KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

VERSLAGEN

V - 350

R.C.J. Somerville

Outline of a primitive-equation numerical weather prediction system.

De Bilt 1980

Publikatienummer: K.N.M.I. V - 350 (MO)

Koninklijk Nederlands Meteorologisch Instituut,
Meteorologisch Onderzoek,
Postbus 201,
3730 AE De Bilt,
Nederland.

U.D.C.: 551.509.313 :
551.509.38

CONTENTS

1. Introduction
2. Analysis
3. Initialization
4. Numerics
5. Physics
6. Prognostics
7. Diagnostics

OUTLINE OF A PRIMITIVE-EQUATION NUMERICAL WEATHER PREDICTION SYSTEM

Richard C. J. Somerville

1980

1. Introduction

In planning the development of a new operational numerical weather prediction model, it is most important to consider the model not in isolation but as part of a forecasting system. The model is the central element in a chain which begins with observational data and ends with the final forecast products. The other vital elements in the system may thus be regarded as generalized input and output processes with respect to the model.

It is usual to divide the model input into two closely related subtopics: objective analysis and initialization. The model itself is conventionally separated into two subjects also: numerical and physical aspects. Finally, the output from the model is postprocessed not only into prognostic information for users but also into what may be called diagnostic information which feeds back into the forecasting system.

These six areas will be considered in this report and will be referred to for brevity by the following concise terms: analysis, initialization, numerics, physics, prognostics, and diagnostics. Of course, there are other elements of the system as well (e.g., observations, telecommunications, operational and computational aspects), but they are beyond the scope of this study. Instead, this report attempts only to provide some suggestions for planning a research and development program to produce a new numerical weather prediction system for the KNMI. Of course, the opinions expressed here are those of the author alone.

From a scientific viewpoint, the constraints on this planning exercise have been quite general. The new prediction system should offer the promise of superior performance in comparison to the existing system, in most or all of the main areas of application, which include general short-range forecasting guidance, as well as input to statistical forecasting techniques and to wave and surge models. The existing model is based on filtered equations, and it has been established that the new model is to be based on the primitive-equation limited-area model developed at the European Centre for Medium-Range Weather Forecasts (ECMWF). From a practical viewpoint, it is vital that the development and operation of the new system be feasible at the KNMI, with realistic constraints on computer and human resources.

No requirement has been imposed to force the new system to be compatible or consistent with the existing one in any way. On the other hand, there is no prohibition against using some elements of the existing system in developing the new one, if this should seem appropriate. By contrast, most new prediction systems historically are forced to develop in a gradual and evolutionary manner from their predecessors. The freedom to depart from this traditional pattern of development is thus a rare and valuable opportunity.

To preserve this freedom, it seems essential that the new forecasting system be designed and developed as a modular system. By this term is meant simply that to the extent practical, each element of the system (objective analysis scheme, initialization procedure, physical parameterizations, etc.) should be conceived as a module with standardized interfaces to the other parts of the system. Insofar as this modular principle is followed, it will then be possible to replace the individual elements of the system with alternative ones with a minimum of disruption to the system as a whole. This strategy will facilitate speedy implementation of the system in its initial configuration, as well as subsequent improvements.

In the final analysis, a numerical weather prediction system is largely an ensemble of computer programs. In practice, it is not possible to design and write complicated but efficient programs with such foresight and generality that replacing any one program will never require any changes in the others. If this idealized goal of complete modularity is kept in mind, however, then the task of developing a new forecasting system, conceived along the lines described in this report, will be greatly facilitated.

2. Analysis

Comparisons of alternative analysis schemes (e.g., Otto-Bliesner et al., 1977) have not yet progressed to the point where it can be proven convincingly that one analysis technique is consistently superior to all others in the sense that forecasts based on this one technique are always significantly better than other forecasts based on the same data analyzed differently. The Cressman (1959) correction scheme is still widely used in operational practice, for example, in many countries.

In principle, however, many advantages are to be found in multivariate statistical techniques based on the optimum interpolation method (e.g., Schlatter, 1975; Schlatter et al., 1976). A linear analysis by Phillips (1976) estimates the expected forecast error as a function of observational error covariance and the true spectrum, for various analysis techniques. Phillips concludes that analysis methods which do not explicitly take into account observational error statistics are ineffective in making use of data with errors comparable to or larger than the "first guess" field, while efficient optimum interpolation schemes can in principle derive useful information even from poor data.

It may be remarked that a characteristic of modern numerical weather prediction is that the forecast model is also part of the observing system, in that it provides a "first guess" for the analysis as well as a means of insuring time continuity between observations in various "four-dimensional data assimilation" modes. Thus one may expect improvements in the forecast to be reflected in the analysis, as well as vice versa. On the other hand, it is perfectly feasible to begin a primitive-equation forecast from a simply balanced field based on a Cressman-type analysis, and it may be wise for practical reasons to defer temporarily the introduction of a multivariate optimum interpolation analysis scheme which will itself require a considerable development effort. This question of priorities is discussed further in the next section in connection with initialization.

3. Initialization

The principal purpose of initialization procedures is to make adjustments in the initial state so that the model will not generate unrealistically large high-frequency oscillations. For many years, this subject has been clouded by the subtle and precarious nature of the required balance between the mass and wind fields and by the apparent insensitivity of many actual forecasts to the choice of how to initialize or even whether to initialize, given a sufficiently smooth analysis. One viewpoint is that the relatively slow and weak typical interaction (between the meteorologically significant signal and the unwanted fast gravity-inertial waves) relegates the importance of initialization to a mainly cosmetic role, so that the noise can be simply reduced by some dissipative mechanism in the model or even filtered out at the end of the forecast ("finalization"). On the other hand, it might reasonably be hoped that a correct initialization procedure would be advantageous in improving the simulation of vertical motion-dependent condensation and precipitation processes, in allowing for realistic diabatic effects, and in reducing the period at the beginning of the forecast during which the model fields are coming into quasi-equilibrium with one another, a period in which skill often appears to be decreasing rapidly. In general, one might expect the importance of initialization to increase as models become more complex and physically comprehensive, as the data base evolves from conventional surface-based and radiosonde observations toward a more heterogeneous observing system, and as one begins to ask more of the forecast models— not merely velocities and temperatures in the free atmosphere, but also predictions of cloudiness, surface temperature and stress, precipitation, etc.

It is thus fortunate and timely that the initialization problem, at least as usually posed, appears to have been solved in principle by the introduction of so-called nonlinear normal mode procedures(Baer, 1977; Baer and Tribbia, 1977; Machenhauer, 1977). The essential feature of such procedures lies in the separation of the gravity and Rossby modes by projecting the data onto the normal modes of the model itself, and then altering the initial time tendencies of the gravity modes so as to suppress the unwanted oscillations. The basic idea has now been refined and applied successfully to several models (e.g., Andersen, 1977; Daley, 1978). It was implemented for the ECMWF global model by Temperton and Williamson (1979), who remark that, "The computational feasibility of normal mode initialization depends crucially on the fact that the model's linearized equations are separable in the vertical and horizontal directions. Application to global (or hemispheric) grid-point models based on anything other than a regular latitude-longitude grid seems to be ruled out.

"Limited-area models based on a regular latitude-longitude grid can probably be handled; the corresponding normal modes should be derived under the assumption that the time-derivations are zero at any non-periodic boundaries (unless these are regarded as free-slip walls), and non-homogeneous boundary conditions could be absorbed into the nonlinear forcing. Limited-area models based on stereographic projections may possibly be amenable to nonlinear normal mode initialization provided that the domain is rectangular—any other shape would not allow the required separability. It would presumably be necessary to set up the linearized equations assuming a constant Coriolis parameter and constant map factors (or at least a function of one coordinate only), and to absorb deviations from these constant values into the "nonlinear" terms; for this reason the procedure is likely to work only for limited-area models of modest horizontal extent. Further study is required before a firmer assessment can be made of the applicability of normal mode initialization to limited-area models."

The ECMWF limited-area model has not yet been widely used or documented, and from the viewpoint of operational implementation, the model is largely untested and unsupported by needed data-processing software, such as analysis, initialization, and post-processing and display programs. However, in addition to operational fine-mesh models, such as the LFM at the U.S. National Meteorological Center, a number of research models have been developed and tested more or less thoroughly. Prominent among these are a model developed at Pennsylvania State University (Anthes and Warner, 1978; Warner et al., 1978) and one developed at Drexel University (Perkey, 1976; Perkey and Kreitzberg, 1976).

A typical configuration for the Penn State model is six layers at sigma levels 0.0, 0.25, 0.4, 0.55, 0.7, 0.85, 1.0 with a top at 250 mb, time step of 180 sec, grid length of 120 km, and a mesh of 30 by 35 points. The Penn State group has recently been working on initialization procedures, and finds evidence that improved initialization has a positive impact on precipitation forecasts (Tarbell et al., 1979). This result is not inconsistent with earlier work by the Drexel group (e.g., Perkey, 1976), which showed a strong sensitivity of the precipitation forecast to the initial moisture specification.

Thus it seems reasonable to conclude that development of nonlinear normal mode initialization procedures for the KNMI system will in the long run be worthwhile and beneficial, although the forecasting system will be usable and useful prior to the implementation of such an initialization scheme. Initialization is thus somewhat similar to analysis, in that introducing state-of-the-art technology deserves a high priority from a scientific viewpoint and promises eventual significant benefits in practice, but from a practical viewpoint the development of advanced techniques may reasonably be postponed in favor of work of greater urgency. As resources permit, the development of improved analysis and initialization schemes may be carried out in parallel with the implementation of the new system in its initial configuration, i.e., one based on the existing Cressman analysis of heights and a simple balanced wind field. As new analysis and initialization modules, quite possibly developed in collaboration with cooperating centers elsewhere, become available, they can be readily introduced into the system and tested.

4. Numerics

A. Horizontal Resolution

Integration time is determined primarily by horizontal resolution, which in the past has generally meant that computer speed and operational schedules, rather than scientific requirements, limited the grid size to relatively coarse values in the 300 to 400 km range. Improvements in numerical methods (e.g., semi-implicit time schemes) and computers (e.g., the Cray-1) have only recently made it feasible to use sufficiently fine grids so that horizontal truncation is no longer the leading error source of large-scale forecasts (Somerville, 1976, 1979).

For example, the ECMWF global model (48 grid points between pole and equator, 15 levels, 15 minute time step) requires less than 3 Cray central processor hours (and less than 4 wall clock hours, since the input/output overhead time is small) for a 10-day forecast. Thus a core-contained limited-area version (such as the 15-level, one-degree grid on a domain of 40 x 60 points, currently being tested at ECMWF), once the appropriate Fourier filter for high latitudes is incorporated so that a time-step of about 7 minutes (semi-implicit) is possible, should run in less than 5 minutes per day on the Cray.

Such high model speed permits testing and developing work via remote access to the ECMWF machine at an acceptable rate, assuming reasonable turnaround times and system reliability as well as a sufficient allocation of machine time. In fact, the pacing item for model development is likely to be the availability of scientists and programmers rather than either hardware or software limitations.

Model speeds in this range should also allow operational running of the model on computers which are between one and two orders of magnitude slower than the ECMWF Cray system. The operational feasibility, however, depends critically on many factors including model domain, vertical resolution, horizontal resolution, model physics, programming efficiency, machine speed and memory size, data processing and analysis and initialization procedures, and operational scheduling requirements. So many of these factors are unknown at present that it is not possible to state what domain and resolution will be practicable, but it does seem very likely that the fundamental proposition, that a model of this type at KNMI will be usable operationally, is true. It is clear, however, that an important aspect of model development will be the examination of trade-offs between speed and model performance to determine optimal domain and resolution requirements.

In this connection, horizontal resolution will dominate the determination of running time, which varies essentially inversely as the cube of grid size. It is thus fortunate that the one degree grid (actually 0.9875 degrees) now being used in the ECMWF configuration may be unnecessarily small. It may well turn out, for example, that with presently available data, a point of diminishing returns is reached when the grid size is reduced to only 1.5 degrees. Obviously, this question must be decided by numerical experimentation.

B. Vertical Resolution

The ECMWF model, like nearly all modern models, uses a form of sigma (pressure normalized by surface pressure) as a vertical coordinate (Phillips, 1957). This terrain-following pressure coordinate tends naturally to concentrate the vertical resolution in the lower atmosphere even if the model layers are equally spaced to minimize formal vertical truncation error, because pressure decreases most rapidly with height at low heights. However, much experience has shown that it is generally advantageous to concentrate the layers even more in the boundary layer than would be the case with equal spacing. Once again, only numerical experimentation can decide such a question definitively.

C. Lateral Boundary Conditions

Despite much research, there is not yet any general agreement as to the proper method of formulating lateral boundary conditions for limited-area models. The ECMWF model uses a relaxation technique due to Davies (1976) and Källberg (1977), which seems to work well in its current implementation (Gibson and Källberg, 1977), despite its lack of mathematical sophistication. This result is consistent with recent results of Perkey and Baumhefner (1979) at NCAR who found that the simple damped one-way interaction scheme of Perkey and Kreitzberg (1976) was not inferior in practice to the theoretically better technique of Williamson and Browning (1974). From a pragmatic viewpoint, it is probably true that the more important source of boundary errors is improper information originating in the large-scale model within which the limited-area model is (at least conceptually) imbedded, rather than the details of the mathematical formulation of the boundary condition. Thus the route to decreased boundary-induced error lies in improving the data which are to be imposed at the boundary, or in expanding the domain to prevent propagation of errors from the boundary to the region of interest within the time period of interest, rather than in devising more ingenious formulas (Miyakoda and Rosati, 1977).

D. Recommended Numerical Experimentation

The basic structure of the ECMWF model is conventional and well-understood. There is no reason to question the fundamental adequacy of the second-order finite-difference scheme with the Arakawa "C" grid in latitude-longitude-sigma coordinates. A great deal of valuable development work on the limited-area version of the model, including the lateral boundary conditions, has already been done at ECMWF. To be able to start with the ECMWF code represents a very significant advantage for the KNMI and a great saving in time and resources.

The highest priority should be given first to the task of becoming familiar with the model as supplied by ECMWF and of modifying and adapting this code so that it is running and usable at the KNMI. Recent essential numerical improvements (e.g., Fourier filter, semi-implicit scheme) should be incorporated, but it should not be absolutely necessary to include the latest ECMWF physical refinements. In fact, for testing sensitivity to resolution, there is a great deal to be said for using a nearly adiabatic and nearly inviscid model version, because many typical parameterizations are themselves very resolution-sensitive.

Once the code is running smoothly, and an affordable resolution and domain have been determined, the most urgent task will be to establish input/output linkages between the model module itself (including lateral boundary conditions) and the existing KNMI programs, so that the model can be run from the operational analysis and the model output can be processed.

At this point it will be possible to use real-data forecasting to test sensitivity to resolution, domain, initialization and the source of the lateral boundary conditions (which can be supplied both by ECMWF and by the KNMI filtered model). In these tests, the operational filtered model will be running from the same analysis and will be an obvious standard of comparison.

There are certain to be considerable difficulties encountered in bringing a large and unfamiliar code to a usable status on a new system, and in establishing linkages with elements of the present operational environment, and so during this phase it will probably be prudent to postpone any contemplated major modifications of the ECMWF code, with the possible exception of relocating the pole on the model grid for better grid homogeneity in the region of interest. More drastic changes, such as the introduction of fourth-order-accurate horizontal space differencing, which is generally more cost-effective in reducing truncation error than simple grid refinement, should be deferred until the model in its present form has been implemented optimally.

5. Physics

The area of greatest potential value for improving the quality of numerical weather forecasts may well be the parameterized "physics," i.e., the representations of small-scale source and sink terms for momentum, heat and water in the conservation equations. In fact, a strong case can be made for the proposition that strictly numerical improvements are unlikely to produce dramatic increments in forecast skill without simultaneous improvements in the data base and in the model physics (e.g., Williamson, 1978; Somerville, 1980).

This area of great opportunity, however, is also one of great difficulty, as is well known. Difficulties arise largely because of inadequate basic physical understanding of phenomena such as those of the boundary layer and of moist processes and their interaction with larger scales. Progress in parameterization thus depends directly on progress in many areas of atmospheric science, and so it is naive to expect dramatic breakthroughs on short time scales. One strong reason for developing a primitive-equation model at the KNMI, however, is to have a convenient vehicle for developing and testing parameterizations which are especially appropriate to local forecasting requirements.

One strategy which suggests itself for the immediate future, therefore, is to begin numerical experimentation by testing and tuning the ECMWF parameterizations which are likely to have the greatest impact on short-range forecasts. These include the boundary layer (Louis, 1979) and moist convection (Kuo, 1974) as well as the large-scale condensation criteria. There is considerable evidence (e.g., Dobosy and Somerville, 1979) of the short-range sensitivity of numerical predictions to boundary-layer parameterizations, but no concensus as to the best approach to follow (e.g., Deardorff, 1972; Arya, 1977). Similarly, the ECMWF cumulus parameterization represents one well-regarded conventional approach, but, as in the case of the boundary layer, there are many alternatives (e.g., Arakawa and Schubert, 1974) well worth exploring. Each of these subjects has a large literature of its own, which cannot be surveyed here, and is currently in a state of very active development in many countries.

It may be possible to postpone the development of other parameterizations, such as radiation and turbulent diffusion, until after initial implementation of the model. It is difficult to escape the conclusion, however, that some sort of radiation scheme, almost certainly including diurnal effects and depending on prognostic cloud and water vapor fields, will eventually be needed for input to the surface heat balance calculation. The need for including diffusion terms outside the boundary layer, and the best form (if any) for such terms, is much less clear (Williamson, 1978). Numerical tests of model sensitivity to these parameterizations will eventually be necessary.

6. Prognostics

An indirect advantage of an operational primitive-equation forecasting system is that such a system makes possible the development of techniques for post-processing the model predictions so as to produce forecast information optimally designed for local applications. Among the most useful of such techniques is the use of model variables as input to statistical regression schemes which produce forecasts couched in terms of local weather. The "Model Output Statistics" system is an example of this approach (e.g., Glahn and Lowry, 1972; Klein and Glahn, 1974). Useful automated techniques have been devised for even such traditionally troublesome variables as quantitative precipitation (Bermowitz and Zurndorfer, 1979).

Recently, promising techniques have been proposed for combining statistical and dynamical forecasting methods in a more sophisticated manner than the use of linear regression as post-processing for a dynamical forecast in the manner of "Model Output Statistics." For example, Lorenz (1977) has demonstrated that nonlinear regression has attractive possibilities, while Faller and Lee (1975) and Faller and Schemm (1977) have explored the idea of statistically correcting as often as every time step during the dynamical integration, rather than once at the end. These techniques have not yet been tested on comprehensive primitive-equation models or applied in operational practice, and so their usefulness remains an open question.

Although it is very difficult to make accurate judgments as to the relative priority which each of these areas deserves in the development of the KNMI system, one should be alert to the possibility that effort diverted from model development to postprocessing may be very worthwhile, especially once the model configuration has stabilized and the model skill is approximately comparable to that of other operational models.

7. Diagnostics

In addition to the usual forecast verification programs, which will necessarily evolve over a considerable period of time, one other topic deserves special mention. Among the most useful diagnostic tools for the study and improvement of primitive-equation models is a set of programs for computing and displaying the model energy budget (e.g., Baker et al., 1978; Hauser and Miller, 1978). Such programs provide essential insight into the energetics of the model atmosphere and are invaluable aids in examining the effects of changes in model physics and numerics. As the complexity of models increases, the need for powerful diagnostic aids becomes greater. Although the initial effort to prepare comprehensive energetics analysis programs is not small, it is very worthwhile. Such programs, once they are operational, can be maintained fairly easily and tend to have long useful lifetimes.

Acknowledgments

This report resulted from the author's visit in November and December 1979 to the Koninklijk Nederlands Meteorologisch Institut (KNMI). During this period the author was on leave from the National Center for Atmospheric Research (NCAR) and from the Scripps Institution of Oceanography (SIO) of the University of California, San Diego. Financial support for the visit was shared by KNMI and by NCAR, which is sponsored by the United States National Science Foundation. The author is grateful to all of these institutions for making this visit possible and wishes especially to thank Prof. Dr. Ir. H. Tennekes and Dr. A.P.M. Baede, respectively Director of Scientific Research and Head of Meteorological Research Section A at KNMI, for inviting him, and Dr. W. A. Nierenberg, Director of SIO, the author's primary employer, for arranging a leave of absence.

The author benefited greatly from discussions with many scientists at K.N.M.I., including Dr. H. M. van den Dool, Drs. A. W. den Exter Blokland, Dr. L. M. Hafkenscheid, Dr. L. C. Heijboer, Dr. G. P. Können, Drs. W. J. A. Kuipers, Drs. J. van Maanen, Drs. J. Oerlemans, Ir. J. D. Opsteegh, Dr. J. Reiff, Drs. J. W. Sanders, Prof. Dr. C. J. E. Schuurmans, Mr. H. Timmerman, Drs. A. P. van Ulden, and Dr. J. Wieringa. The opinions expressed in this report, however, are solely the responsibility of the author.

Author's Address:

Prof. R. C. J. Somerville
Head, Climate Research Group, A-024
Scripps Institution of Oceanography
University of California, San Diego
La Jolla, California 92093
U.S.A.

REFERENCES

LEGEND:

AM	Annalen der Meteorologie (Neue Folge)
BAMS	Bulletin of the American Meteorological Society
BLM	Boundary Layer Meteorology
BPA	Beiträge zur Physik der Atmosphäre
ECMWF	European Centre for Medium Range Weather Forecasts
IR	Internal Report
TR	Technical Report
FCNWP	Fourth Conference on Numerical Weather Prediction, American Meteorological Society
JAM	Journal of Applied Meteorology
JAS	Journal of the Atmospheric Sciences
MWR	Monthly Weather Review
QJRMS	Quarterly Journal of the Royal Meteorological Society

Andersen, J. H., 1977: ECMWF, IR 15.

Anthes, R. A., and T. T. Warner, 1978: MWR, 106, 1045.

Arakawa, A., and W. H. Schubert, 1974: JAS, 31, 674.

Arya, S. P. S., 1977: MWR, 105, 215.

Baer, F., 1977: BPA, 50, 350.

Baer, F. and J. Tribbia, 1977: MWR, 105, 1536.

Baker, W. E., et al., 1978: MWR, 106, 311.

Bermowitz, R. J., and E. A. Zurndorfer, 1979: MWR, 107, 122.

Cressman, G. P., 1959: MWR, 87, 367.

Daley, R., 1978: Tellus, 30, 201.

Davies, H. C., 1976: QJRMS, 102, 405.

Deardorff, J. W., 1972: MWR, 100, 93.

Dobosy, R. J., and R. C. J. Somerville, 1979: BPA, 52, 190.

Faller, A. J., and D. K. Lee, 1975: MWR, 103, 845.

Faller, A. J., and C. E. Schemm, 1977: MWR, 105, 37.

Gibson, J. K., and P. Kållberg, 1977: ECMWF, IR 7.

Glahn, H. R., and D. A. Lowry, 1972: JAM, 11, 1203.

Hauser, R. K., and A. J. Miller, 1978: MWR, 106, 607.

Kållberg, P., 1977: ECMWF, IR 3.

Klein, W. H., and H. K. Glahn, 1974: BAMS, 55, 1217.

Kuo, H.-L., 1974: JAS, 31, 1232.

Lorenz, E. N., 1977: MWR, 105, 590.

Louis, J.-F., 1979: BLM, 17, 187.

Machenhauer, B., 1977: BPA, 50, 253.

Miyakoda, K., and A. Rosati, 1977: MWR, 105, 1092.

Otto-Bliesner, B. L., et al., 1977: MWR, 105, 1083.

Perkey, D. J., 1976: MWR, 104, 1513.

Perkey, D. J., and D. Baumhefner, 1979: FCNWP, 172.

Perkey, D. J., and C. W. Kreitzberg, 1976: MWR, 104, 744.

Phillips, N. A., 1957: J. Meteor., 14, 184.

Phillips, N. A., 1976: BAMS, 57, 1225.

Schlatter, T. W., 1975: MWR, 103, 246.

Schlatter, T. W., et al., 1976: MWR, 104, 765.

Somerville, R. C. J., 1976: AM, 11, 266.

Somerville, R. C. J., 1979: FCNWP, 182.

Somerville, R. C. J., 1980: JAS, 37, in press.

Tarbell, T. C., et al., 1979: FCNWP, 61.

Temperton, C., and D. L. Williamson, 1979: ECMWF, TR 11.

Warner, T. T., et al., 1978: MWR, 106, 1079.

Williamson, D. L., 1978: MWR, 106, 69.

Williamson, D. L., and G. L. Browning, 1974: JAM, 13, 8.