1

Historical overview of numerical weather prediction

1.1 Introduction

In general, the public is not aware that our daily weather forecasts start out as initial-value problems on the major national weather services supercomputers. 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 Center for Environmental Prediction (NCEP) in Washington, DC, guide forecasts from the US National Weather Service (NWS). NCEP forecasts are performed by running (integrating in time) computer models of the atmosphere that can simulate, given one day's weather observations, the evolution of the atmosphere in the next few days.[1] Because the time integration of an atmospheric model is an *initial-value problem*, the ability to make a skillful forecast requires both that *the computer model be a realistic representation of the atmosphere*, and that *the initial conditions be known accurately*.

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; 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.

1 In this book we will provide many examples mostly drawn from the US operational numerical center (NCEP), because of the availability of long records, and because the author's experience in this center facilitates obtaining such examples. However, these operational NCEP examples are only given for illustration purposes, and are simply representative of the evolution of operational weather forecasting in all major operational centers.

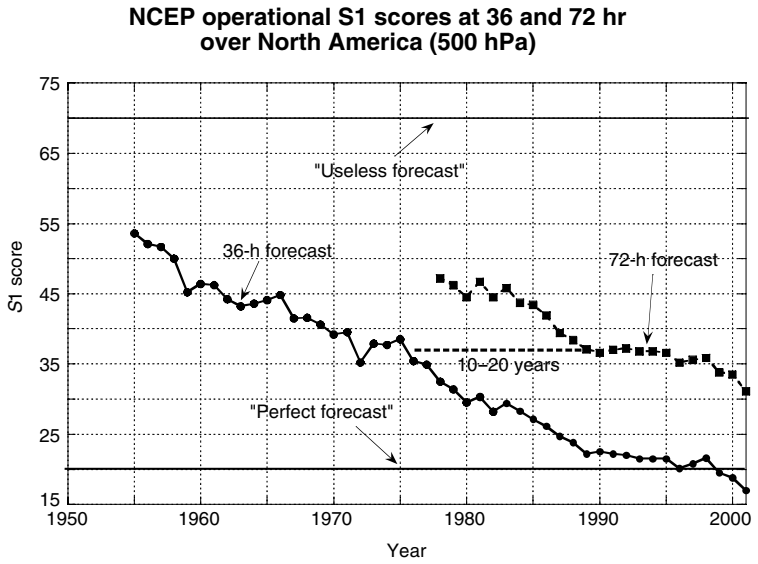
Figure 1.1.1(a) shows the longest available record of the skill of numerical weather prediction. The "$S1$" score (Teweles and Wobus, 1954) measures the relative error in the horizontal gradient of the height of the constant pressure surface of 500 hPa (in the middle of the atmosphere, since the surface pressure is about 1000 hPa) for 36-h forecasts over North America. Empirical experience at NMC indicated that a score of 70% or more corresponds to a useless forecast, and a score of 20% or less corresponds to an essentially perfect forecast. This was found from the fact that 20% was the average $S1$ score obtained when comparing analyses hand-made by several experienced forecasters fitting the same observations over the data-rich North American region.

Figure 1.1.1(a) shows that current 36-h 500-hPa forecasts over North America are close to what was considered essentially "perfect" 40 years ago: the computer forecasts are able to locate generally very well the position and intensity of the large-scale atmospheric waves, major centers of high and low pressure that determine the general evolution of the weather in the 36-h forecast. The sea level pressure forecasts contain smaller-scale atmospheric structures, such as fronts, mesoscale convective systems that dominate summer precipitation, etc., and are still difficult to forecast in detail (although their prediction has also improved very significantly over the years) so their $S1$ score is still well above 20% (Fig. 1.1.1(b)). Fig. 1.1.1(a) also shows that the 72-h forecasts of today are as accurate as the 36-h forecasts were 10–20 years ago. This doubling (or better) of skill in the forecasts is observed for other forecast variables, such as precipitation. Similarly, 5-day forecasts, which had no useful skill 15 years ago, are now moderately skillful, and during the winter of 1997–8, ensemble forecasts for the second week average showed useful skill (defined as anomaly correlation close to 60% or higher).

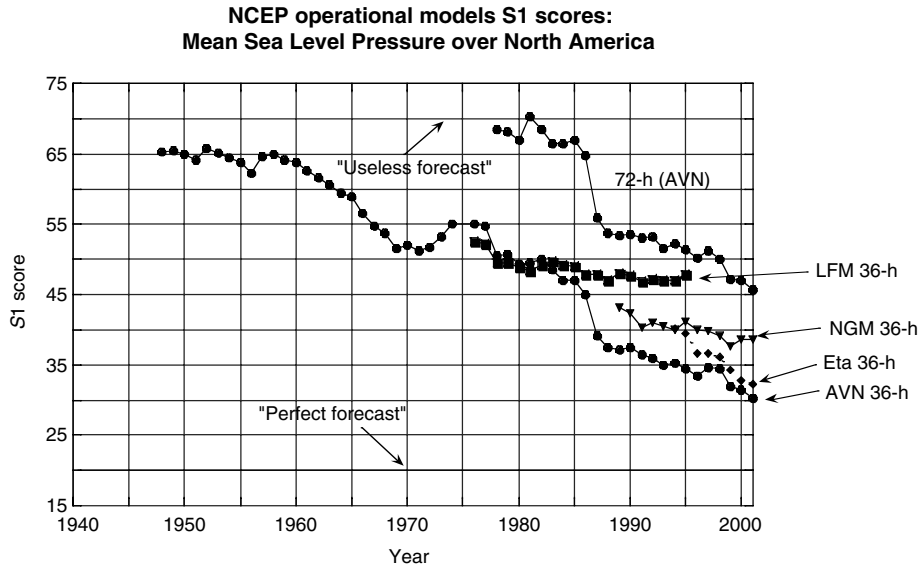
The improvement in skill of numerical weather prediction over the last 40 years apparent in Fig.1.1.1 is due to four factors:

- 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, and radiation) within the models;
- the use of more accurate methods of data assimilation, which result in improved initial conditions for the models; and
- the increased availability of data, especially satellite and aircraft data over the oceans and the Southern Hemisphere.

In the USA, research on numerical weather prediction 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), and in universities and centers such as the



(a)



(b)

Figure 1.1.1: (a) Historic evolution of the operational forecast skill of the NCEP (formerly NMC) models over North America (500 hPa). The $S1$ score 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-h forecasts are currently as skillful as the 36-h were 10–20 years ago (data courtesy C.Vlcek, NCEP). (b) Same as (a) but showing $S1$ scores for sea level pressure forecasts over North America (data courtesy C.Vlcek, NCEP). It shows results from global (AVN) and regional (LFM, NGM and Eta) forecasts. The LFM model development was "frozen" in 1986 and the NGM was frozen in 1991.

Center for Prediction of Storms (CAPS). Internationally, major research takes place in large operational national and international centers (such as the European Center for Medium Range Weather Forecasts (ECMWF), NCEP, and the weather services of the UK, France, Germany, Scandinavian and other European countries, Canada, Japan, Australia, 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 of forecasting has taken place on many fronts, and all countries have benefited from this progress.

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 following chapters.

1.2 Early developments

Jule G. Charney (1917–1981) was one of the giants in the history of numerical weather prediction. In his 1951 paper "Dynamical forecasting by numerical process", he introduced the subject of this book 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*.[2] But it remained for *Richardson (1922)* to suggest the practical means for the solution of this problem. *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

2 The importance of the Bjerknes (1904) paper is clearly described by Thompson (1990), another pioneer of NWP, and the author of a very inspiring text on NWP (Thompson, 1961a). His paper "Charney and the revival of NWP" 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, V. 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 US (Rossby, Eliassen and Fjörtoft).

> 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.
>
> 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 computation 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 it is 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 [at Princeton University] 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 20 May 1910, he computed the time derivative of the pressure in central Germany between 4 and 10 UTC. *The predicted 6-h 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 balanced*, and therefore included fast-moving gravity waves which masked the *initial rate of change* of the meteorological signal in the forecast (Fig. 1.2.1). Moreover, if the integration had been continued, it would have suffered "computational blow-up" due to the violation of the Courant–Friedricks–Lewy (CFL) condition

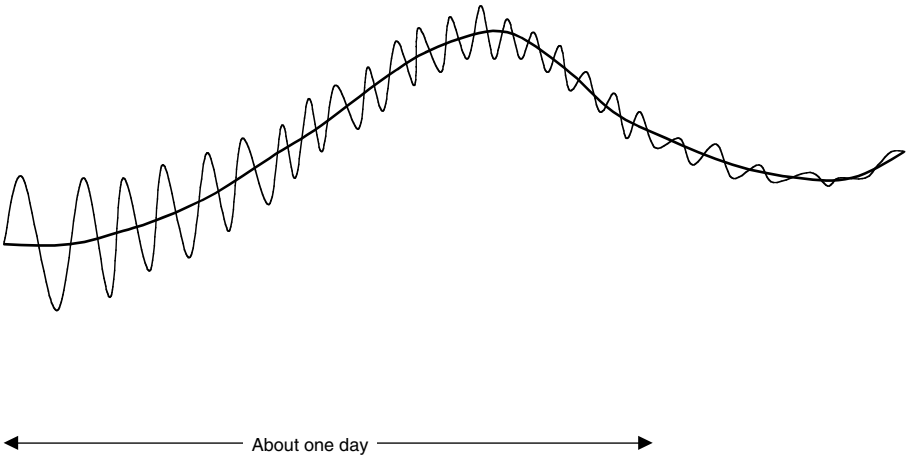


Figure 1.2.1: Schematic of a forecast with slowly varying weather-related variations and superimposed high-frequency gravity waves. Note that even though the forecast of the slow waves is essentially unaffected by the presence of gravity waves, the initial time derivative is much larger in magnitude, as obtained in the Richardson (1922) experiment.

(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 unmeasurable 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 were made by means of geostrophic reasoning from the pressure field alone – forecasts in which the concept of horizontal divergence played no role.

In order to understand better Charney's comment, we quote an anecdote from Lorenz (1990) on his interactions with Jule Charney:

> On another[3] occasion when our conversations had turned closer to scientific matters, Jule was talking again about the early days of NWP. For a proper

3 The previous occasion was a story about an invitation Charney received to appear on the "Today" show, to talk about how computers were going to forecast the weather. Since the show was at 7 am, Charney, a late riser, had never watched it. "He told us that he felt that he ought to see the show at least once before agreeing to appear on it, and so, one morning, he managed to pull himself out of bed and turn on the TV set, and the first person he saw was a chimpanzee.

> perspective, we should recall that at the time when Charney was a student, pressure was king. The centers of weather activity were acknowledged to be the highs and lows. A good prognostic chart was one that had the isobars in the right locations. Naturally, then, the thing that was responsible for the weather changes was the thing that made the pressure change. This was readily shown to be the divergence of the wind field. The divergence could not be very accurately measured, and a corollary deduced by some meteorologists, including some of Charney's advisors, was that the dynamic equations could not be used to forecast the weather.
>
> Such reasoning simply did not make sense to Jule. The idea that the wind field might serve instead of the pressure field as a basis for dynamical forecasting, proposed by Rossby, gave Jule a route to follow.[4] He told us, however, that what really inspired him to develop the equations that later became the basis for NWP was a determination to prove, to those who had assured him that the task was impossible, that they were wrong.

Charney, R. Fjørtoft, and J. von Neuman (1950) computed a historic first one-day weather forecast using a barotropic (one-layer) filtered model. The work took place in 1948–9. They used one of the first electronic computers (the Electronic Numerical Integrator and Computer, ENIAC), housed at the Aberdeen Proving Grounds of the US Army in Maryland. It incorporated von Neuman's idea of "stored programming" (i.e., the ability to perform arithmetic operations over different operands (loops) without having to repeat the code). The results of the first forecasts were quite encouraging: Fig. 1.2.2, reproduced from Charney (1951) shows the 24-h forecast and verification for 30 January 1949. Unlike Richardson's results, the forecast remains meteorological, and there is a pattern correlation between the predicted and the observed pressure field 24-h changes.

It is remarkable that in his 1951 paper, just after the triumph of performing the first successful forecasts with filtered models, Charney already saw that much more progress would come from the use of the primitive (unfiltered) equations of motion as Richardson had originally attempted:

> The discussion so far has dealt exclusively with the quasi-geostrophic equations as the basis for numerical forecasting. Yet there has been no intention to exclude the possibility that the primitive Eulerian equations can also be used for this purpose. *The outlook for numerical forecasting would be indeed dismal if the quasi-geostrophic approximation represented the upper limit of attainable accuracy, for it is known that it applies only indifferently, if at all, to many of the small-scale but meteorologically significant motions.* We have merely indicated two obstacles that stand in the way of the applications of the primitive equations:

He decided he could never compete with a chimpanzee for the public's favor, and so he gracefully declined to appear, much to the dismay of the computer company that had engineered the invitation in the first place" (Lorenz, 1990).

4 The development of the "Rossby waves" phase speed equation $c = U - \beta L^2/\pi^2$ based on the linearized, non-divergent vorticity equation (Rossby *et al*., 1939, Rossby, 1940), and its success in predicting the motion of the large-scale atmospheric waves, was an essential stimulus to Charney's development of the filtered equations (Phillips, 1990b, 1998).

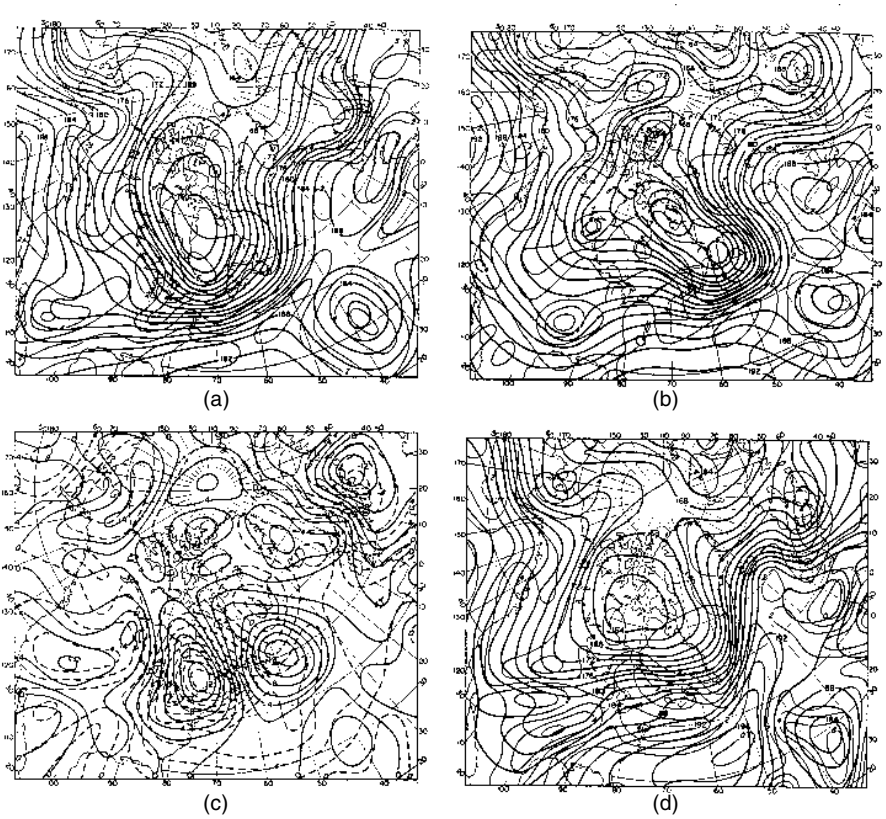


Figure 1.2.2: Forecast of 30 January 1949, 0300 GMT: (a) contours of observed z and $\zeta + f$ at $t = 0$; (b) observed z and $\zeta + f$ at $t = 24$ h; (c) observed (continuous lines) and computed (broken lines) 24-h height change; (d) computed z and $\zeta + f$ at $t = 24$ h. The height unit is 100 ft and the unit of vorticity is $1/3 \times 10^{-4}$ s^{-1}. (Reproduced from the *Compendium of Meteorology*, with permission of the American Meteorological Society.)

> First, there is the difficulty raised by Richardson that *the horizontal divergence cannot be measured with sufficient accuracy. Moreover, the horizontal divergence is only one of a class of meteorological unobservables which also includes the horizontal acceleration. And second, if the primitive Eulerian equations are employed, a stringent and seemingly artificial bound is imposed on the size of the time interval for the finite difference equations. The first obstacle is the most formidable, for the second only means that the integration must proceed in steps of the order of fifteen minutes rather than two hours.* Yet the first does not seem insurmountable, as the following considerations will indicate.

He proceeded to describe an unpublished study in which he and J.C. Freeman integrated barotropic primitive equations (i.e., shallow water equations, Chapter 2) which include not only the slowly varying quasi-geostrophic solution, but also fast gravity waves. They initialized the forecast assuming zero initial divergence, and compared the result with a barotropic forecast (with gravity waves filtered out). The results were similar to those shown schematically in Fig. 1.2.1: they observed

that over a day or so the gravity waves subsided (through a process that we call geostrophic adjustment) and did not otherwise affect the forecast of the slow waves. From this result Charney concluded that numerical forecasting could indeed use the full primitive equations (as eventually happened in operational practice). He listed in the paper the complete primitive equations in pressure coordinates, essentially as they are used in current operational weather prediction, but without heating (nonadiabatic) and frictional terms, which he expected to have minor effects in one- or two-day forecasts. Charney concluded this remarkable paper with the following discussion, which includes a list of the physical processes that take place at scales too small to be resolved, and are incorporated in present models through "parameterizations of the subgrid-scale physics" (condensation, radiation, and turbulent fluxes of heat, momentum and moisture, Chapter 4):

> *Nonadiabatic and frictional terms have been ignored in the body of the discussion because it was thought that one should first seek to determine how much of the motion could be explained without them. Ultimately they will have to be taken into account, particularly if the forecast period is to be extended to three or more days.*
>
> Condensational phenomena appear to be the simplest to introduce: one has only to add the equation of continuity for water vapor and to replace the dry by the moist adiabatic equation. Long-wave radiational effects can also be provided for, since our knowledge of the absorptive properties of water vapor and carbon dioxide has progressed to a point where quantitative estimates of radiational cooling can be made, although the presence of clouds will complicate the problem considerably.
>
> The most difficult phenomena to include have to do with the turbulent transfer of momentum and heat. A great deal of research remains to be done before enough is known about these effects to permit the assignment of even rough values to the eddy coefficients of viscosity and heat conduction. Owing to their statistically indeterminate nature, the turbulent properties of the atmosphere *place an upper limit to the accuracy obtainable by dynamical methods of forecasting*, beyond which we shall have to rely upon statistical methods. But it seems certain that much progress can be made before these limits can be reached.

This paper, which although written in 1951 has not become dated, predicted with almost supernatural vision the path that numerical weather forecasting was to follow over the next five decades. It described the need for objective analysis of meteorological data in order to replace the laborious hand analyses. We now refer to this process as data assimilation (Chapter 5), which uses both observations and short forecasts to estimate initial conditions. Note that at a time at which only one-day forecasts had ever been attempted, Charney already had the intuition that there was an *upper limit* to weather predictability, which Lorenz (1965) later estimated to be about two weeks. However, Charney attributed the expected limit to model deficiencies (such as the parameterization of turbulent processes), rather than to the chaotic nature of the atmosphere, which imposes a limit of predictability even if the model is perfect

(Lorenz, 1963b; Chapter 6). Charney was right in assuming that in practice model deficiencies, as well as errors in the initial conditions, would limit predictability. At the present time, however, the state of the art in numerical forecasting has advanced enough that, when the atmosphere is highly predictable, the theoretically estimated limit for weather forecasting (about two weeks) is occasionally reached and even exceeded through techniques such as ensemble forecasting (Chapter 6).

Following the success of Charney *et al.* (1950), Rossby moved back to Sweden, and was able to direct a group that reproduced similar experiments on a powerful Swedish computer known as BESK. As a result, the first operational (real time) numerical weather forecasts started in Sweden in September 1954, six months before the start-up of the US operational forecasts[5] (Döös and Eaton, 1957, Wiin-Nielsen, 1991, Bolin, 1999).

1.3 Primitive equations, global and regional models, and nonhydrostatic models

As envisioned by Charney (1951, 1962) the filtered (quasi-geostrophic) equations, although very useful for understanding of the large-scale extratropical dynamics of the atmosphere, were not accurate enough to allow continued progress in NWP, and were eventually replaced by primitive equation models (Chapter 2). The primitive equations are conservation laws applied to individual parcels of air: conservation of the three-dimensional momentum (equations of motion), conservation of energy (first law of thermodynamics), conservation of dry air mass (continuity equation), and equations for the conservation of moisture in all its phases, as well as the equation of state for perfect gases. They include in their solution fast gravity and sound waves, and therefore in their space and time discretization they require the use of smaller time steps, or alternative techniques that slow them down (Chapter 3). For models with a horizontal grid size larger than 10 km, it is customary to replace the vertical component of the equation of motion with its hydrostatic approximation, in which the vertical acceleration is considered negligible compared with the gravitational acceleration (buoyancy). With this approximation, it is convenient to use atmospheric pressure, instead of height, as a vertical coordinate.

The continuous equations of motions are solved by discretization in space and in time using, for example, finite differences (Chapter 3). It has been found that the accuracy of a model is very strongly influenced by the spatial resolution: in general, the higher the resolution, the more accurate the model. Increasing resolution, however, is extremely costly. For example, doubling the resolution in the three space dimensions also requires halving the time step in order to satisfy conditions for computational

5 Anders Persson (1999 personal communication) kindly provided the notes on the historical development of NWP in the USA and Sweden reproduced in Appendix A.