QUALITY ASSURANCE IN ATMOSPHERIC MODELING

BY THOMAS T. WARNER

The rapid growth in the number of atmospheric model users is motivation for reviewing best practices in atmospheric modeling and emphasizing the scientific and technical preparation that is necessary to use the modeling tools effectively.

A formal definition of quality assurance that is applicable to this discussion is as follows: the maintenance of a desired level of quality in a service or product, especially by means of attention to every stage of the process of delivery or production. A lack of such quality assurance in the atmospheric modeling process can result from many causes. One is that some model users are less well trained and less experienced than others and lack an appreciation of the sensitivity of model solutions to the numerous decisions that must be made when configuring a model for a particular application. Another is that demands for quick results can lead to a less-than-thorough model setup and verification. A related factor is

the availability of state-of-the-science community models; this represents a great potential benefit to the community, but there is the risk that the models will not be used wisely. This paper suggests ways in which the atmospheric modeling process and culture can be improved, and it is aimed especially at the many novice modelers who are using these tools. The recommendations apply to the use of models for operational forecasting of weather,[1] for climate prediction, for research-oriented case studies, and for the generation of reanalyses. Many of the suggestions are not new ones, having appeared decades ago in references such as Anthes (1983) and Keyser and Uccellini (1987). This paper merely collects the wisdom from these and other sources and includes some additional contemporary advice. Note that there is no attempt here to provide a complete list of references for the discussion topics; the reader should refer to a text on numerical weather prediction (NWP) for this information.

THE INCREASING USE OF ATMOSPHERIC MODELS.
Thirty years ago, atmospheric models were used primarily by research scientists at government and university laboratories and by national

AFFILIATIONS: WARNER*—Research Applications Laboratory, National Center for Atmospheric Research,+ and Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado
*Deceased
+The National Center for Atmospheric Research is sponsored by the National Science Foundation.
CORRESPONDING AUTHOR: Andrew Monaghan, NCAR/RAL, 3090 Center Green Drive, Boulder, CO 80301
E-mail: monaghan@ucar.edu

---

[1] In addition to the large operational forecasting centers, many universities, commercial organizations, and individual countries run models in real time for research, forecaster training, and operational prediction.

weather services for operational prediction. The small cadre of model users had degrees in atmospheric sciences and almost certainly had benefited from formal courses in NWP. Since that time, many factors have contributed to a rapid increase in the number of model users and in the diversity of their scientific and technical preparation. These factors include the following:

- easy access to turn-key community models;
- the ease with which models can be applied, as facilitated by the use of graphical interfaces, online documentation, and training courses;
- rapidly declining costs and increasing user friend-liness of high-performance computing and data-storage hardware;
- the increasing accuracy of models, when used properly;
- a greater awareness of the value of model-generated weather and climate information;
- the greater maturity of coupled secondary models that allow forecasts of atmospheric variables to be used for prediction of floods, infectious-disease outbreaks, electric-power consumption, air-quality-related health warnings, etc.;
- the realization by nearly every nation that it is being affected by climate change and the resulting desire to perform climate downscaling to answer practical questions about future water resources, agricultural productivity, etc.;
- the—perhaps unfortunate—growing expecta-tion by reviewers of grant proposals and journal submissions that models should be a part of most research studies;
- the use of atmospheric models by specialists from other scientific disciplines; and
- the maturation of science in some developing countries.

An example of the increase in the number of users of two particular community models is shown in Fig. 1. The fifth-generation Pennsylvania State University–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) was a community modeling system that was replaced by the Weather Research and Forecasting model (WRF). The ordinate in the figure is the number of subscrib-ers to an e-mail-based news system that supports model users. Because users typically unsubscribe when they are not using the model, the number of subscribers is arguably proportional to the number of active users at any time. Clearly there has been rapid growth, especially over the last half decade, in the
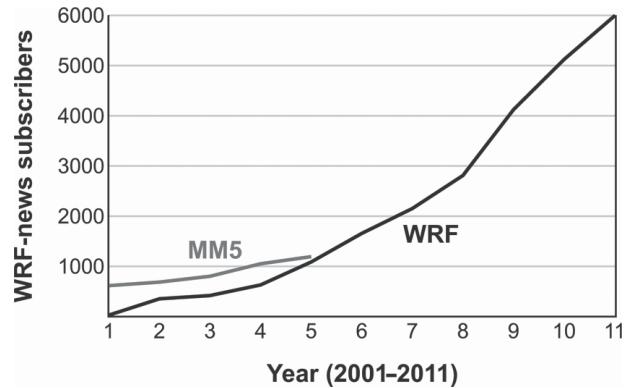

Fɪɢ. I. An example of the increase in the number of users of two particular community models as a func-tion of year. MM5 was a community modeling system that was replaced by the WRF. The ordinate in the figure is the number of subscribers to an e-mail-based news system that supports model users.

number of users of the WRF. This number is almost certainly an underestimate of the actual number of users (every user does not subscribe, but it is unlikely that nonusers subscribe), so the focus should be on the growth rate. The rapid increase in the number of users that is illustrated in this example and that is a result of the above-listed causes is motivation for reviewing best practices in the modeling process. Note that this increase in the number of WRF users may partly be a result of users switching from other models, but there is anecdotal evidence of growth in the number of users of other community systems, such as the Consortium for Small-Scale Modeling (COSMO) model.

REASONS WHY BEST PRACTICES ARE SOMETIMES NOT FOLLOWED IN THE MODELING PROCESS. An increase in the num-ber of model users cannot itself be responsible for the misapplication of models. However, there are many related factors that are causative. For example, new users may come from atmospheric science depart-ments that do not offer an NWP course. Without the benefit of this course, they have to resort to learning best practices through self study, through trial and error, or from more skilled colleagues. This is a sig-nificant problem because models are complex and imperfect tools and their shortcomings should be understood well by every model user. An even more unfortunate situation is one in which model users, in addition to having no NWP training, have no background in atmospheric sciences. This means that they are less likely to recognize when a model solution looks meteorologically unreasonable, because of

either basic model deficiencies or improper model settings. As evidence of this situation, about one-half of the applicants for a recent tutorial on the use of a mesoscale model had no formal training in NWP and about one-fourth had little or no atmospheric science background. Finally, in addition to this lack of training as a problem, time and financial pressures experienced by some model users sometimes prevent them from carefully applying the models, even if the expertise exists to do so.

STEPS FOR IMPROVING MODELING PRACTICES.

Most model users adhere to some of the following good practices, but often steps in the process are omitted in order to save time or because the user is unaware of the importance of the step. This is not a complete list of all the steps in a model application, but rather they are important ones that are often neglected or given insufficient attention.

1) Clearly define the scientific or practical objective. Too often, the model configuration is determined and experiments are performed without first writing down the specific questions to be answered, the expected results, and how the results will be used. As obvious as this step seems, project objectives are often not well defined and articulated from the start, and as a result model configurations must be modified and simulations rerun after goals becomes clearer with time. In fact, a careful definition of the objectives may lead to the conclusion that an analysis of the observations alone, without the use of a model, may be the best and most efficient approach.

2) Based on the above-defined objectives of the modeling project, identify and develop a physical understanding of the atmospheric processes that must be accurately simulated. This information about relevant physical processes will be essential for making decisions about model configuration, such as the necessary vertical and horizontal resolutions, and the most appropriate physical process parameterizations. It is also necessary in order to calculate the accuracy of the model relative to particular processes or variables of greatest importance. If the model is for general-purpose weather prediction, a variety of processes may be important. In contrast, some models have a narrowly defined goal: for example, for use in predicting electricity demand over a local area, in which case processes that control cloud cover and near-surface temperature would be especially critical. This step can involve a significant time

investment in observational-data analysis if the model is being applied in a region where the atmospheric processes have not been previously well studied. Otherwise, it will simply entail a careful analysis of the research literature. Without this step, many subsequent decisions (e.g., about model resolution) in the process will be made arbitrarily and quite possibly incorrectly.

3) Perform a thorough analysis of all available observations. Modelers should remember the often forgotten concept that there are three complementary approaches for studying the atmosphere, involving the use of observations, models, and theory (Hoskins 1983). The user should quality check and study all observations for the proposed simulation period of case studies used for research and for testing model improvements. Using the observations alone, one should perform the best possible overall analysis of the vertical and horizontal structures of the prevailing processes; this could require considerable time. Any conclusions from a study that involves the use of models will be much stronger if a thorough analysis of observations and the use of theoretical concepts are also part of the process. If the modeling is part of a doctorate dissertation, then in some cases it would not be unreasonable to spend a significant number of months just analyzing the observations, before touching the model. This is a justified investment in time, which will improve research efficiency and strengthen the outcomes.

4) Prepare an experimental design. This design should describe the model runs (not necessarily the configuration) that will be needed to accomplish the previously defined objectives. If case studies are being used, how many will there be and how will cases/dates be chosen? Will there be physical process sensitivity studies, and, if so, how will the solutions be compared (e.g., through simple subtraction of solutions or through the use of more complex methods)? Will the verification of the model solution be subjective or objective (statistical)? Specific aspects of the model configuration are defined in later steps, but preparation of this design ensures that the overall process has been thought through. Of course, with research there will inevitably be midcourse corrections in the design as the work progresses. The time required for preparation of the experimental design will depend on the experience of the model user.

In addition to the possibility mentioned in step 1, that the experimental design may include only an

analysis of the observations and no modeling, an additional design decision is whether to use real-data or idealized simulations. Even though real-data simulations are more common, for some project goals it may be more appropriate to use idealized (i.e., synthetic) experimental conditions. Here, the initial conditions and the forcing (e.g., land surface) are defined based on a simple conceptual model, and the model results are much easier to interpret because many processes are not interacting, as will be the case with real-data simulations. Some community models, such as the WRF, have options for using a variety of preconstructed, idealized, test cases, such as for flow over a mountain, a squall line, etc.

5) Define the required horizontal and vertical resolutions of the model, based on knowledge of the typical length scales of the (above established) specific processes that must be simulated well. If air quality in a coastal city must be simulated, then boundary layer processes associated with the land–sea breeze and urban-heat-island circulations will be important. This knowledge would guide the user to perhaps employ i) more model layers within the lowest 1–2 km above the surface in order to resolve the shallow, thermally driven boundary layer circulations and ii) high horizontal resolution to allow simulation of the mesoscale sea-breeze front. The estimate of the required vertical and horizontal grid increments should be based on knowledge of the "effective resolution" (Skamarock 2004) of the specific configuration of the model being used. That is, a number of aspects of a model configuration (e.g., the amount of explicit and implicit diffusion) control the filtering of the model solution. The resolution should be chosen such that all physical processes that are relevant to the study are adequately rendered by the model. Figure 2 shows the effective resolution for one configuration of the WRF, which has less smoothing than many models. Here, the effective resolution is ~7Δx in the context of the kinetic-energy spectrum. Other configurations (e.g., the specified magnitude of the coefficient for the explicit diffusion and the order of the diffusion) of the same model would result in a different effective resolution. Obviously, computational limitations exist for every project, so an outcome of this analysis may be that it is not feasible to accomplish the stated objective with the available time and computing hardware.

6) Avoid the tendency to prematurely run the model, before the above-listed steps have been completed.
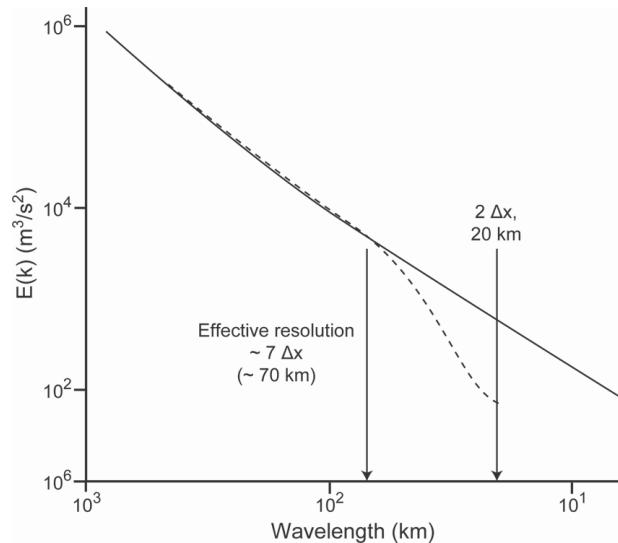

FIG. 2. The influence of explicit and implicit model diffusion on the kinetic-energy spectrum for a WRF forecast having a 10-km grid increment. The expected slope of k⁻⁵/³ is shown as a reference and is reproduced by the model for wavelengths greater than 7Δx. However, the energy between the 2Δx and 7Δx wavelengths has been damped by the model diffusion, resulting in an effective resolution of 70 km, not 20 km (2Δx) (adapted from Skamarock 2004).

Model users very often prematurely begin running the model, imagining that this will lead to an earlier completion of a project. In fact, this leads to inefficiency in the process, wasting time and computing resources, because model simulations will inevitably need to be rerun. The author's experience is that, the sooner that the model is run, the longer a study will take. Avoiding the temptation to run the model before thoroughly understanding the prevailing meteorology, before defining the experimental objectives and design, etc., is simply a matter of willpower.

7) Choose the model start time and the method of model initialization to allow for spinup of the physical processes of interest. Many model users do not realize that, depending on the initialization method and the prevailing physical processes, the model may need time to spin up clouds; precipitation; local ageostrophic circulations, which are associated with orographic or coastal forcing; and boundary layer structure. With some data assimilation methods, this might be enabled by running the model for several data assimilation cycles before using its output. Some methods, such as variational assimilation, may require considerable understanding and time in order to properly define information such as

background and observation error statistics, and they may thus be less attractive for inexperienced users. In contrast, many model users simply throw away the first few hours to one day of the simulation due to spinup issues, before using the model output, but this allows time for error to accumulate. The model start time relative to the study period is also a consideration. For example, when performing a modeling study of the daytime boundary layer, the initialization time of the model can greatly affect the results. Should one start the model the night before the study day, the morning of the study day, or the morning of the previous day to include the residual boundary layer and larger-scale effects from the previous day? Performing sensitivity experiments and reviewing the literature to assess what has worked in other studies is always a good idea.

8) If a limited-area model (LAM) is being used, then run test simulations to evaluate the sensitivity of the model solution to the computational domain size (i.e., lateral boundary location). The solutions from LAMs are notoriously sensitive to the locations of the lateral boundaries, and the use of excessively small domains often results in large errors. Tests should be conducted to define the optimal locations of the boundaries. The lateral boundary problem can take the form of a LAM, with two-way-interacting nested grids, that is parasitically (one way) nested within a coarser (e.g., global) model, or a single-grid LAM can be parasitically nested within the coarser grid. If a single case is being studied, perhaps five simulations will be sufficient to define the sensitivity of the model solution to lateral boundary location, as well as the configuration that produces the best

verifying solution. If the model is to be used for general-purpose operational NWP, each lateral boundary configuration should be tested for a range of weather regimes, flow patterns, and seasons, and a compromise solution should be found for the lateral boundary location. This process could possibly involve over 50 simulations, but this investment is arguably reasonable when establishing an operational modeling system.

Figure 3 shows a jet streak simulated by two versions of a LAM. One simulation (Fig. 3a) employed lateral boundaries that were located a long distance from the geographic area of interest (shown in the figure) and the other (Fig. 3b) had the lateral boundaries located at the edge of the area shown in the figure. The narrower jet streak in Fig. 3a is more realistic, as confirmed by radiosonde observations. In both cases, the lateral boundary conditions were provided by the same coarser-resolution global model. When there is flow over complex orography (exciting vertically propagating gravity waves), different options, if available, for upper-boundary conditions should also be tested.

9) Using sensitivity studies and reviewing the literature, define the most appropriate physical process parameterizations based on the geographic area, the available observations, the horizontal and vertical grid resolutions, and the processes being simulated. Community models often allow a user to select from a list of available parameterizations for convection, radiation, land surface processes, cloud microphysics, and boundary layer turbulence. The user should evaluate the sensitivity of the model solution to the choices of
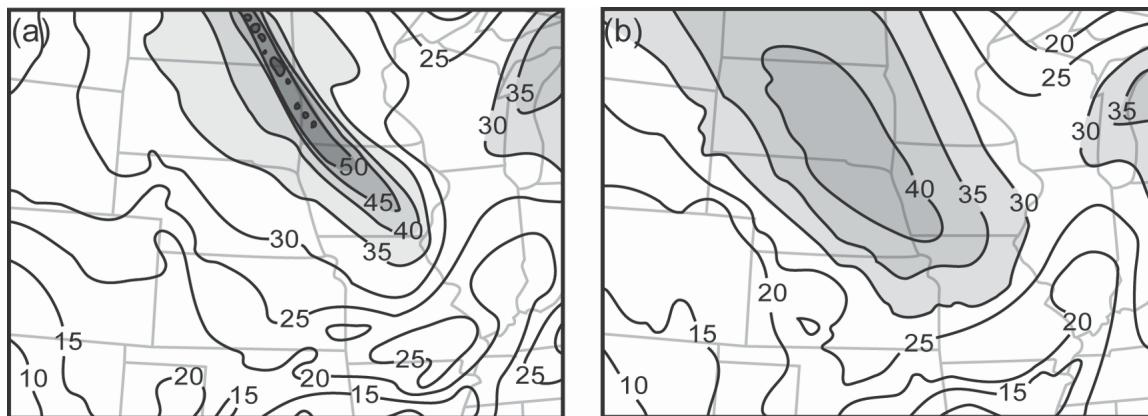

FIG. 3. Shown are 12-h simulations of 250-hPa winds (m s⁻¹) from the 40-km grid increment Eta model initialized at 1200 UTC 3 Aug 1992, based on experiments that used (a) a large and (b) a small computational domain. The isotach interval is 5 m s⁻¹ (from Treadon and Petersen 1993).

parameterizations. This is necessary because the performance of parameterizations can depend on season, the grid increment, the availability of observations, and the meteorological processes that prevail in specific geographic regions. This is why textbooks and model documentations do not attempt to suggest "the best option," something that users often desire. The "default" parameterizations suggested in the user documentation for a particular modeling system will not necessarily provide the best model simulation. Note that, when a model does not produce the desired solution, a common practice is to blindly try different combinations of parameterizations, without attempting to understand the reasons for poor model performance. Ideally, the user should investigate the assumptions inherent in the parameterizations and better understand the reasons for a particular pattern of errors (see Stensrud 2007). This will reduce the chances of getting a right answer for the wrong reason.

How the availability of observations can influence the choice of parameterizations can be easily understood in the context of land surface process parameterizations. In particular, using a complex land surface process model that requires the specification of many properties of the substrate and vegetation may result in a worse simulation than would be obtained from a simple model, if the land surface properties are only poorly observed for the geographic area being modeled.

The list of parameterizations to be tested can be shortened based on a literature review of what has worked best in other model applications for the region. However, if the model grid increment is markedly different than those used in other studies, the parameterizations might need to be retested. Also, newer available parameterizations whose performances have not been thoroughly described in the literature should also be evaluated.

The potential dependence of a model solution on the choice of a parameterization is shown in Fig. 4 in terms of the sensitivity of the accuracy of precipitation forecasts to the choice of the convective parameterization. The rainfall rate is plotted for a spring-season convective event (Fig. 4a), from observations, from four simulations that used different well-known parameterizations for convection, and from one simulation that used no parameterization for convection. At specific times in the simulations, the rainfall rate varied by as much as a factor of 4 among the different parameterizations. Also depicted is the bias score averaged for three warm-season convective events (Fig. 4b), again for each of the four parameterizations and for the use of no parameterization. Both the simulation-average scores on the right, as well as the time-dependent curves show a substantial dependence of the simulated precipitation amount on the parameterization that was employed. If one were to analyze the simulations of these cases in order to determine the best parameterization for convection, then it would be appropriate to go beyond the statistics shown and look at various object-based
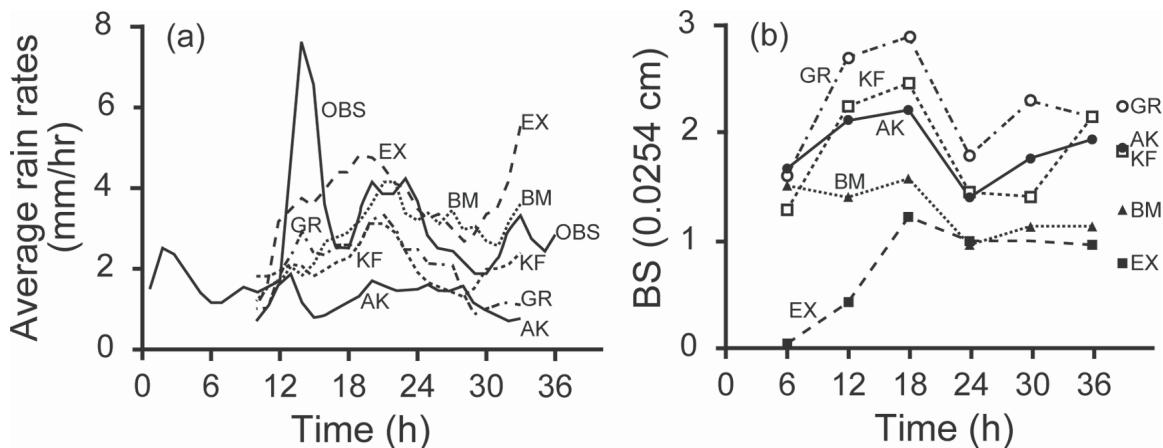

FIG. 4. (a) Average rainfall rate, for a spring-season convective event, based on observations (denoted as OBS) and for five simulations that used different treatments for the convection: four different parameterizations and no parameterization (denoted as EX). (b) Also depicted is the rainfall-rate bias score averaged for three warm-season convective events, again for each of the four parameterizations and for the use of no parameterization. The four convective parameterizations were the Grell (GR), Kain–Fritsch (KF), Betts–Miller (BM), and Anthes–Kuo (AK) schemes (adapted from Wang and Seaman 1997).

metrics such as the size and intensity distribution of the observed and simulated convective elements.

Finally, note that some combinations of parameterizations should be avoided, for example, because one parameterization can require the availability of specific variables from another parameterization. Good model documentation should alert the user to such situations, or the software might even prevent such choices from being made. However, when the documentation is insufficient, it is the user's responsibility to understand the parameterizations well enough to define any incompatibilities. Unfortunately, there is no good summary of these ill-advised combinations.

10) Understand the limitations to the predictability of the phenomena being modeled. Although estimates of inherent predictability limits for various phenomena are only guidelines, unrealistic expectations of the ability of any model to predict features of interest may result in i) erroneous conclusions about a model's predictive ability (or lack thereof) or ii) excessive time spent trying to tune a model to produce a particular solution that is exceptionally sensitive to any aspect of the model setup (e.g., initial conditions). The experimental design should recognize the limits of predictability; for example, should the model evaluation be purely statistical (e.g., does the model produce the right number of hurricanes per year) or deterministic (e.g., does the model produce a realistic evolution of a particular storm)? Of course, to a large degree our expectations about the predictability of various phenomena must be based on the configuration (e.g., resolution) of the model being used. Discussions of predictability limits for different atmospheric phenomena can be found in a number of NWP books, which should be consulted to ensure that there are realistic expectations about model capabilities. Finally, ensemble methods allow for a better assessment of the predictability of a process, compared to the use of a single deterministic run.

11) Establish a verification plan before the model is run and perform a thorough verification, using appropriate metrics, of the model solution using all available observations. The objective and subjective verification of model forecasts or simulations is essential for a variety of reasons. The following list is from Warner (2011):

- Most models are under continuous development, and the only way users and developers can know if routine system changes, upgrades, or bug fixes improve the forecast or simulation quality is to objectively and quantitatively calculate error statistics.

- For physical process case studies, where the model is used as a surrogate for the real atmosphere, the model solution must be objectively verified using observations, and, if the observations and model solution correspond well where the observations are available, there is some confidence that one can believe the model in the space and time gaps where there are no observations. This is a necessary step in most physical process studies.

- When a model is being set up for a research study or for operational forecasting, decisions must be made about choices for physical process parameterizations, vertical and horizontal resolutions, lateral boundary placement, etc. Objective verification statistics are employed for defining the best configuration.

- Only through verification does a forecaster have any chance of developing a sense of a model's systematic weaknesses and how to compensate for them. If the human wants to be in the loop and add value to a model forecast, then this is a requisite step.

- Objective decision-support systems that utilize atmospheric model forecasts as input can benefit from information about the expected accuracy of the meteorological input data.

A verification plan should be established before the model is first run, for a variety of reasons, including the fact that the selection of output frequency and output variables can depend on how the data will be used in the verification process. The verification plan should include the following information:

- Define the variables that will be the focus of the verification process. For general-purpose NWP, all variables may be equally important. When models are being used for a specific purpose, verification of particular variables will be emphasized. For example, for power companies, accurate forecasts of rapid ramp ups and ramp downs of wind speed at ~100 m above ground level are critical for efficient integration of wind power with other energy sources, so these events would be a focus of the verification.

- Describe the verification metrics to be used, whether they are feature (event) based or more traditional ones (e.g., bias, mean absolute error). The metrics should be specific and sensitive to

the weather features that are most important to the users of the model products. In the above mentioned example of ramp events, feature-based metrics could quantify the timing and amplitude errors of the forecast ramp events.

- Establish a process for defining the simulation period(s) to be used in the verification. This is an easy step if the model is being used for a research project to study a specific meteorological event. If an operational general-purpose NWP model is being prototyped or modified, the verification process can entail i) the emulation of the operational forecast cycling for a long multiseason period or ii) the simulation of a number of standard test cases/events/periods. One should avoid the natural temptation to choose only extreme weather events for evaluation of model performance; it is also important to evaluate it with more commonly occurring cases.

Most verification of high-horizontal-resolution models entails the use of traditional metrics, such as mean absolute error and root-mean-square error (RMSE), even though the results can be misleading and not necessarily demonstrate the real value of a forecast. See Davis et al. (2006a,b) for a discussion of the sometimes preferable event- or feature-based verification. Figure 5 illustrates a common problem when traditional metrics are used with high-horizontal-resolution models. In this example, model forecasts have both phase and amplitude errors. The solid line represents the observed wind speed in a jet. The dashed line shows a forecast from a model with high resolution, where the correct amplitude of the jet is retained but the maximum is displaced to one side. The dotted–dashed line and the dotted line show forecasts from models that produce a smoother solution. The RMSE of the forecast wind speed will be worst for the model solution that best renders the correct amplitude of the feature.

Performing a good objective verification does not mean that it is unnecessary for the modeler to visually review a wide range of model output variables and assess their general realism. Only through this process can the modeler develop a subjective sense of the model performance and identify error patterns that would not be apparent in the objective performance statistics. This obviously requires a good background in atmospheric science in general and specifically a knowledge of what the atmosphere looks like in terms of analyses of observations (gained from a weather-analysis course) and conceptual models. As a caution, however, using subjective verification in the process
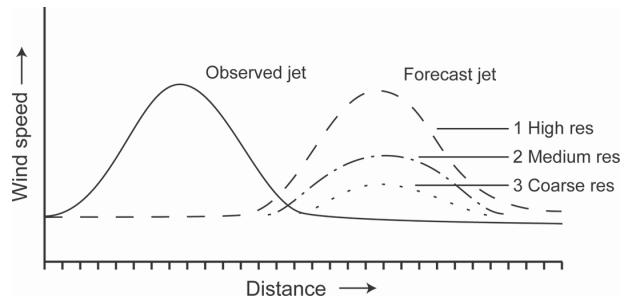

Fig. 5. Illustration of how a smooth forecast can lead to better verification statistics than a forecast with a more correct amplitude in structural features (see the text for discussion; from Warner 2011).

of defining the best configuration of a model (e.g., lateral boundary location and parameterizations) involves a risk that the model configuration will be inadvertently "tuned" to produce results that fit the modeler's incorrect understanding or conceptual model of a process.

If the verification process shows the existence of errors in the model solution that are larger than those reported in the literature in similar situations (phenomenon, geographic region, and season), adjust the model configuration accordingly (resolution and parameterizations) and rerun the simulations. Inadequate verification is a common and unfortunate compromise that is sometimes made by model users who are in a hurry to complete a project or who have unwarranted trust that the model will always perform well. It is arguably the most important step of all. Although costly, it is difficult to imagine a scenario where a thorough verification does not improve results.

12) Be well organized in maintaining a detailed experimental log and the files of model code and output. The natural tendency is to assume that one will remember the details of a model run or remember to write them down later, but such overconfidence often leads to situations where one is missing key information that would allow interpreting or reproducing the results at a later time. Thus, it is important to record the specifications of each model run immediately in an organized format. The same experiment log can also be used to refer to reference file numbers or names that will allow for retrieval of the specific model code used, and it can be used to record the interpretation of the results. This thorough recording of the conditions of an experiment, numerical or otherwise, is central to the scientific method. A related recommendation is that

a systematic approach be used for organizing digital files on a disk, of model code versions and model output files. Failing to have a well-organized overall process will result in lost time, wasted computer resources from rerunning experiments, and confusion.

13) Use good coding practices and well-documented and well-tested software. Many model users employ publically available models and other supporting software packages without modifying them. However, the availability of such software on the web obviously does not mean that it has been well tested, and this is problematic because every model user is ultimately responsible for the veracity of the software that is used in a research project.

Other model users, however, develop and test new model components as part of research projects, and any coding errors negatively affect the quality of the results, perhaps to an even greater extent than not following the above-listed practices. Developing quality code and adequately testing it is a matter of taking one's time and paying attention to detail. There are two ways of testing code, called black-box and white-box testing. In the former, the code is treated as a black box, and the code output is checked for realism. If a model has been modified, this could take the form of applying standard verification metrics to ensure that errors are within expected bounds, or there are ways in which models can be run for simple idealized situations and the model output can be compared with analytic solutions. Alternatively, white-box testing involves evaluating individual segments of new code in order to ensure that they are operating as expected. Opportunities for developing good coding, code-testing, and code-debugging practices can be provided to students in NWP courses by including a laboratory component that requires them to code and run simple models, perhaps based on the one- or two-dimensional shallow-fluid system of equations.

All publically available models undergo periodic major and minor upgrades, and new releases of the software are made available to the community. A historical example of a major upgrade was the release of the WRF, which eventually replaced the MM5. More often, major upgrades involve the implementation of large new segments of code within the same model, and the minor ones may be aimed primarily at fixing software bugs. This cycle of software releases can present the model user with the need for the following two types of decisions:

• For a new research project, is it better to use the newest major revision of the modeling software or the previous version? Even though the new version usually contains improvements of some sort, it also inevitably contains bugs that will be corrected as it is used in the future. It is recommended here that a new research project not employ a very recently released model or a model with a recent major upgrade, unless the upgrade has involved major bug corrections. Using the new model would involve the risk that a significant bug would be uncovered during a project and the work would need to be redone. Let others rush to use the new model and uncover the bugs.

• For an ongoing research project, should the work be periodically transitioned to new minor releases of the model or should it be completed with the original version? The answer here depends on a number of factors, including whether the release takes place near the beginning or the end of the project, the type of experiments that are being conducted, and whether the new release contains significant corrections to bugs. If the model user is in the middle of sensitivity studies, then of course there should not even be small changes in the software. If significant code bugs are corrected in a release, then it is arguable that experiments should be rerun with the new code or at least model runs should be used to determine whether the code change significantly affects the model solution.

14) Employ open-source software tools to improve the efficiency of the modeling process. There are a number of freely available software tools that can be used to make the modeling process more efficient and less prone to error. The following is a list of some of them:

• Employ version-control software to help keep track of the versions of the evolving model software that is used in a project.

• There are community software packages for model verification that have been developed and thoroughly tested. An example is the Model Evaluation Tool (MET) that was designed for the WRF but can be applied to output from any model. The MET will calculate standard or feature-based verification metrics.

• Graphical user interfaces are convenient ways of interfacing with modeling software, to configure and run the model. It simplifies the configuration of the domains, the running and monitoring of

the model, and the visualization of the output. An example is the WRF portal.

- Many open-source software packages are available for analysis and visualization of observations and model output, including the R statistics language and the NCAR command language.

The above mentioned model configuration and optimization process, even when based on compromises, can possibly involve running and carefully verifying more than 50 model runs. The model will sometimes need to be evaluated for a significant number of events or study periods. Ideally, each combination of parameterizations and lateral boundary configurations should be tested for each simulation period. To save time, it may be safe to first perform the tests of different parameterization combinations and then use the best one or two configurations in the lateral boundary sensitivity tests. Regardless of whether shortcuts can be justified, this overall process will require considerable computational and human resources, which should be budgeted, from the start, into the design of a modeling effort.

CONCLUSIONS. This paper provides a discussion of best practices associated with the use of atmospheric models. The suggestions emphasize fairly elementary applications of models and do not encompass the use of the relatively more complex ensemble modeling systems or more advanced data-assimilation approaches. The hope is that it will encourage model users to become more aware, either through self study or through enrollment in a formal course in NWP, of practices that will enable them to use these complex numerical tools more effectively. A corollary is that "on-the-job training" is generally not sufficient to teach undergraduate or graduate students how to properly use models.

An interesting point is that a scientist could, at an intellectual level, agree with the suggestions recommended here and be ready to adopt many of them. However, for this scientist to obtain funding for model-based research, funding agencies need to be prepared to support the sometimes time-consuming process described here. Proposal reviewers need to agree with the process as well. Thus, there perhaps may be institutionalized disincentives for adopting better modeling practices. Further, regarding the human and computational resources required to adopt the best practices described, it is arguable that the efficiencies gained by following some of the recommendations will at least partially offset the additional work involved with adopting others.

REFERENCES

Anthes, R. A., 1983: A review of regional models of the atmosphere in middle latitudes. *Mon. Wea. Rev.,* **111,** 1306–1335.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.,* **134,** 1772–1784.

——, ——, and ——, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.,* **134,** 1785–1795.

Hoskins, B. J., 1983: Dynamical processes in the atmosphere and the use of models. *Quart. J. Roy. Meteor. Soc.,* **109,** 1–21.

Keyser, D. A., and L. W. Uccellini, 1987: Regional models: Emerging research tools for synoptic meteorologists. *Bull. Amer. Meteor. Soc.,* **68,** 306–320.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.,* **132,** 3019–3032.

Stensrud, D. J., 2007: *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models.* Cambridge University Press, 459 pp.

Treadon, R. E., and R. A. Petersen, 1993: Domain size sensitivity experiments using the NMC Eta model. Preprints, *13th Conf. on Weather Analysis and Forecasting,* Vienna, VA, Amer. Meteor. Soc., 176–177.

Wang, W., and N. L. Seaman, 1997: A comparison study of convective parameterization schemes in a mesoscale model. *Mon. Wea. Rev.,* **125,** 252–278.

Warner, T. T., 2011: *Numerical Weather and Climate Prediction.* Cambridge University Press, 548 pp.