11 Techniques for analyzing model output

11.1 Background

This chapter describes methods for (1) the graphical display and interpretation of model output, and observations; (2) the calculation of derived variables from model output, which can help in the analysis of processes; and (3) the mathematical processing of model output, which can reveal properties and patterns that are not apparent from the dependent variables themselves. The comparison of the model output with observations is a type of analysis of course, but Chapter 9 on model verification is devoted to this subject. Also, the application of post-processing algorithms, for example to remove systematic error, is a special type of mathematical processing of the output, and this subject is treated in Chapter 13.

11.2 Graphical methods for displaying and interpreting model output and observations

Much of the material in this section is covered in courses on meteorological analysis; however, it is provided here because many students of NWP have not had such a course available to them. More in-depth material can be found in texts such as Saucier (1955) and Bluestein (1992a,b).

There have been so many creative ways of displaying model output, and comparing it with observations, that it is impossible to present a thorough treatment here. Nevertheless, some examples will be provided and the student is encouraged to review the literature and become familiar with typical techniques (see chapter Problems 1 and 3). This subject is important because successfully publishing research, whether it is model-based or not, depends on displaying the results in an easily and quickly understood format.

11.2.1 Eulerian analysis frameworks

In the *Eulerian framework*, the values of dependent variables are defined at grid points that are fixed in space. All standard software packages that are used for viewing model output include the option of plotting Eulerian plan views (quasi-horizontal), or maps, of the variables, where the analysis is performed on some reference surface such as pressure and is applicable at a particular time defined in terms of Greenwich time, local time, or forecast lead time. Maps of conditions near the ground may apply at a certain distance above ground level. For the atmosphere above the boundary layer, where diabatic heating associated with surface fluxes is negligible, performing the analysis of the meteorological conditions on isentropic surfaces can allow for a revealing interpretation of processes because, in the absence of diabatic effects, the flow remains on the analysis surface. Note that model output can be plotted and interpreted on isentropic surfaces, even if the model itself uses a different vertical coordinate.

An alternative Eulerian plotting option is the use of vertical cross sections. Here, a specific vertical plane is chosen on which model-output variables or analyses of observations are plotted. The orientation of the plane should ideally by chosen so that it best reveals particular processes or phenomena of interest. For example, Fig. 11.1 shows an east–west-oriented vertical

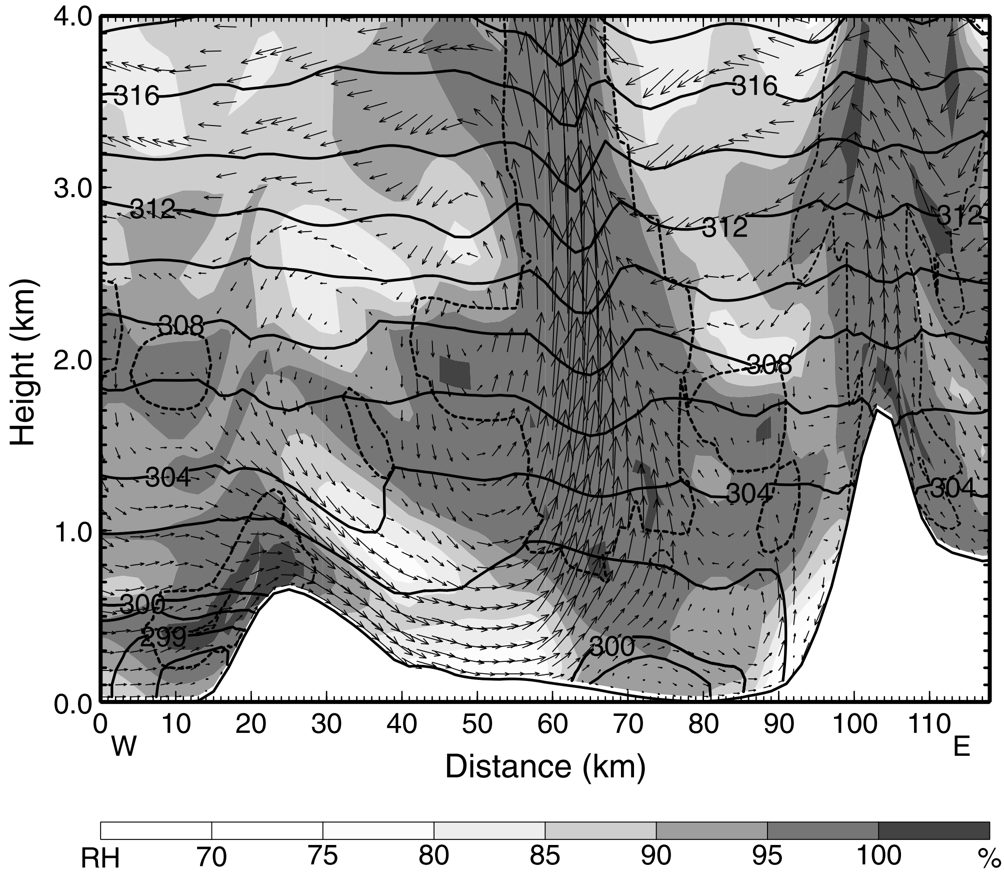

Fig. 11.1 East–west cross section of the lower troposphere for 1400 LT, based on a LAM simulation of processes along the west coast of Colombia. The Andes Mountains are seen at the lower boundary. The model has a 2-km grid increment. Relative humidity is shaded (see the legend at the bottom), with cloud boundaries indicated by dashed curves. Arrows indicate wind components in the plane of the section (zonal and vertical). Potential temperature is contoured at an irregular interval: 2° above 300 °C and 0.5° below 300 °C. The simulation contains what appears to be a hydraulic jump, as the leading edge of the cool maritime air of the sea breeze surmounts the low coastal mountain range and flows eastward into the Atrato Valley, to the foot of the Western Cordillera of the Andes. From Warner *et al.* (2003).

cross section of model-simulated conditions over the Andes Mountains of Colombia. For special applications such as providing weather guidance to aircraft pilots, displays of forecast weather can be produced on vertical surfaces that follow an irregular flight path.

There are numerous other types of Eulerian analyses that show the time evolution of forecast variables at a point, or along a line of points. For example, Fig. 11.2 illustrates a type of display in which the diurnal evolution and seasonal evolution of a meteorological

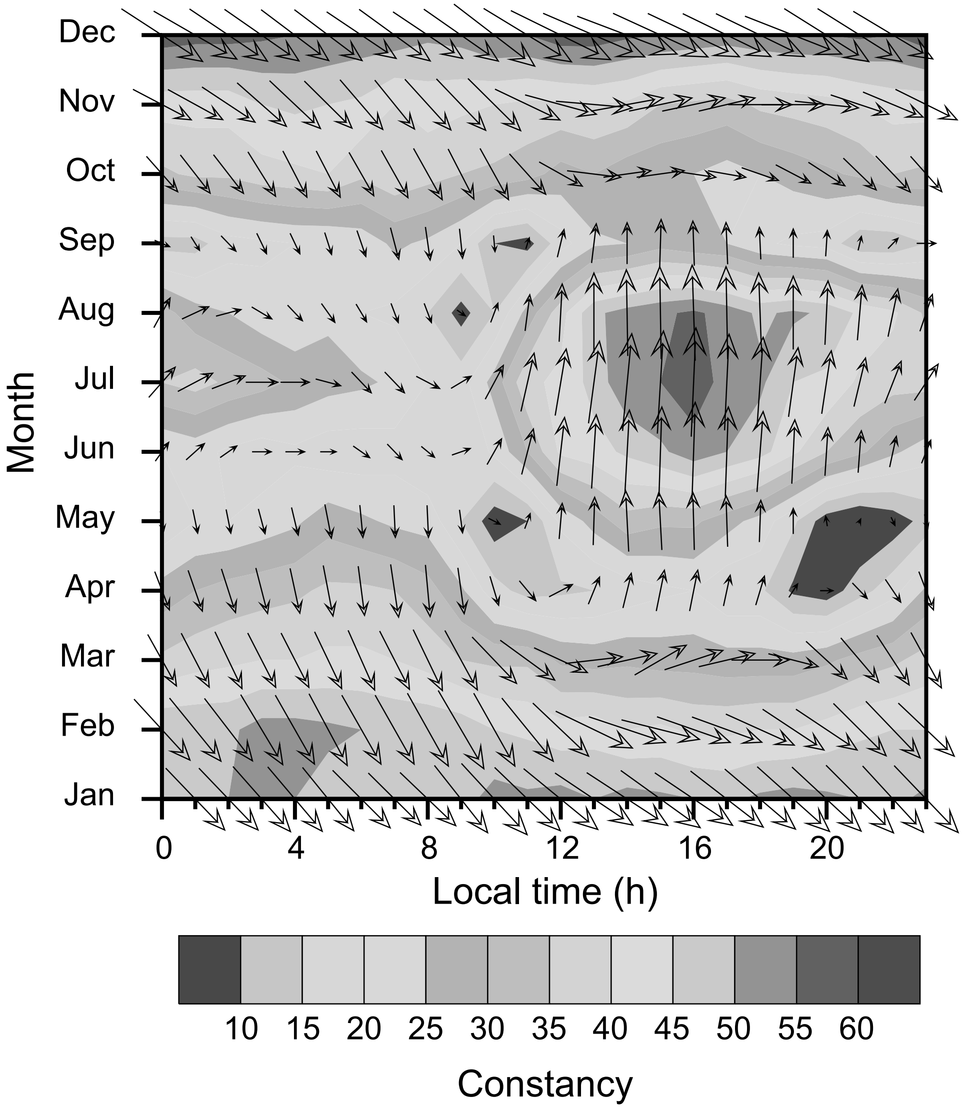

Fig. 11.2 Winds at John F. Kennedy Airport, New York, USA, as a function of time of day and time of year. The gray shades define the constancy of the wind, which is the resultant of vectors summed for a period of time divided by the average wind speed. For this near-coastal station, the onshore winds are strong and constant during summer afternoons. Provided by Ming Ge, NCAR.

variable can be revealed for a single location. In this case, the winds (vectors) are plotted for John F. Kennedy Airport, New York. The vectors illustrate the strong sea breeze during spring and summer afternoons. The gray shades define the constancy of the wind, which is the ratio of the magnitude of the resultant wind vector and the average wind speed.

Two other types of plots are known as Hovmöller diagrams and time–height sections. The *Hovmöller diagram* is a commonly used approach for plotting meteorological data, either generated by a model or based on observations, to highlight the motion of waves or features. The abscissa is latitude or longitude, and the ordinate is time/date. Colors or shading on the diagram indicate the values of some quantity that varies with the position of the wave. If longitude is defined on the abscissa, the value plotted at a longitude–time coordinate is an average value of the variable for a latitude band. An example of a Hovmöller diagram is shown in Fig. 11.3. Here, the GFS-model-simulated precipitation rate

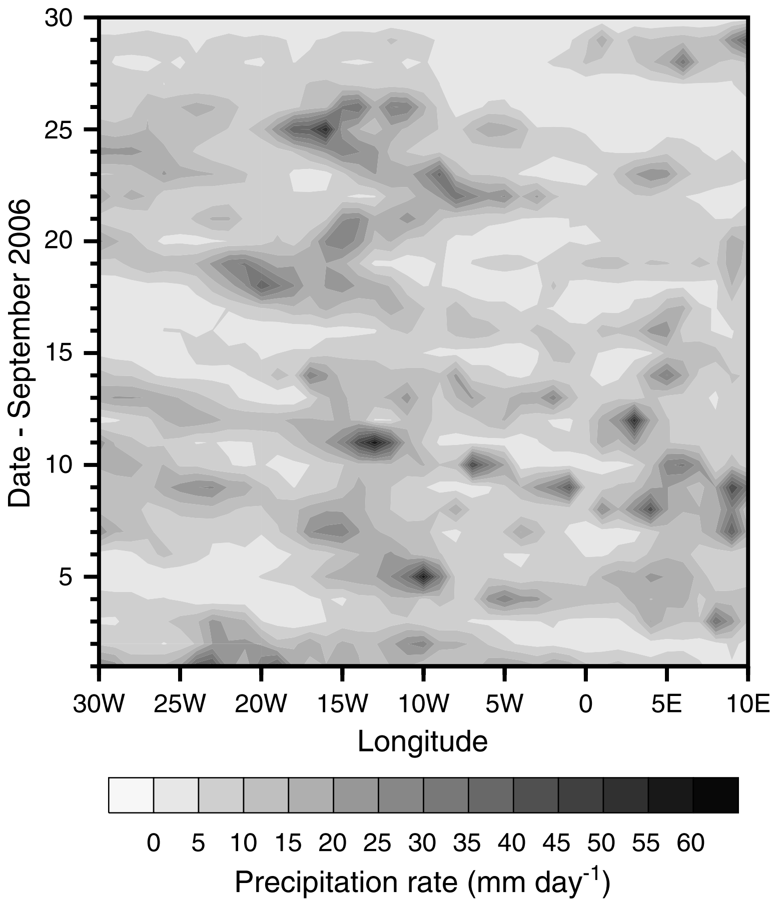

Fig. 11.3 A Hovmöller diagram showing the precipitation rate (mm day^{-1}) from a series of 24-h NCEP GFS-model forecasts. The abscissa spans West Africa, and the time interval plotted is one month. The values plotted apply to the latitude band 5–15° N. The slope of the pattern shows that the precipitation features were moving from east to west. Provided by Erik Noble, NASA GISS.

averaged for 5–15° N latitude in West Africa is shown as a function of longitude and time. The slope of the pattern shows that the precipitation features were moving from east to west, where the period is about 4–5 days between the occurrence of heavier precipitation at a particular longitude.

An example of one type of time–height section is shown in Fig. 4.12a, which depicts the vertical structure of the boundary layer as a function of time of day. A more typical version of this diagram would show isopleths, or colors, or gray shades that define the variation of the vertical profile of a variable at a point in the horizontal, as a function of time.

11.2.2 Tracking the movement of parcels of air, or physical features: the Lagrangian framework

Trajectory analysis

Trajectories are paths followed by parcels of air, and are thus often called parcel trajectories. A graphical display of trajectories applies to the period of time over which the parcel translation takes place. There are two common methods for calculating the parcel movement that defines the trajectory, where the difference is in terms of how the vertical velocity is computed. In what are called kinematic trajectories, the most-common type, the three velocity components are provided by a model, and these are used to define the parcel's three-dimensional motion. For $\boldsymbol{v}$ the velocity vector and $\boldsymbol{x}$ the position vector,

$$\boldsymbol{v} = \frac{d\boldsymbol{x}}{dt}$$

is integrated in time, where displacement in each coordinate direction can be calculated independently. Thus, in the east–west direction,

$$\int_{t_1}^{t_2} u\,dt = \int_{t_1}^{t_2} \frac{dx}{dt}\,dt\,.$$

Solving this numerically, using a time step Δt, we have

$$x_2 = x_1 + \left(\frac{u(x(t_1), t_1) + u(x(t_2), t_2)}{2}\right)\Delta t \approx x_1 + u(x(t_1), t_1)\Delta t$$

for small Δt. Thus, using high-frequency output of the three wind components from a model forecast or research simulation, the path of a parcel can be incrementally calculated. Typically a variety of initial points is chosen if the purpose is to reveal the pattern of fluid motion. For example, Fig. 11.4 shows a large number of trajectories calculated using model-simulated winds within the circulation of a hurricane. The trajectories were initiated in the low-level convergence zone, rise in the eye wall, diverge at upper levels, and provide a visual perspective on the circulation that would not have been possible using

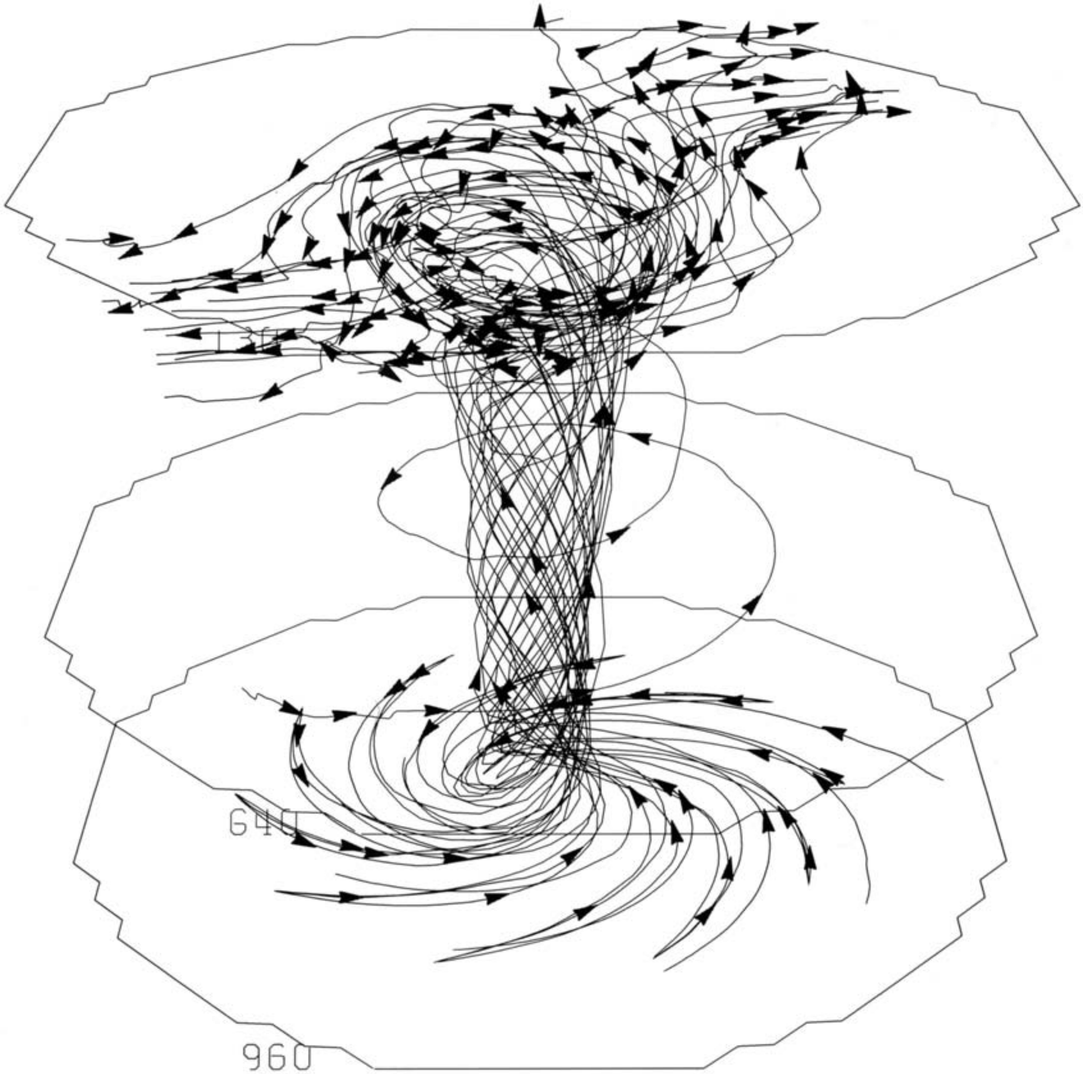

Fig. 11.4 Kinematic trajectories calculated using the model-simulated winds of a hurricane circulation. The lower circular plane is at 960 hPa, the middle one is at 640 hPa, and the upper one is at 130 hPa. Arrowheads are shown every 9 h along the path of each trajectory. From Anthes and Trout (1971).

vectors or other Eulerian displays. The other approach for calculation of trajectories is to assume that the parcels remain on surfaces of constant potential temperature, where these are called isentropic trajectories. The vertical velocity is thus implicitly defined by the horizontal component of the motion and the slope of the surface. As with kinematic trajectories, the model-defined winds are used to compute the horizontal displacement of parcels. Sometimes it is desirable to calculate the origin of a parcel of air that arrives at a particular location, in which case the mathematical process can be reversed and *back trajectories* calculated. Figure 11.5 shows another example of how trajectory analysis can be used to visualize complex spatial patterns in the motion of a model-simulated fluid. A grid was placed over a simple flow pattern in the troposphere, associated with an approximately symmetric trough in the heights. Trajectories were used to define the paths of different regions in the fluid, as defined by the grid, and the resulting distortions in the grid were mapped.

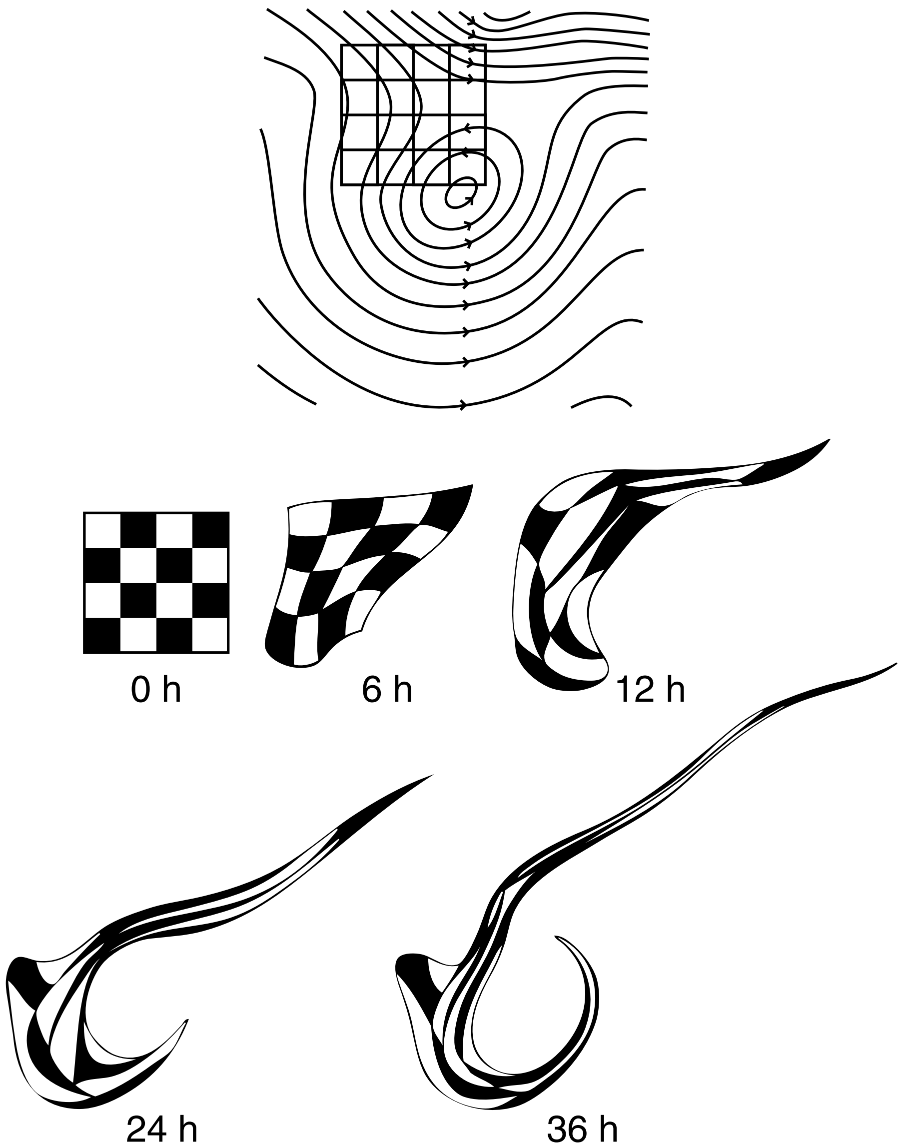

Fig. 11.5 The deformation of a layer of air at 500 mb in a barotropic model of the atmosphere, after the indicated simulation times. The initial streamline pattern is seen at the top, along with the overlaid pattern whose subsequent evolution is traced with trajectories. From Welander (1955).

Physical features can also be tracked, where an example is the image of the tracks of the hurricane centers shown in Fig. 7.1. All that is required to produce this type of analysis is the ability to define the location of a feature in an automated way. This is fairly straightforward for a hurricane, but can be more problematic for extratropical cyclones or convective systems.

Streamline analysis

Streamlines are lines that are drawn parallel to wind vectors at a specific time. They are thus different from trajectories because the lines do not follow the movement of parcels through both time and space. Rather, the lines simply make it easier to view the pattern in the wind direction, compared to the use of vectors or other symbols that only define wind direction at grid points. Note that streamlines are not the same as streamfunction lines, where the latter define the rotational part of the wind. The two may visually appear to be similar, but the spacing of the streamfunction lines is quantitatively related to the wind speed, whereas the spacing of streamlines is arbitrary and is determined by the analyst or analysis software so as to optimize the ease of visual interpretation. Figure 11.6 illustrates

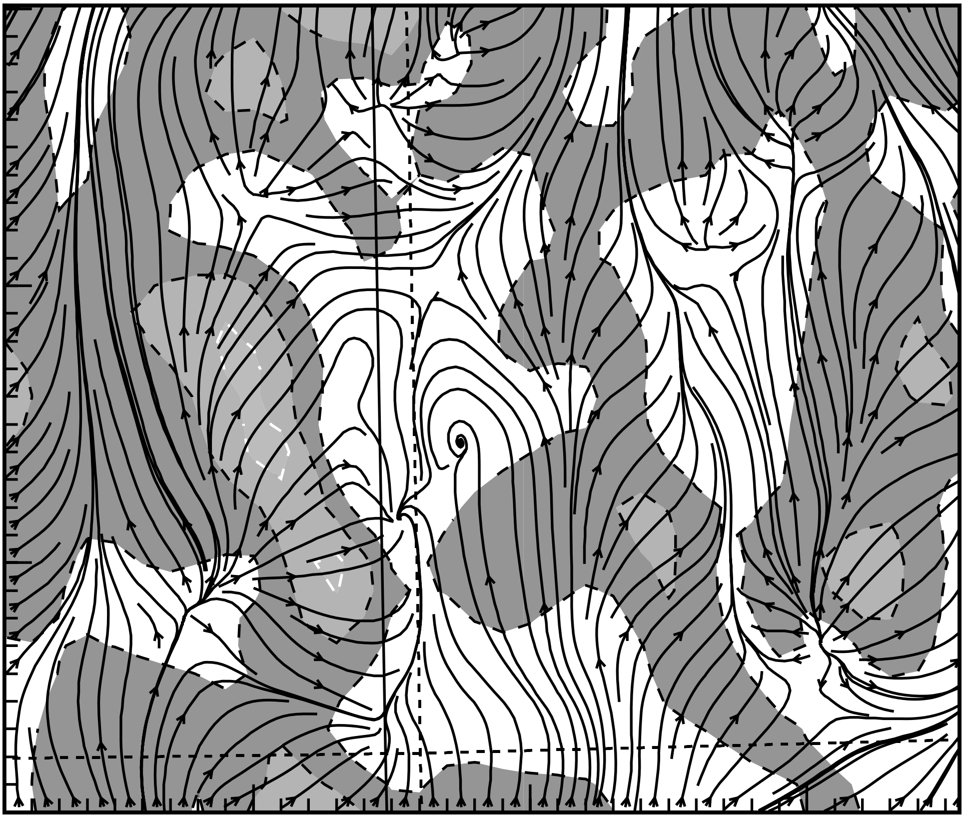

Fig. 11.6 Streamlines for ~15-m AGL based on a mesoscale-model simulation of a region in the western USA. There is much structure to the wind pattern because of the existence of complex orography. The shading shows wind speeds, where white is less than 5 m s^{-1}, and the gray shades have bandwidths of 5 m s^{-1}. Provided by Yubao Liu, NCAR.

~ 15-m AGL streamlines based on a mesoscale-model simulation for a region in the western USA. There is much structure to the pattern because the area is dominated by complex orography. In this display, the speed is also shown in the form of gray-shade bands.

Isochrone analysis

Isochrones are lines (literally, lines that apply at a particular time) that define the location of a geometrically simple feature in the atmosphere, based on model output or observations. An example with which we are nearly all familiar is the synoptic-scale midlatitude front. But, in an isochrone analysis the frontal location would be shown for multiple times in order to characterize changes in its shape and position. Other features that could be similarly analyzed include convection-related gust fronts or outflow boundaries, sea-breeze fronts, dry lines, the edge of elevated mixed layers, the leading edge of a precipitation shield, etc. A necessary characteristic of the feature is that its geographic position must be definable in a simple way, so that when it is drawn for multiple times the image does not become too complex to interpret. An example is shown in Fig. 11.7 of isochrones of frontal position.

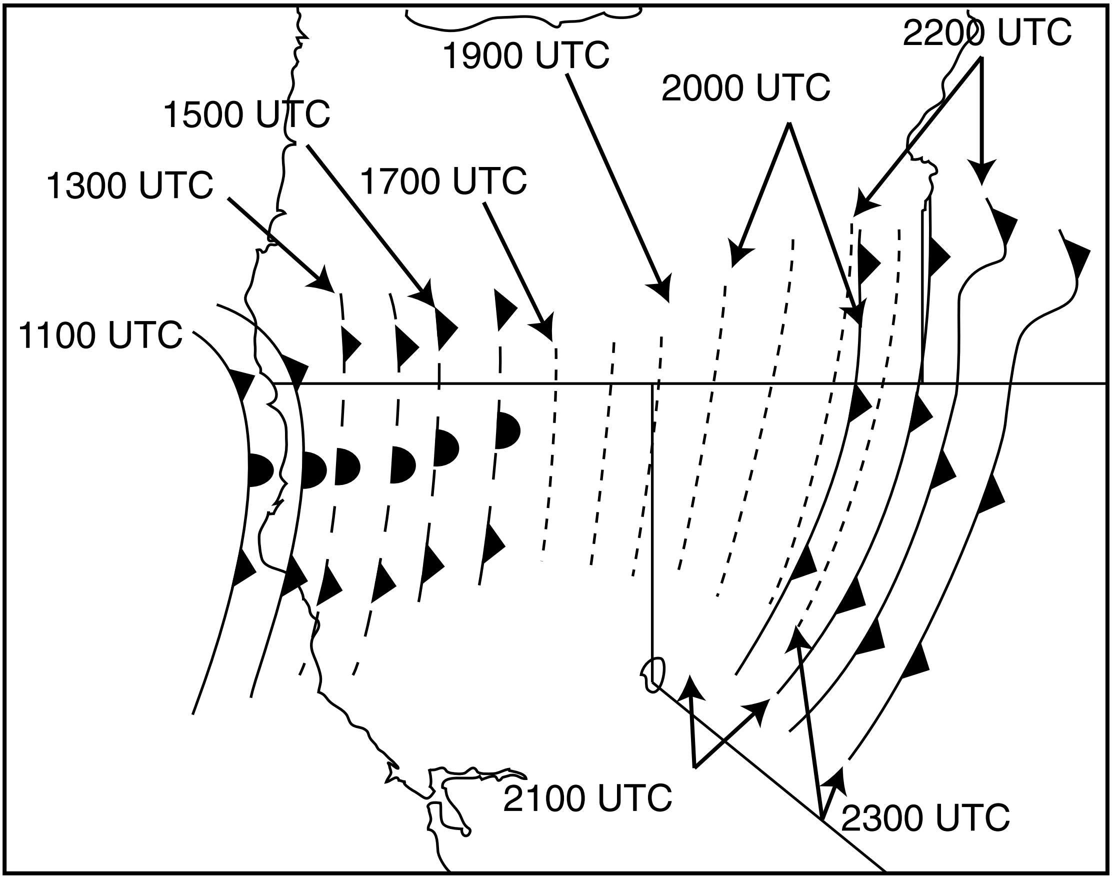

Fig. 11.7 Isochrones of a sequence of frontal positions over the northwest USA. Dashed occluded fronts are decaying, and dashed lines without frontal symbols represent trough positions. Adapted from Steenburgh *et al.* (2009).

11.2.3 Miscellaneous special plotting diagrams

A variety of special data-plotting diagrams can be used to help the modeler interpret output. Many of these have been described already in the context of the verification of forecasts, but a few are worth mentioning again. For example, there are many types of thermodynamic diagrams, such as the skew *T*–log *P* plot, that can be used to display the vertical distribution of variables based on model output and observations. From these diagrams can be inferred many important properties of the atmosphere, such as static stability and associated variables such as CAPE and CIN, and the structure of the boundary layer. Another type of special display approach is the Taylor diagram in Fig. 7.4.

11.3 Mathematical methods for analysis of the structure of model variable fields

Rather than discussing the mathematical foundations for the following methods for analyzing model output, basic concepts and applications will be stressed, as will sources of additional information.

11.3.1 Grouping atmospheric structures by pattern analysis

A variety of techniques exist for grouping atmospheric structures, whether the structures be defined based on model output, analyses of observations, or observations themselves. This process involves the automatic identification of recurring weather patterns in a large data set (e.g., composed of model forecasts or analyses) and the association of variable fields in the data set with one of the patterns. A manual and qualitative equivalent of this process would be for an analyst to sort through a large number of weather maps, say of the sea-level pressure, and put each map into a pile according to the locations of troughs and ridges, the amplitude of the wave pattern, the strength of the average pressure gradient, etc.

Applications of such an analysis process are many. They include the following.

- Define a regional climatology based on model-generated reanalyses – For each variable, the climatology would consist of gridded fields that represent the different prevailing patterns, accompanied by the frequency of occurrence of each pattern. Note that cataloguing the extreme patterns and their frequency may be just as important as documenting the more-common patterns.
- Verification of a model's treatment of regime transitions – Sequences of weather patterns that appear in the analyses of observations are compared with sequences that prevail in model forecasts. Differences can provide insight about regimes that are not represented by the model.
- Conventional model verification statistics can be computed separately for different prevailing weather regimes. This can offer insight into the different components of the model that contribute to the errors. For example, Fig. 9.7 shows model bias calculated

for two of the most-common warm-season weather regimes in the area of Athens, Greece: strong Etesian flow from the north, and weak large-scale flow with a prevailing sea-breeze circulation.

Disadvantages of these automated pattern-sorting techniques are as follows.

- The patterns are not sorted using any dynamical constraints, so there will likely be different underlying processes within a category.
- The patterns represent a composite of the individual analyses, and thus they may not be internally consistent in a kinematic or dynamic sense. Thus, a "typical day" is sometimes chosen from the archive, where the pattern for that day closely matches the composite. This analysis for the selected day will be internally consistent.
- Because there is no predefined number of groups, this is an arbitrary choice that must be made by the analyst without any knowledge of the number of natural clusterings. Thus, some trial and error will be involved.

Two of the most common approaches for such pattern analysis are referred to as *Self-Organizing Maps* (SOMs) and *cluster analysis*. See Wilks (2006) for a summary of meteorological applications of cluster analysis, and Marzban and Sandgathe (2006, 2008) for examples of applications related to model verification. And, Kohonen (2000) describes the general method of SOMs, and an example of one of its many applications for analyzing model simulations of weather and climate output is provided by Seefeldt and Cassano (2008).

Figure 11.8 shows an example of a SOMs analysis for a small number of categories. Six patterns have been arbitrarily chosen for this analysis of 0000 UTC 700-hPa winds, based on one year of model-generated reanalyses for an area of the Middle East. The patterns are distinct in terms of wind speed and/or direction, and the frequency of occurrence of each classification ranges from 10.1% to 26.1%. The frequency refers to the percentage of the analyses that fall into each category. Because the number of categories was arbitrarily chosen, further analysis could involve repeating the process with additional degrees of freedom to determine if there exists considerable variance within any of the original categories.

11.3.2 Finding coupled patterns in model or observational data

A different group of methods is aimed at finding coupled patterns in data from models or observations. Bretherton *et al.* (1992) and Wilks (2006) describe and compare three of the most commonly used methods: Principal Component Analysis (PCA), Canonical Correlation Analysis (CCA), and Singular Value Decomposition (SVD). Additional comparisons are found in articles by Hannachi *et al.* (2007) and Tippett *et al.* (2008).

Principal component analysis

This is also referred to as Empirical Orthogonal Function (EOF) analysis, where the objective of the mathematical procedure is to transform a data set containing a large number of

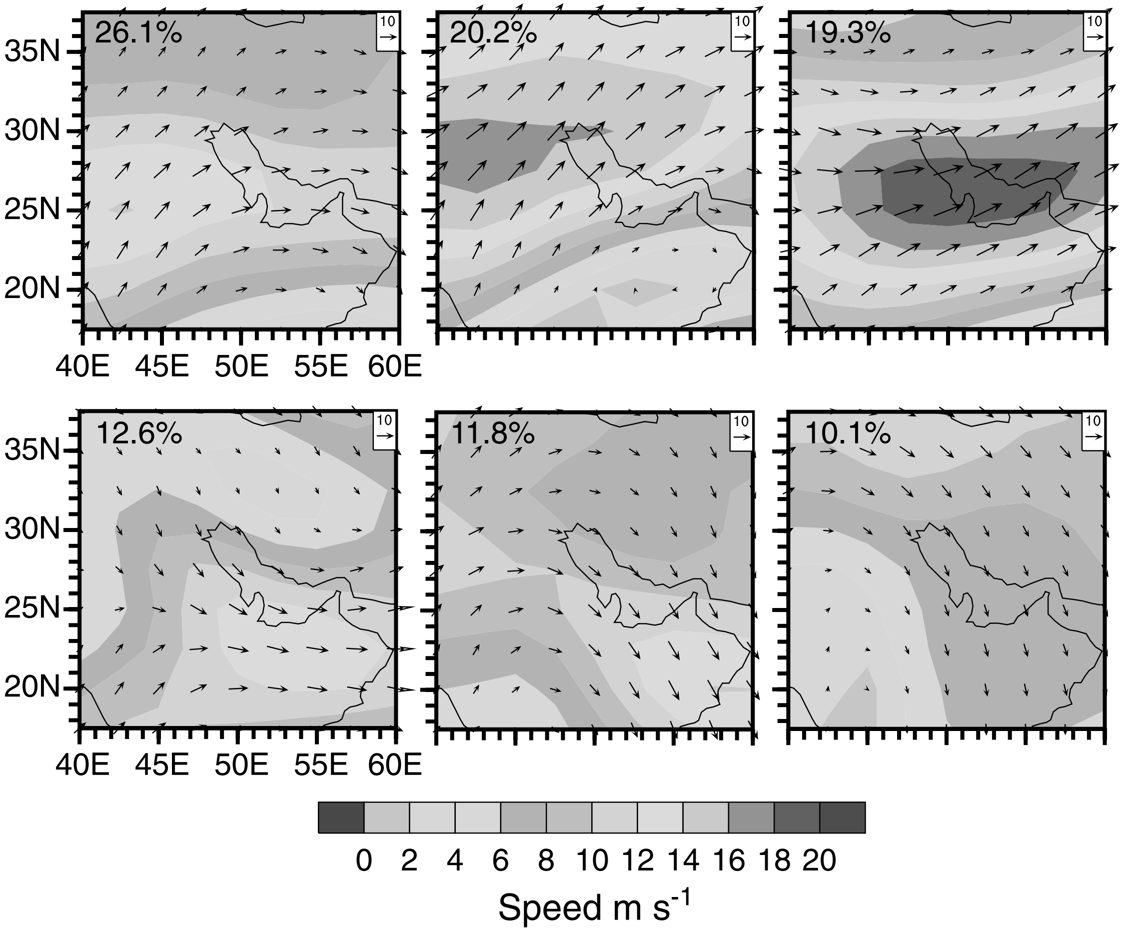

Fig. 11.8 Example of the use of SOMs for an analysis of 0000 UTC 700-hPa winds, based on one year of model-based analyses for an area of the Middle East. The percentage of the analyses that fall into each category is shown. Provided by Ming Ge, NCAR.

correlated variables into one containing many fewer uncorrelated variables, called principal components. The new variables are linear combinations of the original ones, with the first component being the linear combination that contains the largest variance, the second being the combination that contains the second largest variance, etc. Two or more variables can be combined in a PCA to reveal relationships between the fields.

There have been many applications of PCA in the atmospheric sciences. For example, Teng *et al.* (2007) use a form of PCA to show that an AOGCM supports three distinct circulation regimes, having a persistence period of about 7 days, and that analyses of observations have very similar regimes and persistence periods. The impact of greenhouse warming is interpreted in the context of changes in these regimes. Other studies include one by Smith *et al.* (2008) who use PCA to compare the diurnal cycles in a climate-model simulation and observations.

An early description of the PCA method in the context of applications in the atmospheric sciences was by Kutzbach (1967). Exhaustive treatments of PCA can be found in Preisendorfer (1988), which is aimed at geophysical applications, and in Jolliffe (2002), which is a more-general treatment.

Canonical correlation analysis

Canonical correlation analysis is applied to two multivariate data sets and identifies coupled variability between them. The two data sets can apply to the same time period, or there can be a lag between them. In the latter case, the relationships between the lagged variable fields can be used for statistical weather prediction. Indeed, this was the first meteorological application of CCA, where in most of the subsequent efforts the forecast time scales have tended to be interseasonal. For example Barnett and Preisendorfer (1987) relate seasonal-mean SST anomalies over the Pacific Ocean to surface air-temperature anomalies over the USA during the following season. See Bretherton *et al*. (1992) and Wilks (2006) for other examples.

Singular value decomposition

Singular value decomposition is similar to CCA in that it isolates combinations of variables in two fields that tend to be related to one another. See Bretherton *et al.* (1992) for references to example applications.

11.3.3 Spectral analysis

We have seen in Fig. 3.36 that the spectrum of model output can be computed, and interpreted to define the effective-resolution of the model. And, Section 9.9.2 discusses how a model solution can be verified in terms of its spatial spectrum. Here it will be shown that the same type of spectral decomposition can be used to interpret model output in creative ways. There is a variety of different wavelength bands that the modeler might desire to isolate using spectral analysis, but a common one is that which is associated with diurnal forcing. For example, Fig. 11.9 illustrates the variability (each dot is a location) in the amount of spectral power in this diurnal band. For time series of observations of 10-m AGL winds at each of 28 locations in a mountain valley, a spectral analysis separated the spectral energy into three bands: periods longer than the diurnal, a period of about 24 h, and periods of less than the diurnal. The diurnal power for each station is plotted against the average diurnal amplitude of the oscillation of the u component of the wind at that station. This spectral decomposition has allowed us to learn that there is a large station-to-station difference in the diurnal power, and that the greatest diurnal power is located at those stations near a canyon or near a mountain slope (open circles) where the thermally forced circulation should be the strongest.

A recent, related method is called wavelet analysis. If there is a long time series of values of model-simulated or observed variables, performing a Fourier transform will convert the series to frequency space. But, frequently the time series is not stationary in the sense that frequencies change with time. An option for defining the frequency spectrum for short

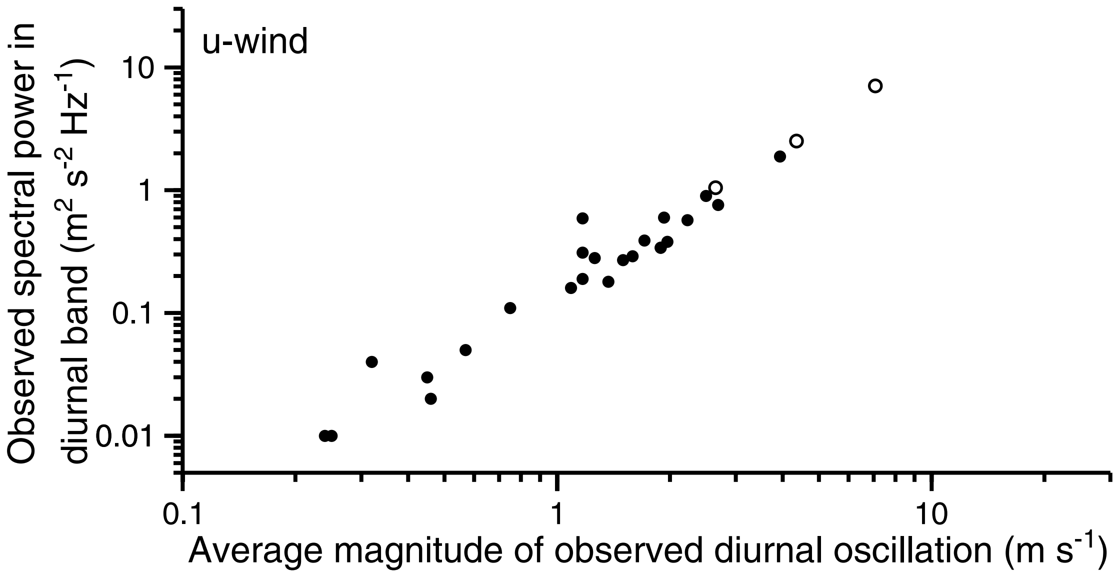

Fig. 11.9 Example of the amount of spectral power in the diurnal band in observations of 10-m AGL winds at each of 28 locations in a mountain valley. The diurnal power for each station is plotted against the average diurnal amplitude of the oscillation of the *u* component of the wind at that station. Stations close to mountain canyons have open circles. Adapted from Rife *et al.* (2004).

periods, in order to characterize this variability, is a short-period Fourier transform. However, wavelet analyses are more suitable for this, and are capable of providing the time and frequency information simultaneously, thus giving a time–frequency representation of the signal. Torrence and Compo (1998) offer a practical guide to wavelet analysis in the atmospheric sciences, and Wilks (2006) provides additional references.

11.4 Calculation of derived variables

The model dependent variables have been used to calculate many derived variables that are useful in understanding atmospheric processes. For example, based on the model winds, the vorticity and divergence can be computed. Or, frontogenesis terms can be calculated. Showing the geostrophic and ageostrophic wind vectors can reveal interesting circulations. See one of the previously noted standard references on meteorological analysis for further examples.

11.5 Analysis of energetics

The analysis of model energetics was discussed in Chapter 3 in the context of ensuring that there are no erroneous sources or sinks of energy in the model. However, the calculation of energy terms and conversions based on the gridded output from model simulations, forecasts, or reanalyses can reveal physical processes as well as differences among models. Energy can be partitioned into kinetic, internal, and potential components, and into components that are associated with the mean flow and eddies that are superimposed on the mean flow. The potential energy can be separated into both available and unavailable components; the available part is convertible to kinetic energy while the unavailable part is related to the equilibrium base state of the system and cannot be converted.

The time-dependent equations for the different energy components are derived from the same fundamental equations that are the basis for the model. The rates of change in the energy components are then obtained by applying the model-defined values of the dependent variables in the terms on the right side of the energy equations.

Two general approaches to the calculation and display of energy-conversion terms involve the use of grid-averaged values as well as local, instantaneous formulations. As an example of the latter method, referred to as the "local energetics" (Orlanski and Katzfey 1991) approach, the eddy kinetic energy tendency would be represented on a grid, and displays of that quantity would be provided.

Energetics analyses have been performed in many types of studies. A few of the classes of studies, with references, are listed below.

- Midlatitude cyclogenesis – Orlanski and Katzfey (1995), Lackmann *et al.* (1999), Lapeyre and Held (2004), Moore and Montgomery (2004, 2005)
- Storm tracks in IPCC AOGCM simulations – Laîné *et al.* (2009)
- Madden–Julian oscillation in a climate model – Mu and Zhang (2006)
- Hydrostatic and geostrophic adjustment to thermal forcing – Fanelli and Bannon (2005)

SUGGESTED GENERAL REFERENCES FOR FURTHER READING

Bluestein, H. (1992). *Synoptic-dynamic Meteorology in Midlatitudes. Vol. 1: Principles of Dynamics and Kinematics*. New York, USA: Oxford University Press.

Bluestein, H. (1992). *Synoptic-dynamic Meteorology in Midlatitudes. Vol. 2: Observations and Theory of Weather Systems*. New York, USA: Oxford University Press.

Saucier, W. J. (1955). *Principles of Meteorological Analysis*. Chicago, USA: University of Chicago Press.

PROBLEMS AND EXERCISES

1. Survey a variety of journal articles, and list and describe the various types of plots that are used to display model output and compare it with observations.
2. Using the shallow-fluid model employed in the problems of Chapter 3, calculate the trajectory of a parcel of air at the surface of the fluid, as a gravity wave propagates past.
3. Access the websites of national and international modeling centers, both those doing operational forecasting and research, and describe the types of plots that are used to display model products.
4. What fundamental differences should exist between analysis methods that are designed for use by operational forecasters and those designed for researchers?
5. Using a hypothetical wind-field pattern of your own creation, illustrate Lagrangian and Eulerian approaches to its characterization.