurements made at single points in space or single instants in time. Additional examples of universal functions and/or constants, which are of interest to a community that encompasses but goes far beyond atmospheric science, include the physical properties of water and air

*Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado.

[†]Radiation Sciences Branch, Atmospheric Sciences Division, NASA Langley Research Center, Hampton, Virginia.

Corresponding author address: David A. Randall, Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.

E-mail: randall@redfish.atmos.colostate.edu

In final form 30 August 1996.

©1997 American Meteorological Society

(such as the saturation vapor pressure of water vapor as a function of temperature) and the optical properties of water vapor, carbon dioxide, and ozone.

b. Case studies

Most of the time, we make measurements in a case-study mode, accumulating data on particular sequences of atmospheric events, which can be compared in detail with simulations of the same cases. Operational weather forecasting makes use of this type of measurement, and most atmospheric measurement programs such as the First ISCCP (International Satellite Cloud Climatology Project) Regional Experiment (FIRE), the Atmospheric Radiation Measurements Program (ARM), and the Tropical Ocean–Global Atmosphere Coupled Ocean–Atmosphere Experiment (TOGA COARE) also fit into this category. The simplest application of case-study measurements is to quantitatively document and/or describe what nature is doing. This is particularly interesting when the measurements deal with physical situations or physical variables that lie outside the range of previous measurements. For example, over the past several decades satellites have provided data on the global distribution of cloudiness (e.g., Schiffer and Rossow 1983) and the effects of clouds on the earth's radiation budget (e.g., Ramanathan et al. 1989). Of course, efforts are being made to understand, model, and interpret these data, but the first order of business has been simply to piece together a quantitative description of the geographical, seasonal, and interannual variations of cloudiness and the effects of clouds on the longwave and shortwave radiation at the top of the atmosphere.

Occasionally, data that have been collected in a case-study mode can definitively show whether an important idea is right or wrong. This rarely happens in the atmospheric sciences, but once in a while it does occur. For example, Matsuno (1966) theoretically predicted the existence in the atmosphere of mixed Rossby gravity waves and Kelvin waves. A short time later, both types of waves were discovered in the data (Yanai and Maruyama 1966; Wallace and Gousky 1968).

3. Models

A model essentially embodies a theory; this is true even for numerical models. A model (or a theory) provides a basis for making predictions about the outcomes of measurements. Atmospheric models can be

conceptually grouped in various ways; one such classification follows.

a. Elementary models

The disciplines of fluid dynamics, radiative transfer, atmospheric chemistry, and cloud microphysics all make use of models that are essentially direct applications of basic physical principles to phenomena that occur in the atmosphere. Many of these “elementary” models were developed under the banners of physics and chemistry, but some are products of the atmospheric science community. Elementary models tend to deal with microscale phenomena (e.g., the evolution of individual cloud droplets suspended in or falling through the air, or the optical properties of ice crystals), so that their direct application to practical atmospheric problems is usually thwarted by the sheer size and complexity of the atmosphere. Because of their generality, elementary models often predict the existence of universal constants or functions.

b. Forecast models

A model that predicts the deterministic evolution of the atmosphere or some macroscopic portion of it can be called a “forecast model.” A forecast model could be, as the name suggests, a model that is used to conduct weather prediction, but there are other possibilities; for example, it could be used to predict the deterministic evolution of an individual turbulent eddy. Forecast models can be tested against real data, documenting, for example, the observed development of a synoptic-scale system or the observed growth of an individual convective cloud, assuming of course that the requisite data can be collected.

We are often interested in computing the statistics of some atmospheric phenomenon, for example, the statistics of the general circulation. It is now widely known that there are fundamental limits on the deterministic predictability of the atmosphere due to sensitive dependence on initial conditions (e.g., Lorenz 1969). For the global-scale circulation of the atmosphere, the limit of predictability is thought to be on the order of a few weeks, but for a cumulus-scale circulation it is on the order of a few minutes. For timescales longer than the deterministic limit of predictability for the system in question, only the statistics of the system can be predicted. These statistics can be generated by brute-force simulation, using a forecast model but pushing the forecast beyond the deterministic limit and then computing statistics from the results. The obvious and most familiar example

is simulation of the atmospheric general circulation (e.g., Smagorinski 1963). Additional examples are large-eddy simulations of atmospheric turbulence (e.g., Moeng 1984) and simulations of the evolution of an ensemble of clouds using time and space domains much larger than the time and space scales of individual clouds (e.g., Krueger 1988). Monte Carlo simulations of radiative transfer (e.g., McKee and Cox 1974) can be said to “forecast” the paths of individual photons, and statistics are computed from a large number of such paths; we classify these Monte Carlo radiative transfer models as “forecast models” for purposes of the present discussion.

The statistics predicted by these various models can be compared with statistics based on atmospheric measurements, most commonly with measurements collected in the case-study mode. For example, statistics computed from the global meteorological observing network can be used to compute general circulation statistics, and these can then be compared with statistics computed from simulations of the general circulation.

As a second example, statistics computed from aircraft data can be compared with the results of large-eddy simulation models. These models are being used to predict universal functions that arise in connection with atmospheric turbulence (e.g., Moeng and Wyngaard 1986, 1989). Through such applications, the large-eddy models are reaching or at least aiming for a stature comparable to that of elementary models. Analogous applications of general circulation models can be imagined but few have been reported up to this time.

Forecast models are now also being used to make predictions of the time evolution of the statistics of the weather, far beyond the limit of deterministic predictability for individual weather systems. Examples are seasonal weather forecasts, which deal with the statistics of the weather rather than day-to-day variations of the weather and are now being produced by several operational centers, and climate change forecasts, which deal with the evolution of the climate over the coming decades and longer. In the case of seasonal forecasting, the predictability of the statistics of the atmospheric circulation beyond the two-week deterministic limit arises primarily from the predictability of the sea surface temperature, which has a much longer memory of its initial conditions than does the atmosphere. In the case of climate change predictions, the time evolution of the statistics of the climate system are predictable to the extent that they are driven by predictable changes in some external forcing. For

example, projected increases in greenhouse gas concentrations represent a time-varying external forcing whose effects on the time evolution of the statistics of the climate system may be predictable. Over the next several decades measurements will make it very clear to what extent these predictions are right or wrong, and many readers of this paper will live long enough to learn the results of these interesting measurements. A more mundane example is the seasonal cycle of the atmospheric circulation, which represents the response of the statistics of the atmospheric general circulation to the movement of the earth in its orbit; because the seasonal forcing is predictable many years in advance,¹ the seasonal cycle of the statistics of the atmospheric circulation is also highly predictable, far beyond the two-week limit of deterministic predictability for individual weather systems.

Over the past decade, an interesting development has occurred: Statistics are being computed from the archived forecasts of numerical weather prediction models. This represents an important new way in which the models can be used to learn about the atmosphere. Because the archived forecasts are initialized with real data and fall within the range of deterministic predictability, they are by no means pure simulations or “model products,” but at the same time they do contain errors that arise in part from weaknesses in the formulation of the model and/or the data-analysis scheme. Archived forecasts thus represent a middle ground between pure analyses of data and pure simulations of general circulation statistics. For example, general circulation statistics can be computed from an ensemble of n -day forecasts, and compared with observed (or “zero-day forecast”) general circulation statistics. Archived forecast datasets are being applied to studies of atmospheric predictability, to investigate the systematic errors of forecast models (e.g., Lorenz 1982; Heckley 1985; Arpe and Klinker 1986).

c. Models that simulate statistics directly

Most radiative transfer models describe the statistical behavior of extremely large numbers of photons. “Higher-order closure models” have been developed to simulate directly the statistics of small-scale atmospheric turbulence (e.g., Mellor and Yamada 1974). Analogous models for direct simulation of the statistics of the large-scale circulation of the atmosphere may be possible (e.g., Green 1970). These are ex-

¹We can say without exaggeration that it is “known.”

amples of models that predict statistics directly; the dependent variables are the statistics themselves, and there is no need to average the model results to generate statistics after the fact. The predictions of such models can be compared with statistics computed from measurements made in case-study mode. Models that predict statistics directly can, in principle, predict the forms of universal functions.

Conventional wisdom holds that the results of models that predict statistics directly are less reliable than statistics computed indirectly from forecast model results, so that, for example, large-eddy simulation results are generally considered more reliable than the results of higher-order closure models. The reason is that higher-order closure models generally contain more empirical parameters than large-eddy models. The role of such empirical parameters is discussed later.

d. Toy models

We also build highly idealized models that are not intended to provide quantitatively accurate or physically complete descriptions of natural phenomena, but rather to encapsulate our physical understanding of a complex phenomenon in the simplest and most compact possible form, as a kind of modeler's haiku. For example, North (1975) discusses the application of this approach to climate modeling. Toy models are intended primarily as educational tools; the results that they produce can be compared with measurements only in qualitative or semiquantitative ways.

A hypothesis is a prediction, based on a model or theory, of the outcome of a measurement. As discussed in the next section, modelers need measurements first and foremost to test hypotheses. Of course, modelers also need measurements to suggest ideas about what is important and interesting, and what sorts of model output to look at.

4. Hypotheses

Consider the following two statements.

- 1) Such and such happens always under the following specific conditions.
- 2) Such and such happens sometimes under conditions something like these.

It is an unfortunate fact that, in atmospheric science, "hypotheses" often take the form of statement num-

ber 2. Usually this represents a watered-down version of an earlier hypothesis with the much more acceptable form of statement number 1.

As mentioned above, a hypothesis is supposed to predict the outcome of a measurement. Statement number 1 is a genuine hypothesis in this sense. It can be proven wrong or "falsified." It can never be proven right, because there is always the possibility that some future measurement will be inconsistent with the hypothesis (Popper 1959). Statement number 2 does not firmly predict the outcome of a measurement, so it is not really a hypothesis at all.

As an example, it was suggested by Randall (1980) that when a specific criterion is satisfied the process of cloud-top entrainment instability (CTEI) leads to rapid entrainment and a decrease in stratiform cloud amount. Based on this idea, a "CTEI hypothesis" was formulated, namely, that CTEI occurs on the equatorward sides of the subtropical marine stratocumulus sheets and that as a result of CTEI the stratiform cloud gives way to broken cumuli. Although some laboratory experiments have apparently produced a physical process analogous to CTEI (Siems et al. 1990), and although some numerical models have apparently simulated CTEI (e.g., Moeng et al. 1995), atmospheric measurements have yet to document a bona fide case of it. As a result, the rapidly dwindling number of scientists who are still willing to entertain the CTEI hypothesis at all are now reduced to hypothesizing that CTEI happens sometimes, under conditions that are perhaps not very clearly defined.

As a hypothesis, this fallback position, which has the form of statement number 2 above, is not very satisfactory because it does not predict the outcome of a specific measurement, and so it is not falsifiable. A better hypothesis, and one that is becoming increasingly popular, is that CTEI never happens. This could be falsified by a single set of measurements.²

A hypothesis can take the form of a prediction of the observed statistics of a system. For example, we might hypothesize that when a coin is flipped, "heads" will come up half the time in a sufficiently large sample of measurements.³ Hypotheses about the statistics of the atmosphere are quite common in many branches of atmospheric science.

Is a model a hypothesis? When a model is used to

²Don't hold your breath.

³If this hypothesis proves to be false, then we may formulate a new hypothesis about the nature of the coin.

perform a calculation, it yields a prediction about the behavior of the atmosphere. To phrase this in terms of hypotheses, we can say that “we hypothesize that the prediction produced through calculations with the model is true.” Atmospheric measurements can (or should) be able to falsify this hypothesis. If a prediction produced by a model is shown to be in conflict with measurements, then the model itself can be said to have been falsified.

Figure 1 summarizes the preceding discussion of the relationships among measurements, models, and hypotheses. The white boxes in the figure represent measurements or statistics computed from measurements. The gray boxes represent models or products generated by models. The thick lines represent hypotheses, which bring together models and measurements to allow scientifically useful conclusions to be drawn.

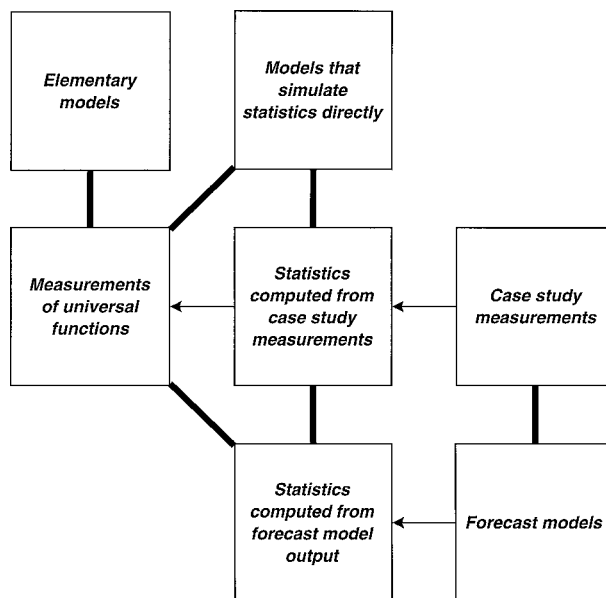


FIG. 1. Diagram illustrating the relations between measurements and models. The white boxes represent measurements or statistics computed from measurements. The gray boxes represent models or products generated by models. The thick lines represent hypotheses, which bring together models and measurements.

5. Tuning

When modelers discuss their needs for data, they rarely speak in terms of falsification. Instead, they often invoke certain buzzwords of the field.

- “Validation.” Modelers say that they need data to “validate” their models. Taken literally, this means that they need data to prove that their models are “right.” As discussed above, it is both possible and useful to falsify a model, but it is impossible to prove that a model is right. An interesting discussion of this and related points is given by Oreskes et al. (1994).
- “Tuning.” Tuning consists of adjusting coefficients in a model to improve the agreement between the model results and measurements. Modelers sometimes say that they need data so that they can tune their models. A more detailed discussion of tuning is given below.
- “Calibration.” Modelers sometimes say that they need data to “calibrate” their models. Model calibration is the same as tuning, except that for reasons to be explained in the next section, the word tuning has certain negative connotations, while calibration has positive connotations. We all appreciate that instrument calibration is a good thing. Surely, then, model calibration is also a good thing, unlike, for example, tuning.⁴

All physical parameterizations have some empirical components. “Good empiricism” is

- Directly based on observations. Parameterizations are sometimes described as “empirical,” even though no data have been used to generate them. In such a case, the parameterization is one that its creator intuitively feels is reasonable but does not know how to justify otherwise. An important point is that some parameterizations contain what may be called empirical quantities that are not directly measurable even in principle. The values of such parameters can only be set by tuning, as discussed below.
- Universally applicable. Ideally, the empirical content of a model should not be case dependent; it should apply universally. Although this goal can be difficult to meet, there are a few well-known examples of universal empirical formulas in the atmospheric science literature. Among the best examples are the Monin–Obukhov similarity functions, which were mentioned earlier. Dimensional analysis tells us that the Monin–Obukhov similarity functions depend only on certain dimensionless combinations. The functions can be and have been determined empirically, and there is good reason to believe that they are the same on Mars as they are in Kansas.

⁴This is called spin control.

- Applied before the model is run. The empirical parameters of a model should be measured and then set, on the basis of these measurements, before the model is used to make a prediction. The parameters should not be adjusted a posteriori to improve the agreement between the model predictions and other data.

Tuning consists of not following the last precept above, that is, it involves adjusting parameters after a model is run to improve the agreement between the model results and data. Tuning is bad empiricism. Calibration is bad empiricism with a bag over its head.

The problem with tuning is that it artificially prevents a model from producing a bad result. As discussed above, the most scientifically valuable thing that can come out of a comparison of measurements with model results is to show that the model has failed.

It might be argued that tuning is justified in the service of numerical weather prediction, because a good forecast is an end in itself, regardless of how it has been obtained. The trouble with this “end justifies the means” argument is that in the long run better scientific understanding is the key to making better forecasts. As mentioned earlier, the forecast archives of the operational numerical weather prediction centers represent an enormously valuable resource for scientific research and, in particular, for evaluating the parameterizations that are used in the forecast models. The most valuable cases are the bad forecasts in which a parameterization has failed to perform in agreement with the observations. Tuning the parameterizations of a forecast model interferes with research progress to the extent that it eliminates or obscures such cases.

Figure 2 summarizes the role of tuning in various situations that arise in the difficult life of a typical modeler. Suppose that a particular process has been included in a model through a parameterization that has some empirical content. The horizontal axis represents a measure of the importance of the process, and the vertical axis represents a measure of how well the process is understood.

If an important process is well understood, there certainly is no problem, and tuning is not needed. Even if such a process is not very well understood, there is no need to include it by tuning, since it is not very important anyway. These are the entries on the left side of the diagram.

In contrast, if a process is very important but poorly understood, then there may be no choice but to tune

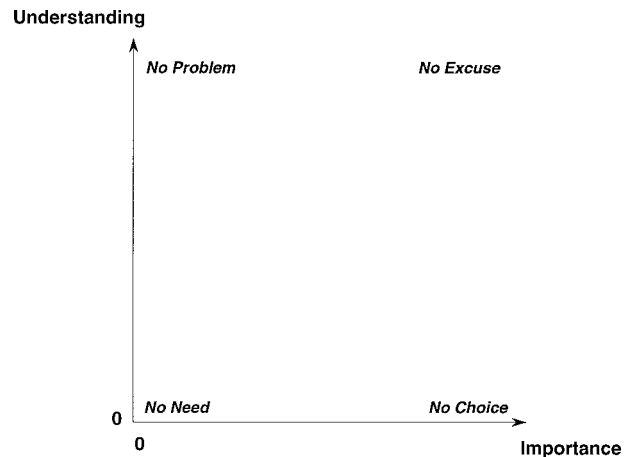


FIG. 2. A diagram illustrating the role of tuning in a model, according to the importance of the process and the degree to which it is understood.

the model. For example, when the first general circulation models (GCMs) were built in the 1960s, it was recognized that the surface fluxes of sensible heat, moisture, and momentum are crucially important for a successful simulation of the general circulation of the atmosphere, but very little was known about how these fluxes could be parameterized. There was no choice but to tune the boundary-layer parameterizations of the early GCMs. As a second example, many GCMs now include simple parameterizations of stratiform cloud microphysics because of a growing appreciation of the great importance of microphysical processes for climate. Unfortunately, however, our understanding of these processes, particularly as they operate on large scales, is very inadequate. As a result, there is at present no choice but to tune the parameters of the cloud microphysics parameterizations. For example, Fowler and Randall (1996) discuss the need to tune the autocorrelation threshold of a bulk microphysics parameterization used in a GCM. As our understanding of cloud microphysics improves over time, we will gradually move from the lower right-hand corner of the tuning diagram toward the upper right-hand corner. If we eventually arrive at a good understanding of how to parameterize cloud microphysics for GCMs, then there will be no excuse for continued tuning. Such a migration should be a key goal of both modelers and observationalists.

Many real parameterizations fall somewhere in the middle of the tuning diagram, so that decisions about whether or not to tune are less straightforward than in the extreme limiting cases discussed above.

One implication of the above discussion is that although empiricism will always be a necessary part of parameterization, tuning is not necessary, at least in principle. Genuinely empirical parameters, that is, those that can be measured and are universally applicable, can be set once and for all before a model is run.

We often hear it said in seminars or informal conversations that “such and such a model has been tuned” or “so and so must have tuned the model in order to get such good agreement.” These accusations are cheap shots; they are very easy to make and very difficult to refute. The fact is that there are many models that contain little or no tuning in the sense defined above. To actually demonstrate that a particular model has not been tuned, however, requires a detailed and exhaustive examination of the model’s formulation. Carrying through such an examination requires patience and expertise; accusing a modeler of tuning requires neither.

6. Modeler-friendly data

Emerging new classes of numerical models, including single-column models (SCMs) and cloud ensemble models (CEMs), are well suited to testing cloud parameterizations against field data (Randall et al. 1996). These measurements provide a basis for case studies.

There are certain minimum requirements for datasets suitable for use with SCMs and CEMs. These include such difficult-to-measure quantities as large-scale vertical motion and advective tendencies of temperature and moisture. The planning of field programs typically involves discussions of how these quantities can be determined from the data collected. Recent technological advances such as wind profilers have improved the situation considerably, but problems remain. Operational analysis–assimilation products can be used, but caution is needed because such “data” are influenced by the parameterizations of the forecast model used.

Many kinds of data can be used with SCMs and CEMs, but first the data must be suitably reduced. It is not realistic to expect modelers to deal with raw radar data or raw satellite data or raw aircraft data; with few exceptions, modelers lack the expertise to perform such tasks and, in any case, most modelers would rather spend their time modeling than reducing and quality controlling someone else’s dataset.

These are facts of life. It follows that modelers rely on observationalists to manipulate measurements into a “ready-to-eat” form that the modelers can easily use. It is also a fact of life, however, that most observationalists would much prefer to spend their time collecting more data, devising better measurement techniques, etc., rather than preparing reduced products for the convenience of modelers.

This situation may improve if we can develop a somewhat standardized format for data reduction. Data arranged in such a format would be readily usable by modelers using SCMs or CEMs in a case-study mode. The adoption of such a standardized format for the minimum data requirements of SCMs and CEMs would provide guidance to observationalists as to “what modelers really want” and would make the goal of providing what modelers want appear more achievable. If a number of such standardized datasets can be built up over time, they can become benchmarks that modelers routinely use to evaluate their models. We think that the atmospheric science community can and should move in this direction.

In addition, there is a trend in the modeling community to produce model-output “diagnostics” that are designed to be directly comparable to data. For example, current satellite remote-sensing algorithms can only detect the uppermost cloud layer, so that an appropriate model diagnostic is the highest simulated cloud top, or suitable statistics thereof.

7. Summary

Scientifically useful comparison of atmospheric data with model results is a very challenging task. We have outlined a classification scheme in which there are two kinds of measurements. The type that is encountered most often lends itself to case studies. Models are conceptual constructs that can be used to make predictions about the outcomes of measurements. Hypotheses can be expressed in terms of model results. The most fundamental use of measurements is to falsify such hypotheses, and thereby to falsify models, not to validate them.

Tuning is the practice of adjusting parameters after a model is run, to improve the agreement between the model results and data. Although parameterizations will always have some empirical content, it is both possible and desirable to avoid tuning, except when our understanding of some important process is weak.

Single-column models and cloud ensemble models offer new opportunities for rapid progress in the development and testing of parameterizations for large-scale models, by providing new ways in which to combine models and measurements. The minimum data requirements of such models could be standardized. In addition, model diagnostics can be tailored for comparison with observables. These developments would allow more fruitful interactions between modelers and observationalists.

Acknowledgments. This paper is based in part on a talk by D. Randall, given as an introductory overview on the topic of “Cloud Measurements and Models,” which was the theme of a workshop held at the Environmental Technology Laboratory (ETL) of the National Oceanographic and Atmospheric Administration, in Boulder, Colorado, during November 1995, and jointly organized by ETL and the Department of Atmospheric Science at Colorado State University. Some parts of the paper are loosely drawn from discussion sessions held during the workshop. The work of S. Frisch and W. Schubert in organizing the workshop is gratefully acknowledged. Dr. Xubin Zeng of the University of Arizona made useful comments. Support has been provided by the National Aeronautics and Space Administration under Grant NAG1-1701 to Colorado State University.

References

- Arpé, K., and E. Klinker, 1986: Systematic errors of the ECMWF operational forecasting model in mid-latitudes. *Quart. J. Roy. Meteor. Soc.*, **112**, 181–202.
- Fowler, L. D., and D. A. Randall, 1996: Liquid and ice cloud microphysics in the CSU general circulation model. Part III: Sensitivity tests. *J. Climate*, **9**, 561–586.
- Green, J. S. A., 1970: Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **96**, 157–185.
- Heckley, W. A., 1985: Systematic errors of the ECMWF operational forecasting model in tropical regions. *Quart. J. Roy. Meteor. Soc.*, **111**, 709–738.
- Krueger, S. K., 1988: Numerical simulation of tropical cumulus clouds and their interaction with the subcloud layer. *J. Atmos. Sci.*, **45**, 2221–2250.
- Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc.*, **50**, 345–349.
- , 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513.
- Matsuno, T., 1966: Quasigeostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25–43.
- McKee, T. B., and S. K. Cox, 1974: Scattering of visible radiation by finite clouds. *J. Atmos. Sci.*, **31**, 1885–1892.
- Mellor, G. L., and T. Yamada, 1974: A hierarchy of turbulence closure models for the planetary boundary layer. *J. Atmos. Sci.*, **31**, 1791–1806.
- Moeng, C.-H., 1984: A large-eddy simulation model for the study of planetary boundary-layer turbulence. *J. Atmos. Sci.*, **41**, 2052–2062.
- , and J. C. Wyngaard, 1986: An analysis of closures for pressure-scalar covariances in the convective boundary layer. *J. Atmos. Sci.*, **43**, 2499–2513.
- , and ———, 1989: Evaluation of turbulent transport and dissipation closures in second-order modeling. *J. Atmos. Sci.*, **46**, 2311–2330.
- , D. H. Lenschow, and D. A. Randall, 1995: Numerical investigations of the roles of radiative and evaporative feedbacks in stratocumulus entrainment and breakup. *J. Atmos. Sci.*, **52**, 2869–2883.
- Monin, A. S., and A. M. Yaglom, 1971: *Statistical Fluid Mechanics*. Vol. 1, *Mechanics of Turbulence*, MIT Press, 769 pp.
- North, G. R., 1975: Theory of energy balance climate models. *J. Atmos. Sci.*, **32**, 2033–2043.
- Oreskes, N., K. Shrader-Frechette, and K. Belitz, 1994: Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263**, 641–646.
- Popper, K. R., 1959: *The Logic of Scientific Discovery*. Hutchinson Education, 479 pp.
- Ramanathan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D. Hartmann, 1989: Cloud-radiative forcing and climate: Results from the Earth Radiation Budget Experiment. *Science*, **243**, 57–63.
- Randall, D. A., 1980: Conditional instability of the first kind upside-down. *J. Atmos. Sci.*, **37**, 125–130.
- , K.-M. Xu, R. J. C. Somerville, and S. Iacobellis, 1996: Single-column models and cloud ensemble models as links between observations and climate models. *J. Climate*, **9**, 1683–1697.
- Schiffer, R. A., and W. B. Rossow, 1983: The International Satellite Cloud Climatology Project (ISCCP): The first project of the World Climate Research Programme. *Bull. Amer. Meteor. Soc.*, **64**, 779–784.
- Siems, S. T., C. S. Bretherton, M. B. Baker, S. Shy, and R. T. Breidenthal, 1990: Buoyancy reversal and cloudtop entrainment instability. *Quart. J. Roy. Meteor. Soc.*, **116**, 705–739.
- Smagorinski, J., 1963: General circulation experiments with the primitive equations. *Mon. Wea. Rev.*, **91**, 99–164.
- Wallace, J. M., and V. E. Koussy, 1968: Observational evidence of Kelvin waves in the tropical stratosphere. *J. Atmos. Sci.*, **25**, 900–907.
- Yanai, M., and T. Maruyama, 1966: Stratospheric wave disturbances propagating over the equatorial Pacific. *J. Meteor. Soc. Japan*, **44**, 291–294.

