12 Operational numerical weather prediction

12.1 Background

The application of models for operational NWP has much in common with their use for answering physical-process questions, and for satisfying practical needs related to the assessment of air quality, evaluating the potential utility of new observing systems with OSSEs, and testing new numerical methods and physical-process parameterizations. Nevertheless, there are some issues that are unique to operational modeling. These will be addressed in this chapter.

It could be argued that the student of NWP should not need this kind of operations-oriented information because only large national modeling centers with experienced staff and large, fast computers are involved in operational prediction. However, there is a rapid growth in the use of operational regional models by consulting companies, universities, and regional governments to satisfy specialized needs. Thus, the student should become familiar with some of the concepts associated with the operational use of models.

Figure 12.1 illustrates the various components of a very simple operational modeling system. It should be kept in mind that the modeling systems that are operated by national weather services have very large infrastructures, and that the one summarized here is more consistent with the many modest-sized, specialized, operational-modeling systems that exist throughout the world. Some of these system components have been discussed before in earlier chapters, for example related to model initialization. To begin with, the system must have real-time connectivity to operational observational-data networks (top box in the figure), where this generally involves separate access to a number of different data providers. The input data types include current land-surface conditions, meteorological observations from *in-situ* and remote sensors, and gridded analyses and forcecasts from operational centers. After observations are received, they must undergo a quality-control process (Chapter 6). If the observations (e.g., satellite radiances) are not in the form of the standard model dependent variables, some data-assimilation processes may require that a retrieval algorithm be applied to obtain these variables. In the figure, the analysis and the data-assimilation processes are listed separately; however, as we have seen, they are often closely coupled (Chapter 6).

If a LAM is being used, LBCs must be provided to the forecast model, and possibly the data-assimilation system, from a global-model forecast (or a forecast from a larger-area LAM). The gridded forecast fields from the global model must be acquired

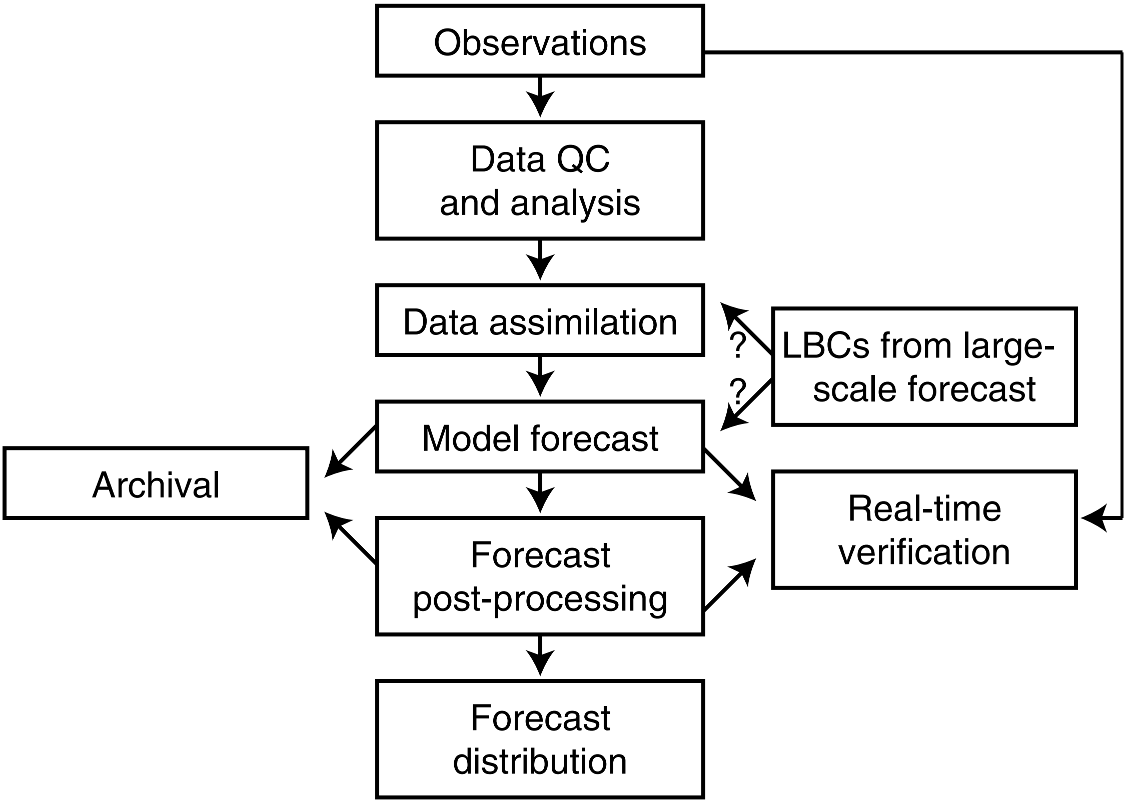

Fig. 12.1 Schematic showing the various components of a simple operational NWP system. See the text for discussion.

from a data service, unless the global model is run at the same facility as the LAM. Obviously, the global-model forecast must be completed prior to the integration of the LAM.

For each forecast of the model, observations that are assimilated, gridded initial conditions, and forecast products are typically written to a storage facility for archival. This archive allows retrospective reruns of the forecasts, to allow modelers to perform experiments to evaluate and correct the causes of especially poor model performance. Forecasts are also often verified in real time on a separate computing platform. The real-time verification statistics are sometimes made available to forecasters so that they can assess the objective accuracy of forecasts from recent cycles of the model.

The post-processing step will be discussed in Chapters 13 and 14, and can include the statistical correction of systematic errors. In addition, special post-processing codes (other models, in some cases) can be applied that derive unforecast variables – such as ocean wave height, river discharge, and air-pollution and dust concentration – from the forecast variables.

The forecasts are disseminated in graphical (analog) and digital form. Forecasters can use a web-based graphical interface to visualize the model output, or the gridded model output can be downloaded to a workstation on which special graphics software is installed.

When the atmospheric-model forecast products are to be used as input to the above-mentioned specialized post-processing models, users can access the digital, gridded forecast products from a data server. Lastly, sometimes specialized model products are produced for specific geographic locations such as major cities. A number of websites provide publicly available software for use in displaying forecast products. One of them is the Unidata site of the University Corporation for Atmospheric Research.

It is worth commenting on the evolving relationship between operational NWP and climate prediction. As will be seen in Chapter 16 on climate modeling, deterministic forecasts with coupled atmosphere–ocean (and other components) models are being produced on interseasonal and interannual time scales. These forecasts are now generated on a regular cycle like weather forecasts, and as the cycle frequency increases, the distinction will become even more blurred between them and operational NWP (see Toth *et al*. 2007). For example, the NOAA Climate Forecast System (CFS) is now running out to ~9 months, with a daily cycle and output every 12 h.

12.2 Model reliability

For research applications of models, the ability of the model to complete the integration with a very high reliability is not an especially great concern because the modeler has the opportunity to rerun the model with the problem corrected – perhaps eliminating a bad observation that made it through the QC process or reducing the time step to correct a violation of the CFL criterion. However, when fatal errors occur in an operational setting, the consequences are more severe, especially if the problem occurs in a model-based, sequential, data-assimilation system where each set of model initial conditions is based on the successful execution of a prior forecast. The resulting unavailability of a forecast in extreme-weather situations can result in the loss of lives. Interruptions to a forecast cycle can occur for the following reasons, among others.

- Violations of the CFL criterion can occur because of the existence of especially strong winds associated with the prevailing meteorological conditions. This could be prevented with an extremely small time step, but it is not especially sensible to use unnecessary computing resources during the 99.9% of the forecasts when the short time step is not needed in order to ensure trouble-free operations during the other 0.1% of the forecasts.
- Models rely on the timely availability of initial conditions. Problems with the data-assimilation system can cause cycles to be missed.
- Limited-area forecast models obtain their LBCs from previously run models, such as operational global models, that span larger areas. If this model does not complete its forecast cycle on time, or if there is a failure in the communication network over which data from the larger model are transmitted, the LAM forecast cannot run.
- The hardware on which the model computations are performed can have a catastrophic failure that will cause the entire system to become unavailable for executing the model, or a sufficient number of processors can become unavailable so that a forecast cannot be produced.

- When using LAMs operationally, for some meteorological situations and grid placements, the LBCs can generate sufficient noise to terminate a forecast. This is especially problematic when LAM grids are relocated operationally to focus the high resolution on specific prevailing meteorological features.
- There are many components of model code, especially related to physical-process parameterizations, that can fail when confronted with an unusual combination of meteorological inputs. Clearly one of the requirements for selecting a parameterization for operational use is its stability over a wide range of inputs.

12.3 Considerations for operational limited-area models

The above discussion highlights a couple of possible reliability problems that are unique to operational LAMs, related to generation of LBC noise and the availability of larger-scale data for defining the LBCs. There are additional issues that have to do with the efficiency of the operational LAM. One is the fact that the large-scale forecast that provides the LBCs must complete before the LAM forecast begins. Thus, there can be a substantial delay, of at least a few hours, in the start time of the LAM forecast, after the initial conditions have been defined.

Another point related to forecast-product timing applies when using a nested system of grids in an operational LAM – a common practice. If the grids are not two-way interacting, i.e. the information passage is only from the coarser grid to the finer grid, the coarse-grid forecast that is run first can be disseminated while the forecasts on the finer grids are still being generated. This sequential output of forecast products, as they are being produced, puts model guidance in the hands of the forecasters faster than if a two-way-interacting grid nest is used. Whether this benefit is worth sacrificing the possible advantages of two-way interacting grids is a situation-dependent decision.

12.4 Computational speed

For operational prediction, it is of obvious importance that model (simulated) time advance much more rapidly than actual (wall-clock) time. This is one of the motivations behind the adaptation of model codes to allow them to run on massively parallel computer architectures. And it is why such a great effort is invested in the development of efficient algorithms for solving the model equations, and allowing the use of longer time steps. The following factors influence the wall-clock time required by a model to produce a forecast of specific duration.

- Before an analysis or data-assimilation process can begin, the system must wait sufficiently long for most of the observations to be available from the network. Different operational systems have different cutoff times for observation availability, after which the data processing is begun and later-arriving observations are not used. This cutoff time is often 60–90 minutes.

- For LAMs, the forecast must wait for the availability of LBCs from a larger model that has to finish executing first.
- The time step defines the rate at which the integration moves forward, given specific computing resources. And, this time step depends on numerical-stability considerations as well as on the amount of temporal truncation error that is acceptable. If the CFL ratio is a constraint, obviously the grid increment is a strong controller of the time step, as is the speed of the fastest wave on the grid.
- The number of points in the computational mesh is a strong constraint. If the computing cost per grid point does not change with the number of grid points, this is a linear relationship. Combining this factor with a requirement for the CFL criterion to be satisfied leads to the often-stated rule that halving the horizontal grid increment increases the computational burden by a factor of eight, given the same area coverage for the grid.
- Operational models are run on computing platforms that range from desktops to massively parallel systems with thousands of processors. In addition to the effect on model-forecast speed of the number and speed of the processors, the efficiency depends on how the total-model speed scales with the number of processors.
- Some atmospheric models that are used for specialized applications do not include the full suite of physical processes. For example, for short-range predictions of convective outflow boundaries, many processes, such as long- and short-wave radiation, do not need to be included.
- Model output is sometimes made available to the forecaster as it is generated. For example, forecast products might be made available to the forecaster at 12-h intervals (forecast time), while the model is still running. Thus, the forecaster does not have to wait for the entire integration to finish before benefiting from information at the shorter lead times.
- There is sometimes a trade-off that must be made, when deciding on the numerical constructs to use in a model, between execution speed (and accuracy) and the "friendliness" of the code. That is, fast numerical methods (e.g., implicit differencing) will allow forecasts to finish more quickly, but the codes can sometimes be cumbersome to work with. This is an issue when an effort is made to unify the operational and university-research modeling communities by using a common model that must be accessible by both graduate students and experienced modelers. How this compromise is made will affect the speed of the forecast.
- Large data input–output loads can slow down the model-execution speed.

12.5 Post processing

There are a few types of post-processing algorithms that are commonly used.

- Correction of systematic error – When models are used in research for the study of physical processes, it is important that the gridded output be faithful to the governing equations. However, for operational prediction the imperative is to have the best guidance for the forecaster. Thus, it is perfectly appropriate to apply statistical-correction algorithms to the raw model output, for example to remove systematic error, even though this will upset the dynamic compatibility of the gridded output. See Chapter 13 for additional discussion.

- Calculation of additional variables with simple statistical or physically based algorithms – As noted above, the post processing also includes the use of the forecast dependent variables as input to statistical algorithms to calculate quantities that are not well forecast by the model, or not forecast at all. The latter have historically included quantities such as freezing rain, fog, turbulence intensity, and visibility.
- Use of secondary models that are coupled to the atmospheric model, for simulating complex processes – These models are discussed in Chapter 14, and include air-quality models, models that predict the elevation of dust from the surface and its transport in the atmosphere, models for the prediction of wildfire behavior, models for predicting the development of agricultural and human infectious diseases, etc.

12.6 Real-time verification

Chapter 9 describes the basic concepts of model verification. Real-time verification is distinct from what is done for research applications because it is performed immediately after observations become available for use in verification. The resulting statistics inform the forecaster about recent errors in model performance, in terms of the error dependence on time of day, lead time, location, etc. A challenge is summarizing and displaying the error statistics in ways that are intuitive, and that can be understood quickly by forecasters who are operating under severe time constraints.

12.7 Managing model upgrades and developments

Many organizations that employ operational modeling systems also perform research that is aimed at carefully verifying the forecasts, and improving the predictive skill of the model in the context of the specific mission of the organization. The computing needs associated with this objective are two-fold. First, proposed model improvements must be thoroughly tested in the setting of the operational system, and this can be most effectively done by having two independent modeling systems running in parallel. One is the operational system, and the other is identical except that the system improvements are included. In this controlled setting, model performance, with and without a system change, can be compared for a long series of forecasts. Such real-time system testing obviously requires a second computer platform that has at least the capability of the primary platform. In addition, while these two systems are running in parallel, allowing a comparison of the performance of the existing and prototype system, researchers need to be able to conduct case studies and other tests in the process of developing future upgrades to the system. This requires a more-modest sized third computing platform, or additional processors on the secondary platform. An important message is that it is not advisable to perform research during idle time on an operational system, because there is too much risk of accidentally corrupting it.

12.8 The relative role of models and forecasters in the forecasting process

Decades ago, when models were not nearly as skillful as they are now, the experience of forecasters was especially crucial to the generation and accurate interpretation of the operational products provided to the public. In the intervening years, models have contributed a growing amount to the "human–machine mix". In the extreme, model products are sometimes translated directly into images and computer-worded forecasts for the public, without expert interpretation. This has led to an ongoing discussion of how forecasters can best add value to the final products, now and in the future. As an illustration of the directions of such conversations, the following points have been made about the relative roles of the models and forecasters.

- Forecasts of "routine" weather should be automated to allow forecasters to focus their efforts on high-impact weather (Sills 2009).
- There should be a greater emphasis on the use of science in operational forecasting, which would be based on improved forecaster knowledge and the use of a more scientific approach to forecasting (Roebber *et al*. 2004).
- Forecasters should be at the "heart of weather prediction", playing a vital role in forecasting high-impact weather.
- Product generation should be automated, with forecasters focussing on analyzing the prevailing meteorology.
- Forecasters can make important contributions to forecast quality by manually modifying model gridded output, using a software system that maintains dynamic consistency among variables (Carroll and Hewson 2005).
- The use of higher-resolution models provides the forecaster with products having ever-increasing complexity. Tools thus need to be available to the forecaster to allow for the easy exploration and analysis of the model output (Roebber *et al*. 2004).
- As the role of forecasters evolves, benefit could be derived from entraining other disciplines that are involved in the cognitive psychology of decision making (Doswell 2004).

See the references in Sills (2009) for additional information.

SUGGESTED GENERAL REFERENCES FOR FURTHER READING

The reader should access the websites of organizations that run operational modeling systems and become familiar with the models, the organizational missions, and the weather products that are provided.

PROBLEMS AND EXERCISES

1. Compile a list of operational modeling systems that are employed by organizations other than national weather services.
2. Why have LAM systems been developed to predict weather for limited geographic areas? Do they compete in any way with national weather services? Are they complementary?
3. Speculate on the role of forecasters as models continue to improve in accuracy.