1 An Overview of Numerical Weather Prediction

1.1 Introduction

In this chapter, we give a historical overview of numerical weather prediction (NWP), which makes possible our daily and weekly forecasts.* Even seasonal forecasts become possible when ocean models are coupled to atmospheric models using similar methodologies to predict El Niño. In general, the public is not aware that our weather forecasts start out as initial-value problems on the supercomputers of the major national and international weather services or that the quality of the operational forecasts has undergone extraordinary improvements since their beginnings in the 1950s. These improvements are one of the most remarkable successes in the history of science, and the goal of this book is to clearly describe the major scientific developments that led to these improvements.

Numerical weather prediction provides the basic guidance for weather forecasting beyond the first few hours. For example, in the USA, computer weather forecasts issued by the National Centers for Environmental Prediction (NCEP) in College Park, MD, guide forecasts from the US National Weather Service (NWS). The NCEP forecasts are performed by "running" (i.e., integrating in time) computer models of the atmosphere that can simulate, given today's weather observations, the evolution of the atmosphere in the next few days. Because the time integration of an atmospheric model is an initial-value problem, in order to make a skillful forecast it is necessary that (a) *the computer model be an accurate representation of the atmosphere* and (b) *the initial conditions be also represented accurately*.

The NCEP (formerly the National Meteorological Center, or NMC) has performed operational computer weather forecasts since the 1950s. From 1955 to 1973, the forecasts included only the Northern Hemisphere (NH); they have been global since 1973. Over the years, the quality of the models and methods for using atmospheric observations has improved continuously, resulting in major forecast improvements. Weather centers have always kept track of the quality of the forecasts by comparing them with what actually happened, i.e., by comparing the forecast maps with the verification maps.[1] Teweles and Wobus (1954) developed a measure of forecast skill known as the

* Note: Sections marked with an asterisk are written at an undergraduate level.

[1] Unfortunately, this is not the case in other sciences that issue periodic forecasts. For example, economic forecasts are not routinely verified, even though they are obviously wrong most of the time.

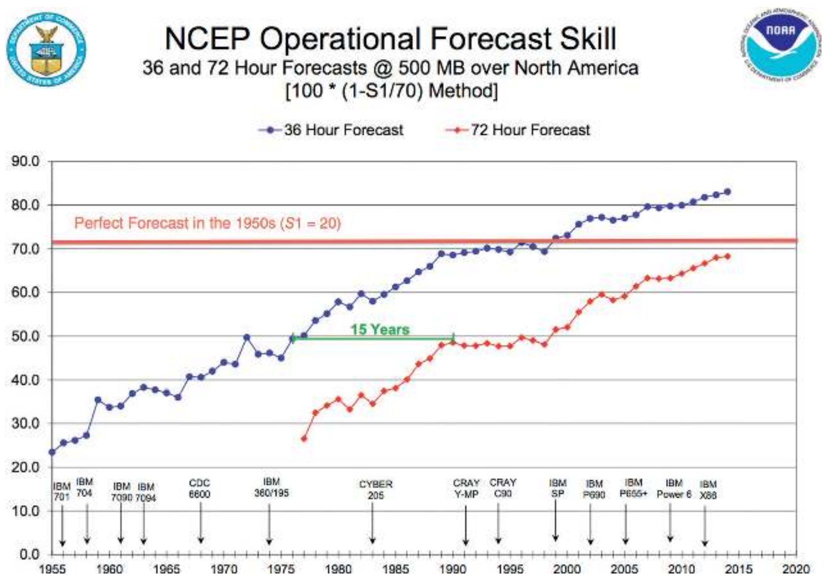

Figure 1.1.1 Historic evolution of the operational forecast skill of the NCEP (formerly NMC) models over North America. The score $100 * (1 - S1/70)$ is based on the $S1$ score that measures the relative error in the horizontal pressure gradient, averaged over the region of interest. The values $S1 = 70\%$ and $S1 = 20\%$ were empirically determined to correspond respectively to a "useless" and a "perfect" forecast when the score was designed. Note that the 72 hr forecasts are currently as skillful as the 36 hr forecasts were 10–20 years earlier (data courtesy S. Lilly, NCEP).

$S1$ score, which measures the relative error in the wind forecast through the estimation of the relative error in the forecast of the pressure gradient. Shuman (1989) pointed out that the $S1$ score was carefully calibrated to reflect the estimation of human forecasters of how useful the computer forecasts actually were. Forecasts with scores of $S1 = 70$ or larger were considered to be useless, and a score of $S1 = 20$ was considered a "perfect forecast," since that was the average score obtained when comparing hand-analyzed maps made by different expert weather analysts using the same observations over the data-rich area of North America. The NCEP still maintains a scaled version of the $S1$ score that reflects these limits: $100 \times (1 - S1/70)$, so that a "perfect score" of $S1 = 20$ is equal to about 71.4 in the scaled $S1$ score. Figure 1.1.1 shows the longest available record of the skill of NWP, measuring the scaled $S1$ score of 36 hr forecasts over North America at the constant pressure surface of 500 hPa (in the middle of the atmosphere, since the mean sea level pressure is about 1,000 hPa, or 1,000 mb). It is remarkable that, using the 1950s standard, the 36 hr forecast has been "perfect" since 1999. In the mid-1970s, NCEP started producing 72 hr forecasts, also shown in Figure 1.1.1, whose skill was comparable to that of the 36 hr forecasts made only 15–20 years earlier. The trend of the skill suggests that in a few years, the 500 hPa 72 hr forecast might also reach the "perfect" level, making it as close to the verifying analysis as two expert human analyses of the same abundant rawinsondes. This figure also includes information about the dates at which new supercomputers were installed at NCEP, since more powerful computers allowed the implementation of more accurate models and methods of data assimilation.

Other major operational centers such as the UK Meteorological Office (Met Office), the Japanese Meteorological Agency (JMA), and the Canadian Meteorological Centre (now Environment Canada, EC) made similar progress, and this shared experience led to the foundation, in 1976, of the European Centre for Medium-Range Weather Forecasts (ECMWF). As indicated by the name, its mission was to improve "medium-range weather forecasts" defined as 3 to 10 days. The ECMWF started issuing 10-day operational forecasts in 1979 and soon became the preeminent NWP operational center. It adopted a different skill measure, the anomaly correlation (AC), which computes the *pattern correlation* between the "anomalies" of a forecast map and of the corresponding verifying analysis (Miyakoda et al., 1976). Anomalies are defined as the difference between a field and its monthly climatology, so that the AC does not give credit to a forecast for just predicting climatology (e.g., predicting that winter is colder than summer, and higher latitudes colder than the tropics).

If the forecast is initialized from the analysis, the AC starts at 100%. As the forecast length increases, it becomes less skillful, and the AC becomes smaller. As they did with the $S1$ score, human forecasters calibrated the skill of the ECMWF forecasts by estimating that in order to provide useful forecast guidance, a forecast needs to have an AC larger than 60%. Figure 1.1.2 shows the historic ECMWF evolution of forecast skill in the NH extratropics measured by the time at which the AC reaches 60%. The figure shows that in 1980, on average, the forecasts "remained useful" until 5 to 6 days. The improvement in the quality of the forecasts is remarkable: Currently the ECMWF forecasts remain useful on average until 8 to 9 days and during some months until 10 days. This evolution in skill is paralleled by other centers (Figure 1.1.3).

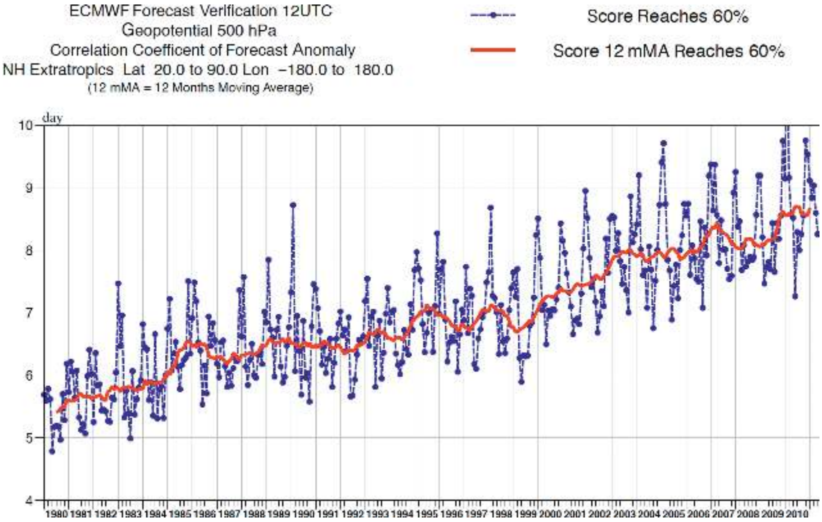

Figure 1.1.2 ECMWF forecast time at which the anomaly correlation (AC) reached the level of 60% (Figure courtesy of ECMWF, under the Licenses of CC-BY-4.0 and ECMWF Terms of Use).

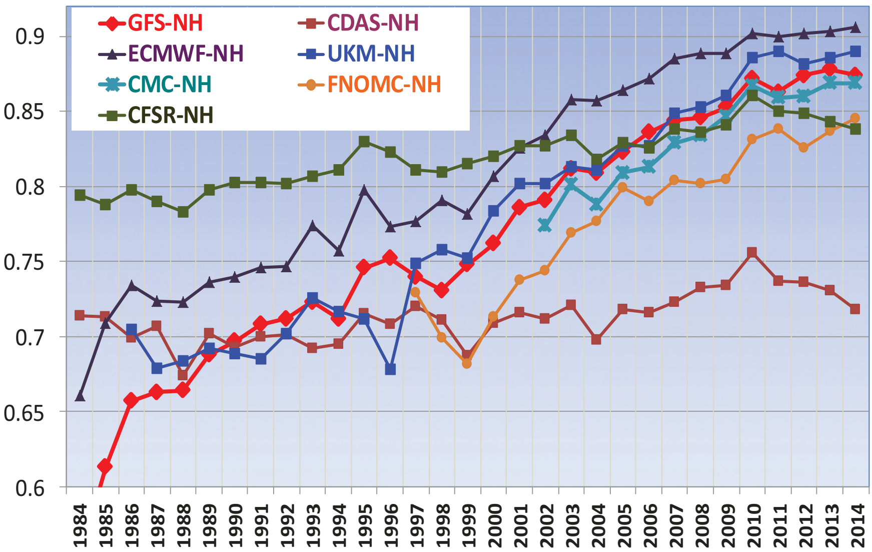

Figure 1.1.3 Evolution of the annual mean of the NH 500 hPa geopotential height 5-day anomaly correlation from 1984 to 2014. Included are the NCEP Global Forecasting System (GFS), the ECMWF, the UKMO, the Canadian Meteorological Centre (CMC), and the Fleet Numerical Meteorology and Oceanography Center (FNMOC, since 1997) operational model. Also included are two NCEP reanalyses run with frozen systems: the Climate Data Assimilation System (CDAS), also known as NCEP-NCAR Reanalysis, which is run with a system similar to the NCEP operational model circa 1995, and the Coupled Forecast System Reanalysis (CFSR), run with a Coupled Forecasting System circa 2005. Note that the two reanalyses, which use a "frozen" model and a "frozen" data assimilation system, show a skill that is almost constant, with only a slight improvement in time associated with the number of observations, and the improvement with time of their coverage and quality (figure courtesy of Fanglin Yang, NCEP).

These remarkable improvements in the skill of NWP are an extraordinary scientific achievement. They also demonstrate the benefits of national and international cooperation in sharing observations and new developments, as well as the impact of the friendly competition among scientists that try different methods in different research and operational centers, until eventually, the most successful two to three methods are selected.

In the USA, research on NWP takes place in the national laboratories of the National Oceanic and Atmospheric Administration (NOAA), the National Aeronautics and Space Administration (NASA), and the National Center for Atmospheric Research (NCAR), as well as in universities and centers such as the Center for Analysis and Prediction of Storms (CAPS) and the Weather and Chaos group at the University of Maryland. Internationally, major research takes place in large operational national and international centers (such as the ECMWF, NCEP, and the weather services of the UK, France, Germany, Scandinavian and other European countries, Canada, Japan, Australia, Korea, and others). In meteorology, there has been a long tradition of sharing both data and research improvements, with the result that progress in the science

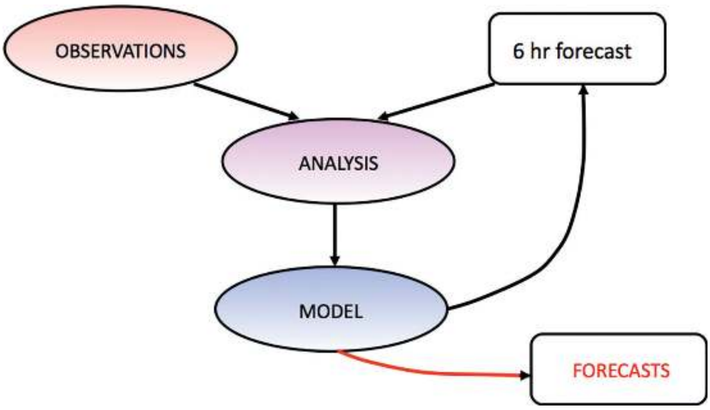

Figure 1.1.4 Schematic of the classic 6 hr analysis cycle that provides the initial conditions to the model. A 6 hr forecast is the first guess or background for the next analysis done 6 hr later and obtained by the optimal combination of the first guess and the new observations. The operational forecasts in red (which are two weeks long at NCEP) are used as guidance by the human forecasters.

of forecasting has taken place on many fronts, and all countries have benefited from this progress.

In order to understand how this scientific progress was achieved, let's consider the major components of NWP, namely the atmospheric *observations*, the *model*, and the *method used to create the initial conditions* for the model forecast. The initial conditions are known as the "analysis" because at the beginning of NWP, they were created by a human analyst who drew isolines of pressure, temperature, and wind, and this "analysis" was then interpolated to the model grid points. It soon became clear that rather than interpolating the analysis based on observations, it was better to interpolate the difference between the new observations obtained every 6 hr and the 6 hr forecast started from the previous analysis. This difference, which represents the new information brought by the observations, is also known as *analysis increment* or *innovation*. Thus was born the analysis cycle (Figure 1.1.4).

This type of "analysis cycle" where short forecasts are combined with new observations to obtain the best estimate of the state of the atmosphere (the analysis) came to be known as *data assimilation*, because the model is forced to "assimilate" observations of the atmosphere and thus to remain close to the real atmosphere. It is clear that in order to improve the forecasts, it is necessary to improve simultaneously the models, the observations, and the methods used to create the analysis (data assimilation). For example, improving the number and quality of the observations alone would not produce a good forecast if at the same time we use a mediocre model or initial conditions that come from an inaccurate analysis. In summary, the remarkable improvement in skill of NWP over the last half-century apparent in Figures 1.1.1–1.1.3 is due to four factors, the first two associated with model improvements, and the last two with improvements in data assimilation and in the quantity and quality of the observations:

- The increased power of supercomputers, allowing much *finer numerical resolution* and *fewer approximations* in the operational atmospheric models;
- The improved representation of *small-scale physical processes* (clouds, precipitation, turbulent transfers of heat, moisture, momentum, radiative transfer) within the models;
- The use of more accurate methods of data assimilation, which result in improved *initial conditions* for the models (analyses); and
- The increased availability and accuracy of observations, as well as advances in the use of satellite and aircraft data over oceans and the Southern Hemisphere.

This book is devoted to the exploration of NWP science (and similar approaches more recently used in ocean forecasting). This chapter contains a historical overview of NWP and an introduction to the main scientific advances that made possible the remarkable improvements we have seen and will continue to see in the future. Chapter 2 includes a derivation of the equations used in NWP and in ocean modeling, as well as the type of wave solutions that they allow. The presence of high-frequency gravity waves and sound waves in the equations of motion impose the use of very short time steps, which would be computationally extremely expensive. Chapter 2 also presents several approximations to the equations of motion that filter gravity and sound waves, such as the use of the hydrostatic or the anelastic approximations, and the limits in accuracy introduced by these approximations. Chapter 3 provides an introductory practical guide to the numerical methods used in atmospheric models, including finite differences, finite volume, and spectral schemes used in space discretization of the dynamical equations. Time schemes appropriate for different types of equations, as well as special methods such as semi-implicit and semi-Lagrangian schemes are also described. We also discuss the type of models (such as the Finite Volume, Cubed Sphere, or FV3) that are replacing the currently widely used global spectral models, since they become inefficient at horizontal grid scales smaller than about 10 km, which also require the use of nonhydrostatic equations. The textbook Durran (2010) covers this subject in more detail. Chapter 4 is a brief introduction to the problem of representing the impact (parameterizing) of those physical processes that take place at subgrid scales and that are not resolved by the model, such as radiation and turbulent transports of heat, moisture, and momentum. Since these subgrid-scale processes cannot be explicitly resolved by the numerical discretizations presented in Chapter 3, and they cannot be ignored beyond a few hours, they instead need to be parameterized. The textbook Stensrud (2007) is devoted to this subject. The new approach of "super-parameterizations" is briefly presented. Chapter 5, perhaps the most important in the book, is devoted to data assimilation, one of the pillars of NWP, where much progress has been made, after the early empirical methods to correct the short-range forecasts with observations were replaced with methods that accounted for the error statistics of the forecasts and observations. The first "statistical interpolation" methods (optimal interpolation and 3D-Var) assumed that the forecast error covariance is constant with time. More advanced methods (4D-Var and ensemble Kalman filter) can now account for the "errors of the day." All these methods require involved matrix algebra, so that they are introduced with "baby" scalar examples that make clear the interpretation of the otherwise very complex equations. We also review some recent developments in applications of data assimilation, where the pioneering NWP research is being explored in other sciences. Chapter 6 is devoted to chaos and predictability, ensemble forecasting, coupled modeling, and an introduction to climate change modeling. Here we respond to questions such as: "Can we take advantage of the variability of chaos and predict with skill some states of the atmosphere longer than others? How could we possibly predict climate change when we know that we cannot predict the weather more than two weeks ahead? Should we model and bidirectionally couple the human system with the Earth system models that we use to predict climate change?"

In this introductory chapter, we give an overview of the major components and milestones in numerical forecasting. They will be discussed in detail in the corresponding Chapters 2–6.

1.2 Early Developments

There are several outstanding scientists whose work was essential to the creation of NWP, but for brevity we just refer to the three most important giant advances:*

Vilhelm Bjerknes (1862–1951) was a Norwegian physicist who developed a "general circulation theorem" for the atmosphere published in 1900 in both *Monthly Weather Review and Meteorologische Zeitschrift*. He felt that empirical and statistical methods used at that time would never lead to a scientific approach for meteorology. Bjerknes set up the scientific basis for weather prediction in a subsequent paper entitled "The problem of weather forecasting as a problem in mechanics and physics," published in 1904 in *Meteorologische Zeitschrift* (Bjerknes, 1904). He did not use equations but posed the problem of NWP as integrating seven equations with seven unknowns using observations as initial data (Bjerknes, 1904; Gramelsberger, 2009). Since numerical solutions were at that time out of the question, he proposed instead the use of graphical methods of numerical integration. With his son Jacob Bjerknes (1897–1975) he established the famous Bergen School of Meteorology where the frontal theory for cyclone development was first developed and became the basis of synoptic meteorology. Jacob Bjerknes' amazing physical intuition made possible the prediction of El Niño: In 1969, he explained for the first time the phenomenon of El Niño as the ocean component of a coupled ocean–atmosphere oscillation, now referred to as the El Niño–Southern Oscillation, or ENSO.

L. F. Richardson (1881–1953) was also a physicist (and a pacifist) who refused to fight during World War I, instead driving a Quakers' ambulance. During the two years he spent at war, he developed the appropriate finite differences for a primitive equations model, solved many theoretical and practical problems in the physical and numerical formulations, and computed by hand the time derivative of the surface pressure for a single point in Germany, using maps created by Bjerknes (Richardson, 1922). Even though his methodology was correct, the result he obtained was catastrophically wrong: He predicted a huge, unrealistic change of 145 hPa in 6 hr, while the observed pressure hardly changed in those 6 hr. The reasons for this failure (the presence of high-frequency gravity waves that dominated the time derivative in the solution) are discussed in detail in Chapter 2. Despite this dramatic failure that kept other scientists from attempting further experiments in NWP, Richardson published an inspiring book (Richardson, 1922), describing in detail his methodology, which, as shown by Lynch (2006), was scientifically impeccable.

Charney, Fjortoft, and von Neumann succeeded for the first time in predicting a moderately realistic 24 hr change in the atmosphere (Charney et al., 1950) by using, instead of the primitive equations as used by Richardson, the conservation of vorticity equation based on the quasi-geostrophic theory Charney had developed (Charney, 1948). This was a spectacular success (Figure 1.2.2) made possible by the access to a computer (the ENIAC) and especially by the choice of a simple but realistic equation to describe the atmospheric dynamics most relevant to weather prediction (the conservation of vorticity). Charney made this choice in order to *filter out the high-frequency gravity waves* that ruined Richardson's results (see Chapter 2).

Jule G. Charney (1917–1981) was one of the giants in the history of NWP. In his paper "Dynamic forecasting by numerical process" (Charney, 1951), he introduced the subject of this book (NWP) as well as it could be introduced today. We reproduce here parts of the paper (with emphasis added):

> As meteorologists have long known, *the atmosphere exhibits no periodicities of the kind that enable one to predict the weather in the same way one predicts the tides*. No simple set of causal relationships can be found which relate the state of the atmosphere at one instant of time to its state at another. It was this realization that *led V. Bjerknes (1904) to define the problem of prognosis as nothing less than the integration of the equations of motion of the atmosphere* (Bjerknes, 1904).[2] But it remained for Richardson to suggest the practical means for the solution of this problem (Richardson, 1922). *He proposed to integrate the equations of motion numerically and showed exactly how this might be done*. *That the actual forecast used to test his method was unsuccessful was in no way a measure of the value of his work*. In retrospect it becomes obvious that the inadequacies of observation alone would have doomed any attempt, however well conceived, a circumstance of which Richardson was aware. The real value of his work lay in the fact that it crystallized once and for all the essential problems that would have to be faced by future workers in the field and it laid down a thorough groundwork for their solution.

[2] The importance of Bjerknes (1904) is clearly described by Thompson (1990), another pioneer of NWP, and the author of the first and still inspiring text on NWP (Thompson, 1961). His paper "Charney and the revival of numerical weather prediction" (Thompson, 1990) contains extremely interesting material on the history of NWP as well as on early computers: It was not until 1904 that Vilhelm Bjerknes – in a remarkable manifesto and testament of deterministic faith – stated the central problem of NWP. This was the first explicit, coherent recognition that the future state of the atmosphere is, *in principle*, completely determined by its detailed initial state and known boundary conditions, together with Newton's equations of motion, the Boyle–Charles–Dalton equation of state, the equation of mass continuity, and the thermodynamic energy equation. Bjerknes went further: He outlined an ambitious, but logical, program of observation, graphical analysis of meteorological data, and graphical solution of the governing equations. He succeeded in persuading the Norwegians to support an expanded network of surface observation stations, founded the famous Bergen School of synoptic and dynamic meteorology, and ushered in the famous polar front theory of cyclone formation. Beyond providing a clear goal and a sound physical approach to dynamical weather prediction, Bjerknes instilled his ideas in the minds of his students and their students in Bergen and in Oslo, three of whom were later to write important chapters in the development of NWP in the USA (Rossby, Eliassen, and Fjörtoft).

For a long time no one ventured to follow in Richardson's footsteps. The paucity of the observational network and the enormity of the computational task stood as apparently insurmountable barriers to the realization of his dream that one day it might be possible to advance the computations faster than the weather. But with the increase in the density and extent of the surface and upper-air observational network on the one hand, and the development of large-capacity high-speed computing machines on the other, interest has revived in Richardson's problem, and attempts have been made to attack it anew.

These efforts have been characterized by a devotion to objectives more limited than Richardson's. Instead of attempting to deal with the atmosphere in all its complexity, one tries to be satisfied with *simplified models* approximating the actual motions to a greater or lesser degree. By *starting with models incorporating only what are thought to be the most important of the atmospheric influences*, and by gradually bringing in others, one is able to proceed inductively and thereby to avoid the pitfalls inevitably encountered when a great many poorly understood factors are introduced all at once.

A necessary condition for the success of this stepwise method is, of course, that the first approximations bear a recognizable resemblance to the actual motions. Fortunately, the science of meteorology has progressed to the point where one feels that at least the main factors governing the large-scale atmospheric motions are well known. *Thus integrations of even the linearized barotropic and thermally inactive baroclinic equations have yielded solutions bearing a marked resemblance to reality*. At any rate, it seems clear that the models embodying the collective experience and the positive skill of the forecast cannot fail utterly. This conviction has served as the guiding principle in the work of the meteorology project at The Institute for Advanced Study [in Princeton, New Jersey] with which the writer has been connected.

As indicated by Charney, Richardson performed a remarkably comprehensive numerical integration of the full primitive equations of motion (Chapter 2). He used a horizontal grid of about 200 km, and four vertical layers of approximately 200 hPa, centered over Germany. Using the observations at 7 UTC (Universal Coordinate Time) on May 20, 1910, he computed the time derivative of the pressure in central Germany between 4 and 10 UTC. *The predicted 6 hr change was 146 hPa, whereas in reality there was essentially no change observed in the surface pressure*. This huge error was discouraging, but it was due mostly to the fact that the initial conditions were *not quasi-geostrophically balanced*, and therefore included large amplitude, fast-moving gravity waves that masked *the initial rate of change* of the meteorological signal in the forecast (Figure 1.2.1). Moreover, if the integration had been continued, it would have suffered "computational blow-up" due to the violation of the Courant–Friedrichs–Lewy (CFL) condition (see Chapter 3), which requires that the time step should be smaller than the grid size divided by the speed of the fastest traveling signal (in this case, horizontally moving sound waves, traveling at about 300 m/s).

Charney (1948, 1949) and Eliassen (1949) solved both of these problems by deriving "filtered" equations of motion, based on quasi-geostrophic (slowly varying) balance, which filtered out (i.e., did not include) gravity and sound waves and were based on pressure fields alone. Charney points out that this approach was justified by the fact that forecasters' experience was that they were able to predict tomorrow's weather from pressure charts alone:

In the selection of a suitable first approximation, Richardson's discovery that the horizontal divergence was an immeasurable quantity had to be taken into account. Here a consideration of forecasting practice gave rise to the belief that this difficulty could be surmounted: forecasts

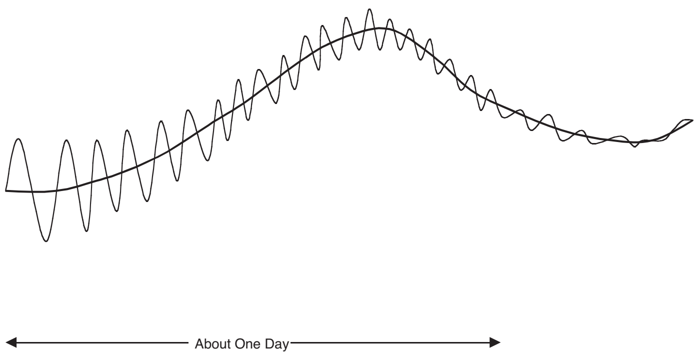

Figure 1.2.1 Schematic of a grid point forecast with slowly varying weather-related variations and superimposed high-frequency gravity waves that decay with time because they are dispersive, as in the shallow water experiment performed by Freeman described by Charney (1951). Note that even though the forecast of the slow waves is essentially unaffected by the presence of gravity waves, the initial time derivative is much larger in magnitude, as obtained by Richardson (1922) when he computed the initial time derivative at a grid point in a primitive equations model.

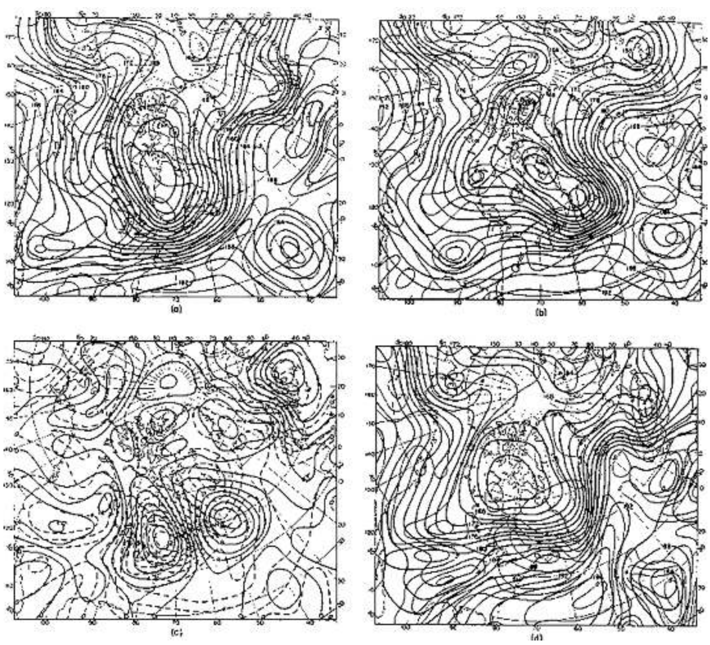

Figure 1.2.2 Forecast of January 30, 1949, 0300 GMT: (a) contours of observed z (height) and $\zeta + f$ (total vorticity) at $t = 0$; (b) observed z and $\zeta + f$ at $t = 24$ hr; (c) observed (continuous lines) and computed (broken lines) 24-hr height change; (d) computed z and $\zeta + f$ at $t = 24$ hr. The height unit is 100 ft and the unit of vorticity is $1/3 \times 10^{-4} s^{-1}$. Note that 1.2.2 (d) is the one day numerical forecast of 1.2.2 (b). ©American Meteorological Society. Used with permission