10 Experimental design in model-based research

The aim of this chapter is to provide a few examples of some common methods for using models in research studies. Other chapters also discuss experimental designs in the context of the specific subject being discussed. For example, there are many places in Chapter 16 describing experimental methods related to modeling studies of climate change. The summary here is far from complete because experimental methods are obviously closely tied to the objectives of a research project, which can vary widely. Nevertheless, the methods summarized are in wide use, and their strengths and limitations should be understood.

10.1 Case studies for physical-process analysis

Model simulations, generally for short time periods, are often used to study some aspect of a meteorological phenomenon. Sometimes the purpose is to better understand the predictability of a process, in terms of the necessary physical-process parameterizations or initial conditions. This is treated in Section 10.7 on predictive-skill studies. More often, the purpose is to use the model to help better understand the dynamics or kinematics of a physical process. The model is integrated from an initialization that is based on observations at the beginning of the study period. A next step in the process is to confirm that a good correspondence exists between the model simulation and the observations that are available during the simulation period. Good verification of the model skill at these observation locations is typically considered to be justification for believing the simulation in the space and time gaps between the observations. A benefit of using the model to fill the observation gaps is that the resulting fields are dynamically consistent, at least with respect to the dynamics embodied by the numerical approximation to the equations. Another benefit is that it is easier to analyze data on a quasi-regular matrix of grid points, compared to using the randomly spaced observations themselves. Additionally, the models respond to fine-scale local surface forcing that often adds information beyond what can be represented by the observations, and nonlinear wave interactions can add smaller scales than those represented in the observations.

The publicly available global or regional reanalyses described in Chapter 16 can be used for case studies. They are model-generated, and have all of the benefits mentioned above. And, these data sets are ready for analysis, requiring none of the investments associated with individual scientists running models. But, even though these data have been generated by trusted models from major forecasting centers, the resolutions are sometimes not sufficient to adequately represent some processes. Thus, it is often necessary to run a LAM to generate a fine-scale gridded data set for a case study. Generally, the LAM will use the large-scale reanalysis for LBCs. In predictive-skill studies, operational forecasts can be used instead. In addition to the benefits of added resolution, the use of a LAM for case studies allows the physical-process parameterizations to be chosen specifically for the phenomenon and geographic area being studied.

The following is a summary of the suggested components, and sequence of tasks, for a physical-process case study. It is especially important to understand the importance of first thoroughly analyzing the observations themselves, before running the model. Studying the observations will provide an understanding of the prevailing horizontal and vertical scales of motion, and the probable processes that are operating. This is essential information that is needed in order to properly set up the model.

- Clearly define the scientific objectives of the case study.
- Identify a candidate case to study, based on reviewing reanalyses or operational analyses, personal observation of a case, or the availability of special field-program observations.
- Obtain, quality check, and study all observations for the proposed study period. Analyze the vertical and horizontal structures in the atmosphere. Perform the best possible overall analysis of the process being studied – this could require months. Avoid the tendency to run the model before this phase is complete; running the model prematurely is a very common mistake (modelers like to model)!
- Determine how you would like the model to improve upon the above analysis of observations.
- Develop an experimental design for the modeling study. For example, will there be sensitivity studies? How will the model simulations be analyzed to satisfy the study objectives – cross sections, trajectory analyses, budget calculations?
- Based on the identified vertical and horizontal scales of motion associated with the processes being studied, choose appropriate horizontal and vertical grid increments. This should be based on the effective resolution of the model (Fig. 3.36), and not simply the grid increment. Evaluate the sensitivity of the model solution to the use of different horizontal and vertical resolutions.
- Based on a review of the literature, estimate the most appropriate physical-process parameterizations for the geographic area, the horizontal and vertical grid resolutions, and the process being simulated. Evaluate the sensitivity of the model solution to the use of alternative physical-process parameterizations.
- Define sources for the best possible model initial conditions, LBCs (if a LAM is being used), and land-surface conditions.
- If a LAM is being used, run test simulations to evaluate the sensitivity of the model solution to the domain size (LBC location).
- Perform a control model simulation, for use in verification.
- Compare the model solution with the observations available during the simulation period. If there are significant errors, adjust the model configuration accordingly (resolution, parameterizations).
- Based on the fact that the model compares favorably with available observations, use the gridded model output as a surrogate for the atmosphere in the physical-process

analysis. Chapter 11 reviews some methods that might be useful for analyzing model output.

The reader should have received the clear message above that there is much work that needs to be done in a physical-process case study before a researcher even thinks about using the model. In fact, the author's experience is that the sooner that the model is used in the process, the longer the study will take.

It is tempting to consider using continuous data assimilation, for example through Newtonian relaxation, to generate the gridded data sets for cases studies. After all, it can be argued that that process is better at integrating observations and model dynamics than one in which a model simulation only uses observations at the initial time. However, the relaxation, or nudging, terms are not physical, so the resulting model-generated data set does not exactly represent the thermodynamic or dynamical balance of the finite-difference equations. However, if this is less important than the correspondence of the model solution and the observations, the observations can be assimilated in this way. Nevertheless, it would still be advisable to first perform the simulation without assimilating observations throughout the study period, to allow those observations to be used as an independent check on the ability of the model equations to reproduce the processes.

When a single extreme weather event is to be analyzed using a case study, or if special field-program observations are available for a short period, it makes sense to use only one example of the process (i.e., one case) in the analysis. And, an in-depth analysis of even a single case can be very time consuming. However, it may be reasonable for some purposes to study a series of cases, to evaluate case-to-case variations in a process, or to make the conclusions more convincing. In such situations, it may be appropriate to focus on one or two aspects of each case, to make the analysis more tractable, rather than perform the analysis with the same level of detail that would be appropriate for a single case.

10.2 Observing-system simulation experiments

An Observing-System Simulation Experiment (OSSE) is a procedure for identifying the potential benefit to operational NWP of a yet-to-be-developed and -deployed observing system or observational strategy. Its use is motivated by the fact that observing systems are often extremely expensive to develop and deploy, so they must be first justified by a quantitative evaluation of the degree to which the possible new observations will improve the forecasts of operational models. This is accomplished by simulating the entire process, beginning with observing the atmosphere and ending with the verification of the forecast. Figure 10.1 shows the components of the OSSE process. The process begins in the upper left with the so-called nature run, where the best possible surrogate for the real atmosphere is generated by a model. Then, the measurement process is simulated by sampling the surrogate atmosphere in a way that is consistent with the existing and proposed

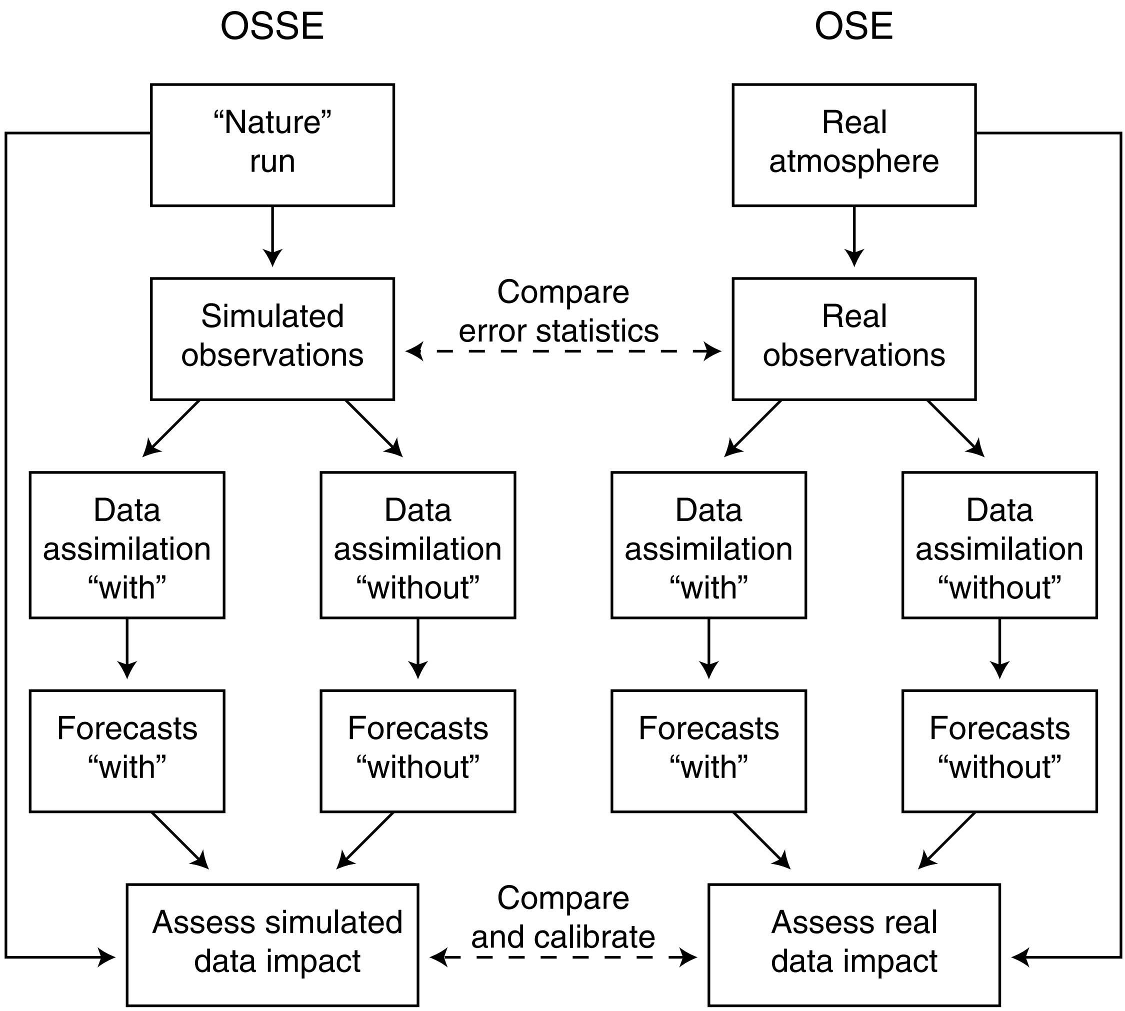

Fig. 10.1 Schematic of an OSSE (left), with the calibration and validation of the OSSE procedure shown on the right (the Observing System Experiment – OSE).

new observing systems. The simulated existing measurements are assimilated by the operational data-assimilation system, and this process is repeated by assimilating both the simulated existing and proposed new observations. Forecasts based on both sets of initial conditions are produced by an operational-class model (i.e., it must run in real time), and they are compared with the nature run in order to assess the impact of the new observations. The process is the same whether the impact is being evaluated of proposed new observing systems, or of new configurations (e.g., locations, numbers) of existing instruments. The OSSEs can be performed with global models alone, or with LAMs nested within global models. More discussion about each of these steps is provided below.

As suggested above, in addition to assessing the impact of potential new observing systems, OSSEs can be used to evaluate new observing strategies. For example, the influence of adding additional observations of a type that is already used can be estimated, as can the effect of moving present observations from one location to another.

10.2.1 The nature run

This simulation of an historical study period employs the highest model resolution and the best representation of physical processes that can be afforded with the available time and computational resources. It should therefore be expected that the nature run will more faithfully represent the real atmosphere than will forecasts from operational models that must run faster, on operational time scales. This is because the higher resolution in the nature run will result in (1) smaller truncation-related errors in approximating derivatives, (2) an ability to explicitly represent some processes rather than parameterize them, and (3) a better rendering of landscape forcing. The results of the nature run are archived at the native grid-point resolution and with high temporal frequency. This nature-run output has also been referred to as the “truth”, “history”, or “reference” atmosphere.

A goal is for the nature-run model to be as different from the operational model used later in the process as the real atmosphere is from the solution from the operational model. If the same model is used for the nature run and for the forecast, the surrogate atmosphere will have the same biases as does the forecast model. Thus, even if the resolution of the nature run is much greater than used in the forecast run, and the representations of some of the physical processes are much better in the nature run, the use of the same dynamical core will lead to common biases. The use of the exact same model configuration for both purposes is referred to as an *identical-twin experiment*. When the exact same model is not used for both purposes, but the models are not as different from each other as the forecast-model solution is from the real atmosphere, it is called a *fraternal-twin experiment*. Given that the nature-run simulation will be used to verify the forecasts, any biases that are common between the nature and forecast runs will make the forecasts appear better than they really are. An alternative is to use completely different models, and therefore different dynamical cores, for the nature run and forecasts, and this would normally qualify as a reasonable approach for an OSSE. But, even then it is well known that model solutions are sometimes more similar to each other than they are to the atmosphere.

10.2.2 Simulating the observations

There are two basic approaches for simulating an observation. The simplest is to interpolate from the nature run’s model grid to the observation location, and add an error that is consistent with the known systematic and random error of the measurement system. However, the most thorough approach is to use what is called an *instrument-forward model* that, as explicitly as possible, represents the interaction of the sensor with its environment, to produce a measurement that may, in fact, not be one of the model dependent variables. The atmosphere from the nature run is used as input to the instrument-forward model, and the output of the model is the simulation of the sensor output. For example, an emulator for a satellite-based sensor would use the nature-run variables as input to software (a model of the sensor) that would emulate the functioning of the sensor's optics and electronics. The output from the sensor, for example radiances, would then be assimilated by the data-assimilation software.

It is worth being reminded that the truth atmosphere was generated by Reynolds-averaged equations, and thus the model solution does not represent the turbulence that exists in the actual atmosphere. Thus, a current area of research is the development of methods for representing the effects of the turbulence on simulated observations in OSSEs.

10.2.3 Data assimilation

If an OSSE is being employed to assess the impact of a completely new sensor, on a satellite that has yet to be launched, it could be 5–10 years before the new data become available. The operational assimilation systems and models in use at that time in the future will inevitably be different than the present-day systems used in the OSSE. Thus, because the impact of an observation on forecast skill depends greatly on the assimilation system and the model, the impact of the new observations as assessed by the OSSE will not reflect the future impact. Even though there is no obvious way to address this problem, it should be recognized as a source of error in the process. A separate point is that, just as the error characteristics of real observations are used in operational data-assimilation systems, the error characteristics associated with the instrument-forward models mentioned above should be used in the data-assimilation software employed in the OSSE.

10.2.4 Forecasting

For the reasons mentioned above in the context of the data-assimilation system, the forecast model used in the OSSE should, as closely as possible, approximate the operational systems that will be in use at the time that the potential new observations will become available. This is because the characteristics of the forecast model will affect the impact of observations. To understand this, remember that the entire modeling process has many components that contribute to the final errors in the forecast, where these include the model initial conditions, the dynamical core, the physical-process parameterizations, and the quality of the lower-, upper-, and lateral-boundary conditions. Large errors in any one of these components can limit forecast skill regardless of the sophistication of the other components. For example, the benefit of accurate, high-resolution data from a new sensor is not going to be realized if the forecast model has coarse resolution or large errors in the representation of the physics. Thus, the value of a new observing system can be limited by the properties of the model used to produce the forecasts. The implication is that the use in the OSSE of a forecast model that is inferior to the one that will be in use when the future observing system is implemented, will probably underestimate the positive impact of the new observations.

10.2.5 Assessing the impact of observations

The forecast-skill measures used here can be conventional ones, or they can be specific to a particular application of the forecast. For example, if the objective of the OSSE is to evaluate the impact of a new type of global satellite-based observing system on 72-h synoptic-scale weather forecasts, measures such as anomaly correlations would be reasonable. However, consider an OSSE that is conducted to estimate the effectiveness of different scan strategies for Doppler radars that are used to initialize forecasts of moist convection. In this case, some of the measures described in Chapter 9 for verification of precipitation forecasts would be more appropriate.

10.2.6 Calibrating the OSSE

The right side of Fig. 10.1 illustrates a way of verifying the ability of the OSSE procedure to properly represent all components of the process, based on the use of current, real observing systems. The process is called an Observing System Experiment, or an Observation Sensitivity Experiment (OSE), where the purpose is to evaluate the impact on forecast skill of the use of an existing observation network. Here, a nature run isn't needed because real observations are used to initialize the forecast. And, the data-assimilation and forecast steps are run both with and without the observations from the measurement system that is being evaluated. The comparison of the forecast variables and real observations during the forecast period defines the contribution of the withheld observations to forecast skill. Then, the OSSE procedure is applied to the same case, where a nature run is generated, and the observations are simulated using the method of choice. Again, the data-assimilation and forecast systems are run with and without the simulated observations. A comparison of the two forecasts with data from the nature run defines the impact of the observation type. If the impact is similar from the OSSE and the OSE, it provides some confidence that the OSSE will reasonably estimate the impact of a hypothetical new measurement system. If there is a difference, OSSE results can be calibrated such that the estimate of the impact of the new measurements will be more realistic. Comparison of the documented, known error statistics from real observations and those from the simulated observations is another way of evaluating the OSSE process.

10.2.7 Examples of OSSEs

Table 10.1 lists examples of the use of OSSEs to assess the impact on model-forecast accuracy of future observing systems. Note that this list represents only a small subset of the hundreds of OSSEs that have been conducted. The observing system whose impact was evaluated, and the associated references, are provided.

Table 10.1 Example applications of OSSEs, arranged approximately chronologically

Purpose of experiment	Observing system, variable	References
Define the sensitivity of the accuracy of heat and moisture budget calculations, for convective situations, to the spatial and temporal density of soundings.	Radiosonde	Kuo and Anthes (1984)
Define the spatial and temporal frequency of soundings needed for calculation of accurate kinematic trajectories.	Radiosonde	Kuo *et al.* (1985)
Estimate the impact of a new surface-based observing system on mesoscale weather prediction.	Profiler network, winds and temperature	Kuo *et al.* (1987), Kuo and Guo (1989)
Assess the impact of potential new satellite observing systems on a global data-assimilation system.	Satellite Doppler lidar wind, microwave temperature and moisture	Hoffman *et al.* (1990), Zagar *et al.* (2008)
Assess the impact of potential new satellite observing systems on analyses and forecasts.	Microwave sensors; rainfall, water vapor, temperature	Nehrkorn *et al.* (1993)
Estimate the impact of a potential new satellite observing system on predictability.	GPS refractivity	Kuo *et al.* (1998), Ha *et al.* (2003)
Evaluate the impact of satellite winds on forecasts.	Satellite scatterometer winds	Atlas *et al.* (2001)
Evaluate the impact on forecast skill of a higher density observation network.	*In-situ* observations and satellite radiances	Liu and Rabier (2003)
Generate improved 13-month nature run with the ECMWF T511 General Circulation Model (GCM)	NA	Reale *et al.* (2007)
Assess the impact of potential new super-pressure balloon data on regional weather analyses and forecasts.	Balloon-borne pressure, temperature, humidity, and wind	Monobianco *et al.* (2008)

10.3 Observing-system experiments

The OSE procedure described above was used to verify the validity of OSSEs. However, OSEs may also be used to quantify the contribution to model forecast skill of existing observations. This may be motivated by the need to eliminate individual observations, or perhaps entire types of observations, as a budget-cutting measure. The process is described in the schematic on the right side of Fig. 10.1. A pair of data-assimilation and forecast cycles is performed, with and without the use of the observations being evaluated. As with OSSEs, the results should be based on an analysis of the sensitivity for each season of the year, for as long a period as possible. The availability of special observing systems during a field program also provides an opportunity for using OSEs to evaluate the impact on forecast skill of adding new types of observations.

10.4 Big-Brother–Little-Brother experiments

These Big-Brother – Little-Brother (BB-LB) experiments have traditionally been used to evaluate the impact of LBCs on the model solution in dynamic-downscaling experiments. The procedure is to first generate a high-resolution, large-grid-area reference simulation, called the BB simulation. This solution is then spatially filtered, so as to retain the scales typical of atmosphere–ocean general-circulation-model simulations. The identical model is then run for a smaller grid that is within the area of the larger grid, using the filtered large-grid simulation for ICs and LBCs. This is the LB simulation. The difference between the BB solution, and the LB solution after it spins up, is entirely attributable to the numerical impacts of the nesting procedure (e.g., the size of the smaller grid, the LBC update frequency, the blending strategy at the boundary) used in the downscaling process. Figure 10.2 shows a schematic of the procedure.

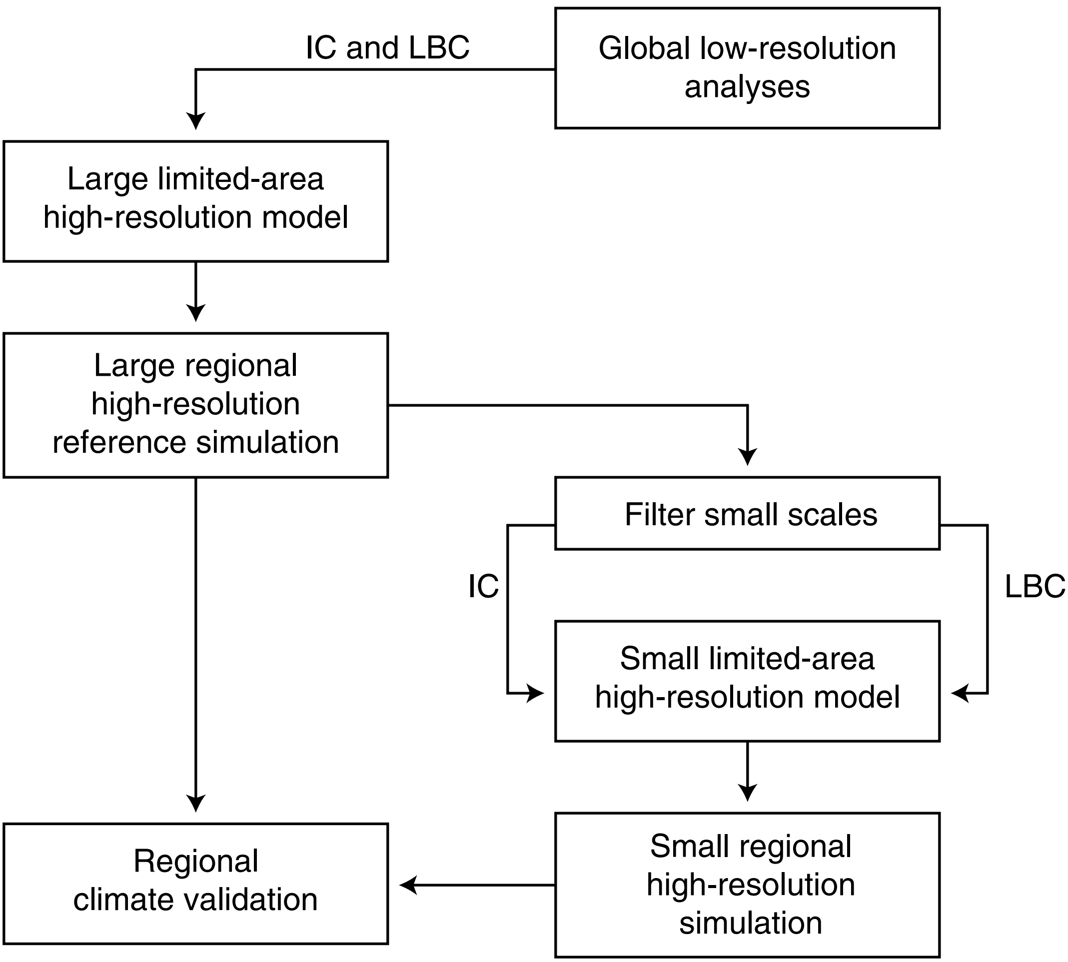

Fig. 10.2 Schematic of BB-LB experiments, as they can be used to test the LBC process for downscaling with regional climate models. Adapted from Denis *et al.* (2002).

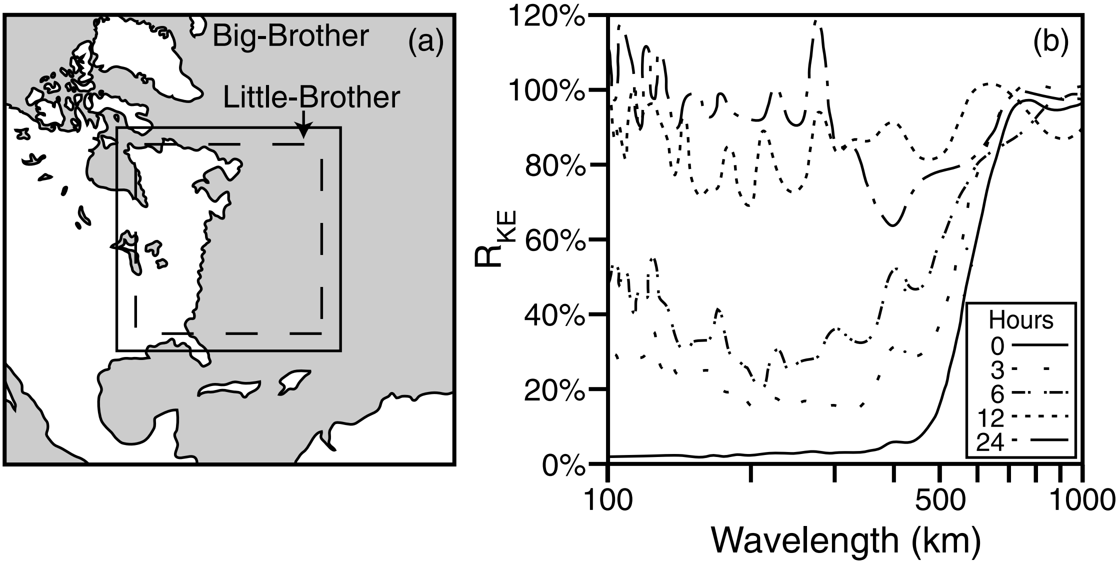

Fig. 10.3 Model grids used in the BB and LB experiments (a) and the spectrum of the ratio of the LB and BB low-level KE for different times during the LB simulation (b). Adapted from Denis *et al.* (2002).

Examples of the use of this type of experiment are in Denis *et al.* (2002), Castro *et al.* (2005), Dimitrijevic and Laprise (2005), Antic *et al.* (2006), Herceg *et al.* (2006), Diaconescu *et al.* (2007), and Køltzow *et al*. (2008). Even though these references focus on climate downscaling, this method can be used for other purposes. For example, Fig. 10.3 is based on the Denis *et al.* (2002) BB-LB experiment, and shows the time required for different spatial scales to spin up after initialization in the LB experiment. Panel (a) illustrates the area coverage of the large (BB) and small (LB) computational grids, and panel (b) shows the spectrum of the ratio of the LB and BB low-level Kinetic Energy (KE) for different times during the LB simulation. The ratio for the ICs (0 hour) of the LB simulation shows the result of the low-pass filter that was applied to the BB simulation. In fact, there was virtually no KE at this time in scales below 500 km on the LB grid. The ensuing KE growth in this part of the spectrum results at least partly from the model atmosphere's response to the Appalachian Mountains near the east coast of North America. The most rapid adjustment in the KE occurs between 6 and 12 hours. By 24 h into the simulation, the LB KE is very similar to that of the BB. This use of BB-LB experiments could define, for a new model application, the time required after initialization for model spin-up. The user of the simulations or forecasts would thus be aware that the model output should not be used during that period after initialization.

10.5 Reforecasts

Chapter 16 describes reanalyses, which are obtained by running the same data assimilation system for a long historical period. Reforecasts are similar to reanalyses, except that forecasts with the same numerical model are produced at regular intervals (e.g., daily) over an historical period using the reanalysis data set for initial conditions. The reforecasts can be deterministic forecasts or ensemble forecasts. These retrospective forecasts, which have been produced with a fixed version of the model, can be used for a variety of purposes. Biases can be calculated to help identify and correct weaknesses in the model. Or, they can be used for calculation of model-output statistics or, similarly, for training algorithms for statistically downscaling the forecasts. Lastly, predictability studies can be conducted. The obvious benefit of reforecasts over archived operational forecasts is the fact that operational models are routinely changing, with implementations of improved physics and numerics, and with code-bugs fixed.

Unfortunately, the computational requirements for generating decades of reforecasts are usually prohibitive for operational centers, given that operational-forecasting demands consume virtually all available resources. Alternatively, it has been possible for individual researchers to conduct a limited number of reforecasts to satisfy some of the above objectives. However, long periods of reforecasts are needed in order quantify model skill at forecasting infrequent, extreme events. References on this subject are Hamill *et al.* (2004, 2006), Hamill and Whitaker (2006), and Glahn (2008).

10.6 Sensitivity studies

A common motivation for performing a modeling study is to define the sensitivity of a model simulation to initial conditions, lateral or lower boundary conditions, or physical-process parameterizations. The following sections summarize some of the methods used for analyzing this sensitivity.

10.6.1 Simple sensitivity studies

A simple and historically common method of performing a sensitivity analysis is to produce a simulation with a control version of the model, and then change some aspect of the modeling process and perform a second simulation. By directly comparing the two simulations, or subtracting them to produce a difference field, an assessment can be made of the sensitivity to the modified process. For example, Fig. 10.4 shows the difference between two simulations, one with the Great Salt Lake and Utah Lake in western North America (the control experiment), and the other with the lakes replaced by the surrounding natural landscape. The purpose was to define the influence of the lakes on the regional wind field. A drawback to this approach to sensitivity analysis is that it only can answer simple questions. For example, in this case there are mountains near the lake shore, so the lake breeze at the time in the figure is influenced by the terrain, so what is seen is the interaction of the lake breeze with the orography, and not the result of the lake breeze alone. Nevertheless, if the objective of the analysis is to answer a practical question – in this case, how would the low-level wind field change if the lakes dried up – rather than to separate the different physical effects, this experimental design serves the purpose.

Table 10.2 lists a few of the many hundreds of model-based sensitivity studies that have been conducted. In addition to physical-process sensitivity studies, others isolate the impact of some aspect of the model configuration, such as resolution, physical-process parameterizations, LBCs, etc. Some of the listed studies, many of which had the aim of improving our understanding of the physical processes that prevailed in a particular meteorological situation, could have used the factor-separation method described in the next section. The last section of Chapter 16 discusses additional sensitivity studies that evaluated the impact of landscape changes on regional and global climate.

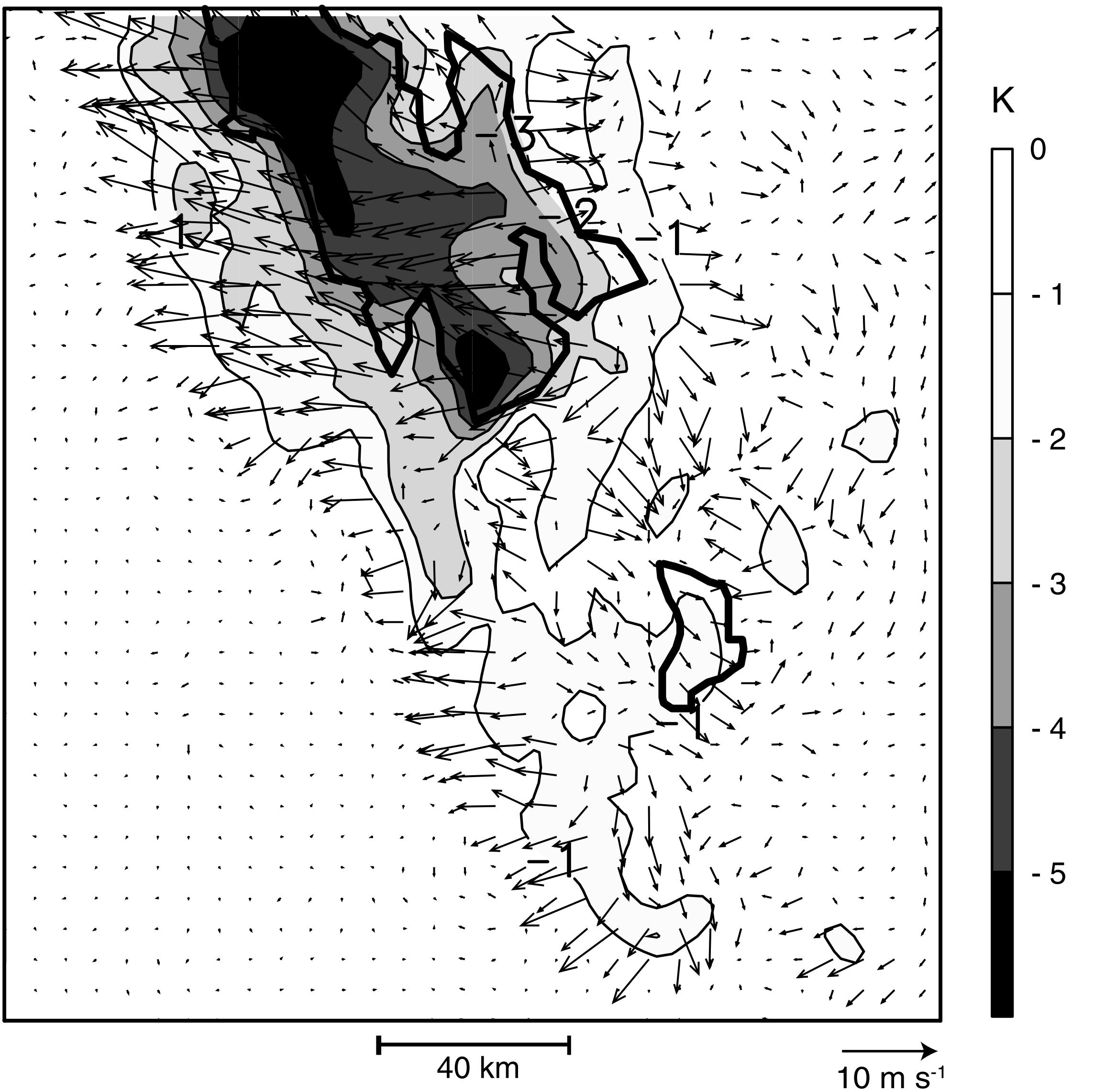

Fig. 10.4 Simulated 10-m-AGL wind (see vector scale) and 2-m potential temperature (shaded) difference fields (control minus no-lake experiment) for 1900 LT 14 July 1998. The potential temperature difference field is analyzed and shaded with a 1 degree interval, and the wind vector difference field is plotted at every second grid point. The heavy solid lines outline the Great Salt Lake and Utah Lake. Adapted from Rife *et al.* (2002).

Table 10.2 Examples of simple sensitivity studies. They include evaluations of the sensitivity of simulations of synoptic- and convective-scale processes to various factors (impact variables), including surface conditions, surface fluxes, latent heating, static stability, resolution, and LBCs.

Processes for which impacts are evaluated	Impact variables	References
Ice storms in the southeastern USA	Atlantic Ocean SSTs	Ramos da Silva *et al.* (2006)
Explosive maritime cyclogenesis in the Atlantic Ocean	Ocean sensible- and latent-heat fluxes, latent-heat release, initial conditions, horizontal resolution	Anthes *et al.* (1983)
Idealized maritime cyclogenesis	Ocean sensible- and latent-heat fluxes, latent-heat release, static stability, baroclinity	Nuss and Anthes (1987)
Convective initiation	Amount of assimilated data, lateral-boundary location, data-analysis procedure	Liu and Xue (2008)
Southern Hemisphere extratropical climate	Sea ice	Menendez *et al.* (1999)
Island convection	Wind speed and direction, surface fluxes, low-level moisture	Crook (2001)

10.6.2 Factor-separation method

This method is similar to the previous approach to performing sensitivity analyses, except that, here, more modeling runs span all the different combinations of the chosen factors, and mathematical manipulation of the model results allows isolation of the contribution of each factor to a specific output variable, such as precipitation. As an illustration, consider the experiment in Stein and Alpert (1993), where the purpose was to evaluate the relative contributions of the surface heat fluxes and irregular orography to modifying the large-scale dynamical production of precipitation in the region of the eastern Mediterranean Sea. During the day, the boundary layer is warmed as a result of the existence of surface heat fluxes, this establishes horizontal pressure gradients where there are variations in surface elevation, and the resulting thermally direct circulations have ascent over the higher elevations. This can generate precipitation. Here, it is not possible to isolate the effect of either the surface heat flux or the orography using just two simulations. For example, comparing the results of a control simulation having all the factors represented, with the results of a simulation that does not have variable orography, will eliminate the thermally forced precipitation over the higher elevations. But, the difference between the two precipitation simulations does not represent the contribution of only the orography because the heat fluxes were necessary as well. Similarly, precipitation from a no-heat-flux simulation subtracted from the control precipitation might produce a similar difference field, but the difference would be associated with the orographic variation as well as the fluxes. The factor-separation method addresses this problem by conducting four experiments: a control simulation with all the factors, a simulation with variable orography eliminated, a simulation with no surface heat fluxes, and an experiment with no variable orography and no surface heat fluxes. From these experiments, the effects on precipitation of the interaction of the heat fluxes and orography can be isolated.

Using notation similar to that in Stein and Alpert (1993), let f represent a variable field that results from a model simulation. For the above study, there are three factors that contribute to precipitation formation: large-scale dynamics (d), surface sensible-heat fluxes (f), and variable orography (o). Subscripts indicate the factors that are represented in a simulation; e.g., f_{dfo} is the precipitation field that results from the inclusion of all three factors in a control simulation. The quantities f_{df}, f_{do}, and f_d are similarly defined, and represent simulations where one or two of the local-forcing factors are not included. Now, let $\hat{f}$ be the variable field after factors have been separated from the solution, where subscripts refer to processes that have been isolated. For example, for the above situation, $\hat{f}_{df}$ is the precipitation that results from two factors, the large-scale dynamics and the heat fluxes; that is, the variable-orography factor has been eliminated. Based on Stein and Alpert (1993), the factor-separated fields can be calculated by the following equations:

$$\hat{f}_d = f_d, \tag{10.1}$$

$$\hat{f}_f = f_{df} - f_d, \tag{10.2}$$

$$\hat{f}_o = f_{do} - f_d, \text{ and} \tag{10.3}$$

$$\hat{f}_{fo} = f_{dfo} - (f_{do} + f_{df}) + f_d. \tag{10.4}$$

The last equation defines the precipitation field that results from the interaction of the surface heat flux and the irregular terrain elevation. Stein and Alpert (1993) describe the general form of the above equations, for an arbitrary number of factors. Note that multiple factors can be lumped together, if it is acceptable for their effects on the model solution to be aggregated in the sensitivity analysis. And, simulations do not necessarily need to be performed for every combination of factors. Even though there are clear benefits to the factor-separation method in terms of enabling the isolation of specific factors and combinations of factors, there are a few drawbacks.

- The method is time consuming. If n factors must be completely separated, 2^n model simulations are required.
- It is often not possible to identify a priori the most important physical factors that contribute to a particular aspect of a model solution. The unidentified factors, as important as they might be, have their effects collectively represented in the simulation that has all the other factors removed.
- Knowing the quantitative effect, on a simulated variable, of interactions among factors provides no insight about the physical processes represented in the interaction. This issue becomes more challenging for larger numbers of factors.

Table 10.3 lists some examples of applications of the factor-separation method used in sensitivity studies, with information about the variable in terms of which the sensitivity is tested, the geographic area, the types of factors whose impact on the model solution is assessed, the numbers of factors, and references. Figure 10.5 illustrates one way of

Table 10.3 Examples of applications of the factor-separation method

Process or variable on which impacts are evaluated	Geographic area	Factors	Number of factors	References
MCS-precipitation forecast skill	High Plains of North America	Physics parameterizations, initial conditions	8	Jankov *et al.* (2005, 2007)
Lee cyclone Sea-Level Pressure (SLP)	Western Mediterranean Sea	Orography, upper-level Potential Vorticity (PV) anomaly, surface sensible-heat flux	3	Horvath *et al.* (2006)
Snowfall	North America	Different Great Lakes	3	Mann *et al.* (2002)
Extreme convective precipitation	Spain	Orography and latent heating	2	Romero *et al.* (2000)
Mesocyclone vorticity	Eastern Mediterranean Sea	Orography, sea-surface fluxes (latent and sensible)	2	Alpert *et al.* (1999)
Cool-season heavy precipitation	Western Mediterranean Sea	Orography and surface latent-heat flux	2	Romero *et al.* (1998)
Quasi-tropical cyclone SLP and precipitation	Western Mediterranean Sea	Orography, surface sensible- and latent-heat fluxes, latent-heat release, PV anomaly	5 (not all are separated)	Homar *et al.* (2003)
Sea-breeze wind	Monterey Bay, California, USA	Coastline, coastal mountain, inland mountain	3	Darby *et al.* (2002)
Lee cyclone SLP	Alps Mountains	Lateral-boundary location, initial conditions, orography	3	Alpert *et al.* (1996)
Cyclone geopotential height, convective instability, wind	Coastal South Africa	Orography, surface sensible-heat fluxes	2	Singleton and Reason (2007)
Extratropical cyclone precipitation	Connecticut and Long Island, New York, USA	Orography, coastal differential friction	2	Colle and Yuter (2007)

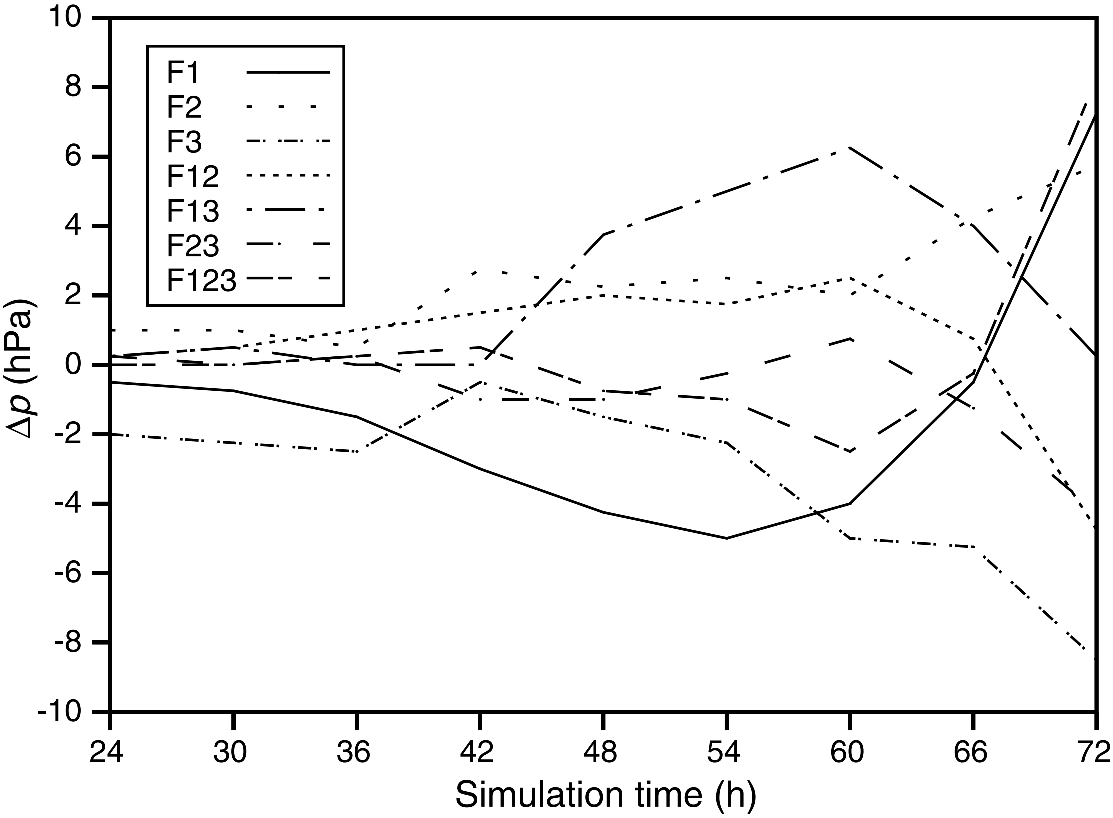

Fig. 10.5 The contributions of three factors to the evolution of the central pressure of a deep cyclone in the Mediterranean Sea. The factors evaluated are the existence of Atlas Mountain orography (factor 1), the surface sensible-heat flux (factor 2), and an upper-level PV anomaly (factor 3). The first 24-h period of the simulations is not shown on the time axis. Factors and factor interactions associated with each line are defined in the legend. Adapted from Horvath *et al.* (2006).

viewing results from a factor-separation experiment, in this case from the Horvath *et al.* (2006) study listed in the table. The factors evaluated here are the existence of Atlas Mountain orography (factor 1), the surface sensible-heat flux (factor 2), and an upper-level Potential Vorticity (PV) anomaly (factor 3). The influence of these factors on the central pressure of a deep cyclone in the Mediterranean Sea is assessed. In the figure, the F1, F2, and F3 curves are calculated using an equation analogous to Eq. 10.2. They thus isolate the individual influences of the three factors on the pressure depth of the storm. The next three curves identified in the legend show the contributions of the synergies between pairs of factors to the evolution of the central pressure (Eq. 10.4). The last curve shows the results of the triple interaction among the factors. Another way to interpret factor-separation results is to compare plan views of a simulated variable, for the same time in each experiment. Or, the factor-separated fields defined by Eqs. 10.1–10.4 can be plotted. For another synoptic-scale cyclogenesis case in the eastern Mediterranean, Fig. 10.6 shows the 36-h

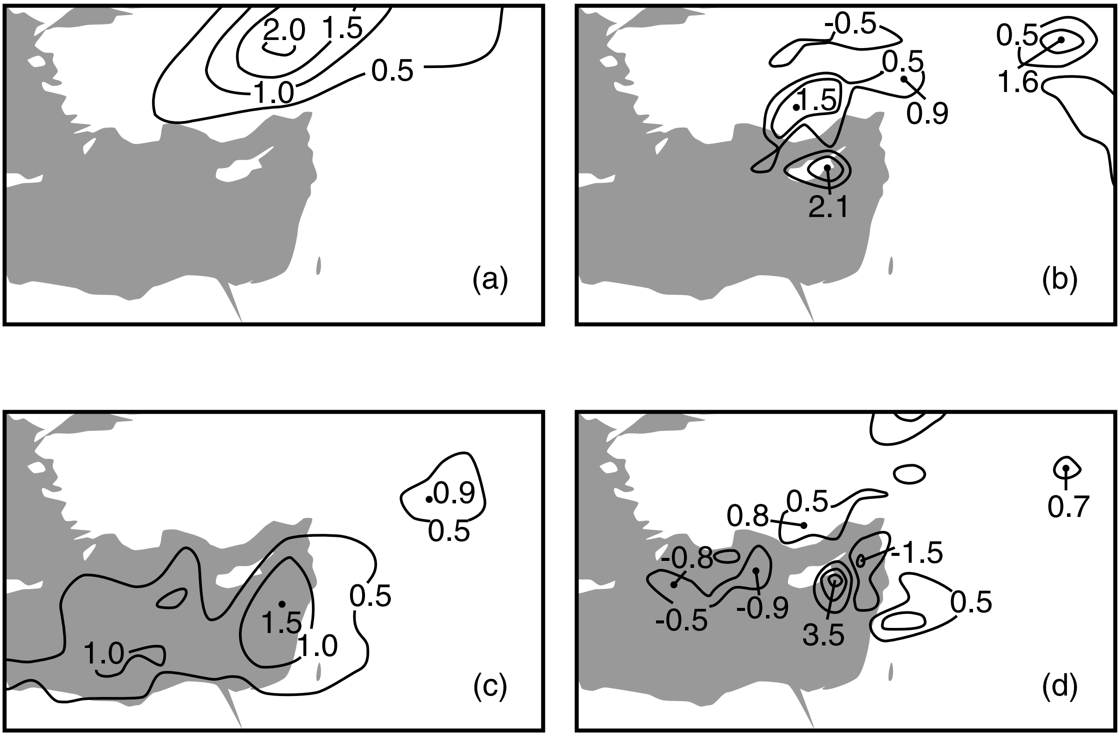

Fig. 10.6 For a synoptic-scale cyclogenesis case in the eastern Mediterranean, the 36-h simulated precipitation totals (cm) associated with large-scale dynamics alone (a), orography alone (b), surface fluxes alone (c), and the interaction between orography and surface fluxes (d). The zero isohyet is not shown. Adapted from Stein and Alpert (1993).

precipitation totals associated with large-scale dynamics alone (a), orography alone (b), surface fluxes alone (c), and the interaction between orography and surface fluxes (d), based on Stein and Alpert (1993). The results are intuitively reasonable. The large-scale dynamics produces a relatively large, smooth area of precipitation along the storm track, the orographic effects are smaller in scale and near mountains, the surface fluxes have their greatest impact over the waters of the eastern Mediterranean and immediately downwind, and the synergistic effects of fluxes and orography are near the orographic forcing in the eastern Mediterranean.

10.6.3 Adjoint methods

Adjoint methods are discussed in Chapter 6 in the context of their use in variational model-initialization procedures. And, Chapter 3 describes variational techniques, employing an adjoint model, that were used to investigate the sensitivity of LAM forecasts to initial conditions and boundary conditions. The adjoint operator produces fields that indicate the quantitative impact, on a particular aspect of the forecast, of any small, but arbitrary, perturbation in initial conditions, boundary conditions, or model parameters. In the types of sensitivity studies described in the previous two sections, modeling systems with different physics, data inputs, LBCs, etc., are run, and the resulting differences in the simulated variables are the measures of sensitivity. The difference fields represent the sensitivity of the variable to the perturbed factors. With the adjoint approach, direct metrics of the sensitivity are provided. For a more in-depth discussion of this technique, the reader should consult Hall and Cacuci (1983), Errico and Vukicevic (1992), Errico *et al.* (1993), and Errico (1997).

10.7 Predictive-skill studies

This common type of study assesses an operational model's skill at weather prediction, typically emulating the operational environment in a retrospective setting. This can be motivated by an interest in testing a forecast model for perhaps some new geographic area, or because a modification to the model has been made and the impact on forecast skill must be evaluated. Such studies are different in a couple of respects from case studies that use research models rather than operational systems. Specifically, because an operational system is being emulated, the resolutions and model physics are chosen such that the model will execute in a sufficiently short period of time for the output to serve as a forecast. Also, unlike physical-process-oriented modeling studies, the LBCs of LAMs must be specified using archived operational global-model forecasts rather than reanalyses of observations. That is, operationally realistic errors should exist in the LBCs. An individual case can be used, or more-meaningful results can be obtained using a long series of forecast cycles.

Modeling system components that are often evaluated in the context of operational prediction are the data-assimilation system, the dynamical core (numerics), the formulation of the LBCs for LAMs, the physical-process parameterizations, and the land-surface specification. Also, the impact of new data types, in OSE or OSSE frameworks, is evaluated in tests that emulate operational models.

Criteria for evaluating a forecast's success in an operational setting should be related to the variables of greatest importance for the ultimate users of the forecast information. For example, if warm-season convective rainfall is an important quantity because forecasts of it are used for agricultural applications, that variable should be included in the verification process.

Every operational forecast center conducts studies such as these using off-line (nonoperational) versions of the operational models. Thus, the fact that many hundreds of such studies have been conducted makes it impractical to summarize the literature. Let a couple of examples suffice. Powers *et al.* (2003) describe the initial testing of a new operational regional modeling system for use over Antarctica, to support aviation, maritime, and land-based activities. And, Liu *et al.* (2008b) summarize the performance of an operational regional modeling system used for five locations in North America.

10.8 Simulations with synthetic initial conditions

The use of real meteorological situations for modeling studies introduces inevitable complexity, sometimes making it difficult to interpret results when many processes are interacting. A solution to this problem can sometimes be achieved by the use of synthetic, or idealized, initial conditions. For example, much can be learned about coastal circulations, urban-heat-island circulations, and mountain-valley circulations through model simulations that use simple, idealized large-scale flow conditions. Initial large-scale winds may be defined as horizontally uniform or calm, and in geostrophic or gradient balance with the mass field. The resulting thermally forced circulations are more easy to interpret when they are superimposed upon the smooth large-scale flow, compared to the situation with real-data simulations where other features exist. Or, if the model's ability to properly simulate Rossby waves is a subject of study, an appropriate large-scale wave in the initial conditions can be prescribed analytically, superimposed on a zonal flow in a channel model. This general approach has been used for simulation of many processes, including tropical cyclones (Frank and Ritchie 1999, Riemer *et al.* 2008), boundary-layer flow over a forest canopy (Inclan *et al.* 1996), conditional symmetric instability (Persson and Warner 1995), and mesocyclones (Klein and Heinemann 2001).

10.9 The use of reduced-dimension and reduced-physics models

The use of these reduced-dimension and reduced-physics models is motivated by the same reasons as the use of synthetic initial conditions described above – simplifying the experimental situation to allow for a more-clear interpretation of results. In addition, the simplification of the modeling framework results in the use of less wall-clock time and computational expense to perform simulations.

10.9.1 Reduced-dimension models

These models include the single-layer, shallow-fluid models (x–y) described in Chapter 2, cross-section models (x–z or y–z), and column models (z). The shallow-fluid models are useful because the computer codes are simple, and there are generally no moist processes, radiation, or turbulence. They are typically used for evaluation of dynamical cores. Two-dimensional, vertical cross-section models often include a fairly complete representation of physical processes (to the extent possible with two dimensions), but the lack of the second horizontal dimension makes the models perhaps two-orders of magnitude less computationally intensive to use. Thus, higher vertical and horizontal resolutions can be used efficiently, and more computationally intensive numerical procedures and process parameterizations can be evaluated. Lastly, one-dimensional, column models are convenient for testing parameterizations of boundary-layer fluxes and growth, radiation, and moist convection – all being somewhat one-dimensional processes in terms of their representation in a model.

Some LAM systems allow a user to select an option that collapses the model to a cross-section configuration. If that option is not available, the user may be able to sufficiently reduce the grid dimension in one direction to achieve the same goal. For example, if only one row or column of grid points is needed to define the lateral boundary conditions at each edge of the grid, it may be possible to specify a grid dimension of three in the collapsed direction – one computational row or column, and boundary values replaced by interior values (zero-order extrapolation).

10.9.2 Reduced-physics models

Obviously these models that employ less than the full suite of physics need to be employed appropriately for situations where the lack of complete physics still allows the experimental objectives to be met. One type of reduced-physics model that we have already seen is the shallow-fluid model. The shallow-fluid system has no moist processes, no radiation, and no turbulence parameterization because it is used primarily for studies that focus on numerical solutions to the equations, and on simple dynamical processes. As noted in the previous section, this is also a reduced-dimension model, with typically no variability in the vertical.

The above-mentioned column models are, out of necessity, reduced-physics models. For example, if studies of the radiative impact of dust on the vertical temperature profile are conducted with a column model, all processes except radiation can be excluded. Similarly, in boundary-layer studies, all processes other than those associated with the land surface and the vertical fluxes of heat, moisture, and momentum can be ignored.

In the context of operational weather prediction and climate-system modeling, there are many examples of some physical processes not being included in the modeling system. For example, for weather prediction on time scales of weeks or less, coupled ocean processes are not represented. On climate time scales, a spectrum of models with different complexities is available for answering specific questions (Randall *et al.* 2007). In addition to quite simple climate models, there are Earth-system Models of Intermediate Complexity (EMICs) that have somewhat simplified physics representations compared with full physics Atmosphere–Ocean General Circulation Models (AOGCMs). Because of their greater simplicity and computational speed, the EMICs can address climate processes and interactions that evolve on time scales too long for AOGCMs. The use of simplified EMICs also allows large ensembles to be employed.

10.10 Sources of meteorological observational data

Unless model initial conditions are idealized, or the verification of the model uses analytic solutions, observations will be needed for initialization and verification in research or operational applications. Observational data are available at no cost from a number of sources. The US NCAR archives operational observations in their Mass-Storage System, but late observations are not added to the data set, nor are data-transmission-related gaps in the record filled. In contrast, complete archives are maintained by the US NOAA National Climatic Data Center (NCDC). In fact, NCDC maintains the world's largest archive of climate data. Satellite products that define atmospheric and land-surface properties are available from the ESA and NASA. Also, there are many regional mesonets that make, primarily, near-surface data available on servers in real time. However, it is always the user's responsibility to ensure that the data, regardless of the source, have been adequately QCed. Reanalyses, and archivals of operational forecasts, can also be obtained from many sources including NCAR, NASA, NOAA, and ECMWF. The best way to investigate how to obtain data from these organizations is to see their websites.

SUGGESTED GENERAL REFERENCES FOR FURTHER READING

Observing-system simulation experiments

Atlas, R. (1997). Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Japan*, **75**, 111–130.

Atlas, R., R. N. Hoffman, S. M. Leidner, *et al.* (2001). The effects of marine winds from scatterometer data on weather analysis and forecasting. *Bull. Amer. Meteor. Soc.*, **82**, 1965–1990.

Big-Brother–Little-Brother experiments

Denis, B., R. Laprise, D. Caya, and J. Côté (2002). Downscaling ability of one-way nested regional climate models: the Big-Brother experiment. *Climate Dyn.*, **18,** 627–646.

Reforecasts

Hamill, T. M., J. S. Whitaker, and S. L. Mullen (2006). Reforecasts: An important data set for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46.

Factor-separation methods

Stein, U., and P. Alpert (1993). Factor separation in numerical simulations. *J. Atmos. Sci.*, **50**, 2107–2115.

Adjoint methods

Errico, R. M., and T. Vukicevic (1992). Sensitivity analysis using an adjoint of the PSU–NCAR mesoscale model. *Mon. Wea. Rev.*, **120**, 1644–1660.

PROBLEMS AND EXERCISES

1. Explain why the impact on model-forecast skill of a new type of observation will depend on the characteristics of the model and data-assimilation system.
2. It is claimed that, if the instrument characteristics are properly specified and utilized in the assimilation process of an OSSE, then the instrument should have a positive or

neutral impact on the forecast quality. In contrast, a negative impact indicates a problem in the OSSE system. Explain why this should be true.

3. An example is provided in this chapter about how a BB-LB experiment can be used to estimate the time required during a simulation for the atmosphere to spin up in response to local forcing. Describe alternative approaches for this, with relative advantages and disadvantages.
4. Why are reduced-dimension models also commonly reduced-physics models?
5. The first section discusses the use of case studies for physical-process analysis. Traditionally, the cases analyzed were only a few days in duration. Speculate about how longer-period model simulations can be analyzed in a practical way, to provide more robust analyses of physical processes.